JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

On the Accuracy of the Adiabatic Separation Method

RANDALL S. CASWELL AND MICHAEL DANOS
National Bureau of Standards Washington D.C. 20234

(Received 31 March 1969)

Numerical experiments performed for a model of two strongly coupled oscillators indicate that the adiabatic separation method yields accurate results even where the condition of adiabaticity is violated to a very high degree, except in those cases where two levels are degenerate in the adiabatic approximation. An accurate solution for those cases can be obtained by diagonalizing the 2 × 2 Hamiltonian submatrix built on the two degenerate adiabatic states. It is conjectured that the adiabatic separation method can be expected quite generally to yield highly accurate results, at least for states belonging to the discrete spectrum.

I. INTRODUCTION

It has been useful in many cases to employ adiabaticity arguments in the treatment of a system with strongly interacting degrees of freedom. In the resulting adiabatic approximation, the degrees of freedom are dynamically decoupled and a separation of the equations describing the system is achieved by neglecting certain kinetic energy terms. Other kinetic energy terms, which are of the same order of magnitude, are kept. Usually, this makes good sense because of physical arguments. Sometimes, however, the adiabatic separation method has been applied in such situations where the physical arguments would cast doubt on the accuracy of the method. Nevertheless, the obtained results turned out to be quite reliable.

In the present paper, we set out to make numerical tests concerning the accuracy of the adiabatic approximation by comparing the presumably very accurate solutions obtained by diagonalizing very large matrices with the solutions obtained from integrating the differential equations resulting from making an adiabatic approximation. In these tests we used a model consisting of two coupled harmonic oscillators. It turned out that, at least in this case, the method works even when the conditions of adiabaticity are not only broken, but even negated.

We were motivated to undertake this investigation and were guided in the choice of the model by the collective model of nuclear physics.[1] In it, several modes of motion are strongly coupled, viz., the surface rotations, the surface vibrations, and the giant resonance oscillation.[2] In addition, in odd A nuclei, the motion of the "last" nucleon must be taken into account. It interacts with the surface modes via the potential well which depends on the shape of the nucleus. This system can be described by coupled anharmonic many-dimensional oscillators. The characteristic frequencies of the different modes of the system have approximately the ratios 0.1:1:15:10. Since these ratios are 0.1 or smaller, except for the last pair, which is uncomfortably close to 1, it has been assumed that the adiabatic separation method should be quite accurate. As already mentioned, this assumption turned out to be quite well founded.

We found, in short, that the adiabatic approximation is quite accurate for an astonishingly large range of the relevant parameters. In terms of the parameter λ specifying the adiabaticity, i.e., $\lambda = \omega_2/\omega_1$, which is the ratio of the harmonic oscillator frequencies, the adiabatic approximation is supposed to be valid in the limit $\lambda \ll 1$. One should expect a complete breakdown of the adiabatic approximation for $\lambda \approx 1$. This did not turn out to be the case even for very strong coupling between the oscillators. Except for level degeneracies which happen at some particular values of λ, the adiabatic solutions remained quite accurate even for $\lambda > 1$. Thus, the only substantial discrepancies are connected with those solutions which are almost degenerate in the adiabatic method. We have not yet been able to pinpoint any reason for this surprising result. Finally, we investigated the level crossings. We found that the discrepancies in these regions can be eliminated to a large extent by mixing the two degenerate states, i.e., by constructing a 2 × 2 Hamiltonian matrix. The solutions of the system thus seem to be quite well represented by a single solution or, in the case of level crossings, by the two closest solutions of the adiabatic approximation.

II. THE MODEL

We take two 1-dimensional harmonic oscillators coupled by a term symmetric in both variables as a model. The simplest nontrivial system of this kind[3] is

[1] For a discussion from the point of view of atomic collisions see, e.g., J. Heinrichs, Phys. Rev. 176, 141 (1968).

[2] See, e.g., M. Danos and E. G. Fuller, Ann. Rev. Nucl. Sci. 15, 51 (1965).

[3] For a system which can be solved exactly see, e.g., P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Vol. II, p. 1718.

described by the Hamiltonian

$$\tilde{H} = \tfrac{1}{2}\hbar\omega_1\left[\left(-\frac{\partial^2}{\partial\xi^2} + \xi^2\right) + \lambda\left(-\frac{\partial^2}{\partial\eta^2} + \eta^2\right) + 4\kappa\xi^2\eta^2\right] \equiv \tfrac{1}{2}\hbar\omega_1 H. \quad (1)$$

Here,

$$\lambda = \omega_2/\omega_1. \quad (2)$$

To obtain the adiabatic approximation, one puts

$$\Psi = \varphi(\eta)\psi(\xi,\eta), \quad (3)$$

which leads to

$$H\Psi = \varphi\left(-\frac{\partial^2\psi}{\partial\xi^2} + \xi^2\psi\right) + \lambda\psi\left(-\frac{\partial^2\varphi}{\partial\eta^2} + \eta^2\varphi\right) - \lambda\left(2\frac{\partial\varphi}{\partial\eta}\frac{\partial\psi}{\partial\eta} + \varphi\frac{\partial^2\psi}{\partial\eta^2}\right) + 4\kappa\xi^2\eta^2\varphi\psi = E\Psi. \quad (4)$$

This equation separates into two uncoupled equations if one drops the term containing the "wrong" derivatives, i.e.,

$$\lambda\left(2\frac{\partial\varphi}{\partial\eta}\frac{\partial\psi}{\partial\eta} + \varphi\frac{\partial^2\psi}{\partial\eta^2}\right), \quad (5)$$

on the basis of the adiabaticity condition $\lambda \ll 1$. Note that the magnitude of the term (5) is of order λ only with respect to the first parenthesis of (4). It is of the same order as the second parenthesis of (4), viz., linear in λ. To emphasize, in the adiabatic method one thus drops selectively only a part of the terms proportional to λ. Therefore, λ is not actually a parameter equivalent to the small parameter of a systematic perturbation procedure. Uncertainties of the order λ, in other words, of the order of the level spacing of the low-energy system, i.e., of order $\hbar\omega_2$, therefore cannot be excluded. As we will see below, they do, in fact, show up.

At any rate, the two separated eigenvalue equations are[4]

$$H_1\psi_n = -\frac{\partial^2\psi}{\partial\xi^2} + (1 + 4\kappa\eta^2)\xi^2\psi_n = E_n(\eta)\psi_n, \quad (6)$$

$$E_n(\eta) = (2n + 1)(1 + 4\kappa\eta^2)^{\frac{1}{2}}, \quad (7)$$

and

$$H_2\varphi_{nm} = \lambda\left(-\frac{\partial^2}{\partial\eta^2} + \eta^2 + \frac{2n+1}{\lambda}(1 + 4\kappa\eta^2)^{\frac{1}{2}}\right)\varphi_{nm} = E_{nm}\varphi_{nm}. \quad (8)$$

The eigenvalues and eigenfunctions of (8) were found numerically for a series of values of the parameters λ and κ.

[4] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1968), 3rd ed., p. 66.

To determine the accuracy of these solutions we solve the Schrödinger equation with the Hamiltonian (1) by constructing the Hamiltonian matrix in the basis of the free-harmonic-oscillator solutions corresponding to $\kappa = 0$ in (1). In other words, for Ψ we write the expansion

$$\Psi_n = \sum_{k,l} a_{kl}^{(n)} u_k(\xi) v_l(\eta), \quad (9)$$

where u and v are the 1-dimensional harmonic-oscillator wavefunctions.[4] Then the elements of the Hamiltonian matrix are

$$\begin{aligned}\langle k'l'|\, H\,|kl\rangle = {} & \delta_{k'k}\delta_{l'l}[(2k+1) + \lambda(2l+1)] \\ & + \kappa\{\delta_{k',k-2}[k(k-1)]^{\frac{1}{2}} + \delta_{k'k}(2k+1) \\ & + \delta_{k',k+2}[(k+1)(k+2)]^{\frac{1}{2}}\} \\ & \times \{\delta_{l',l-2}[l(l-1)]^{\frac{1}{2}} + \delta_{l'l}(2l+1) \\ & + \delta_{l',l+2}[(l+1)(l+2)]^{\frac{1}{2}}\}. \end{aligned} \quad (10)$$

Thus, the matrix splits into four disconnected matrices for the states of the four possible different symmetries. We consider only those states for which both k and l are even.

The matrix was truncated at suitable values of k and l. The following simple truncation procedure was employed. To begin with, for each value of k a certain $l_{\max}$ was chosen. After the diagonalization it was checked whether the amplitudes $a_{kl}^{(n)}$ of (9) at the different $l_{\max}$ had become sufficiently small. In those k-blocks where this seemed necessary, $l_{\max}$ was changed by 1, 2, or 3. Also, it was checked whether adding another value of k with $l_{\max} = 1$ was required. This procedure was continued until a satisfactory convergence of the amplitudes was achieved. It was found that going to excited states, i.e., with increasing n, the convergence very rapidly diminishes. In other words, the number of harmonic-oscillator states which participate in a given state of the system increases very rapidly with increasing eigenvalue. In order to achieve reasonable convergence, therefore, matrices of sizes up to 130 × 130 had to be employed for the case $\kappa = 0.75$. Then the amplitude printout for the ground state ($n = 0$) contained mostly zeros, while for $n = 6$ convergence was just tolerable.

III. RESULTS

For a given $\kappa > 0$, the adiabatic approximation is supposed to approach the exact solution as $\lambda \to 0$. This should be true independently of the magnitude of κ. On the other hand, as κ decreases the oscillators decouple and the solution becomes trivial. After some tryouts most of the numerical calculations were done for $\kappa = 0.75$, which in the terminology of the nuclear

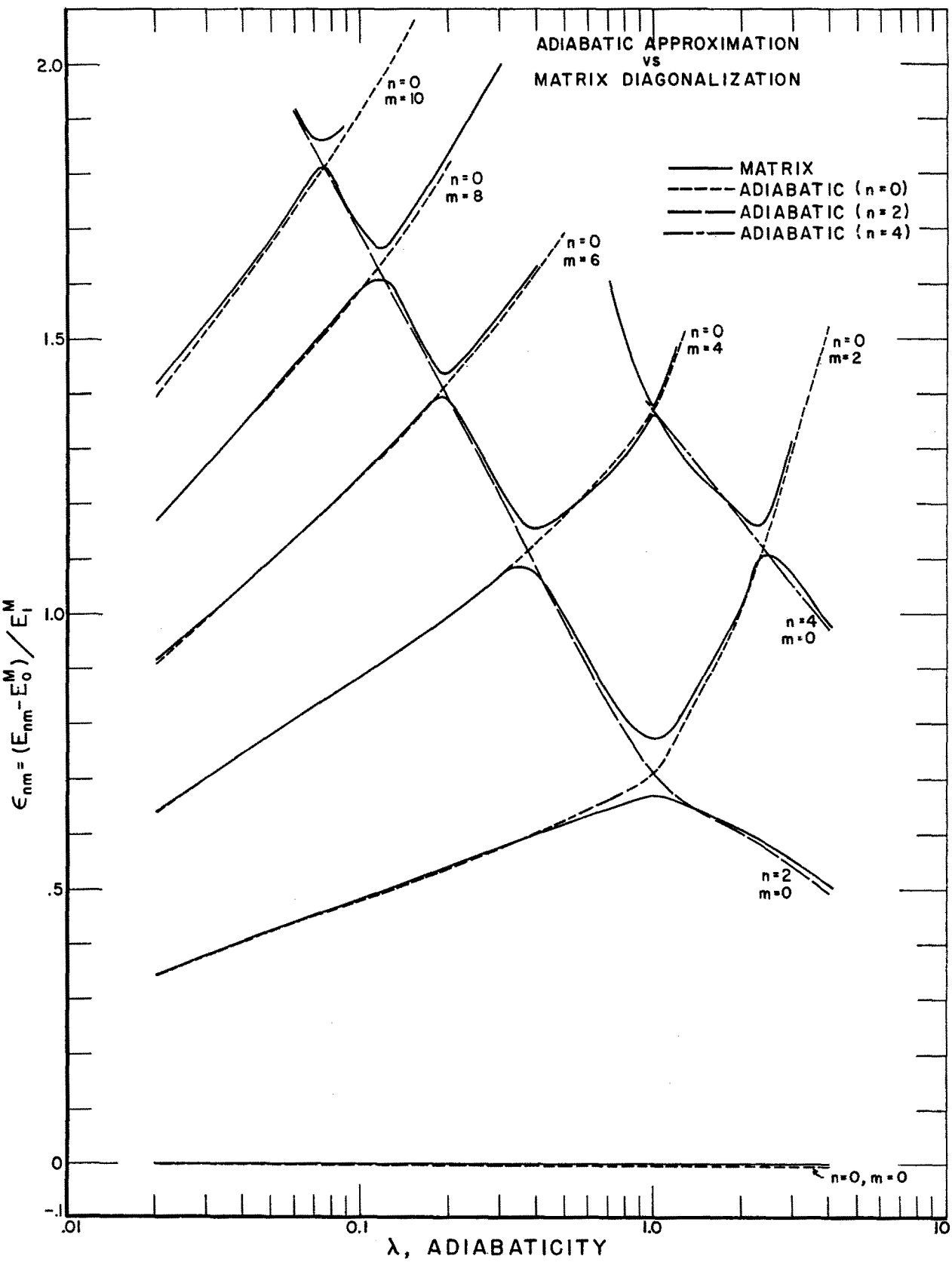


FIG. 1. Comparison of adiabatic approximation and matrix diagonalization solutions for a coupling parameter $\kappa = 0.75$.

collective model corresponds to strong coupling. The strength of the coupling is best judged from the magnitude of the contributions of the several terms of the Hamiltonian to the matrix elements (10). The coupling term, the product of the two braces containing the factor κ, contributes even to the lowest diagonal element an amount comparable in magnitude with that of the unperturbed oscillators. Also, even the lowest off-diagonal element is comparable in magnitude with the adjacent diagonal elements. Thus, a very thorough mixing of the harmonic oscillator states is to be expected here. This indeed turned out to be the case and, as already mentioned, for the considered part of the eigenvalue spectrum, i.e., the lowest six or seven states, a 130 × 130 matrix has to be employed.

The results were plotted in terms of the reduced eigenvalue variable

$$\epsilon_{nm} = (E_{nm} - E_0^M)/E_1^M \qquad (11)$$

in order to be able to represent the results on a graph in such a way that the trivial dependences have been eliminated. In (11), the ground-state energy of the matrix solution is denoted by E_0^M and the energy of the first excited state of the matrix solution is denoted by E_1^M. The form (11) was chosen for the reduced energy in preference to other possibilities, e.g., $(E_{nm} - E_0^M)/(E_1^M - E_0^M)$, in order to minimize the distortions from the straightforward plot of E_{nm}.

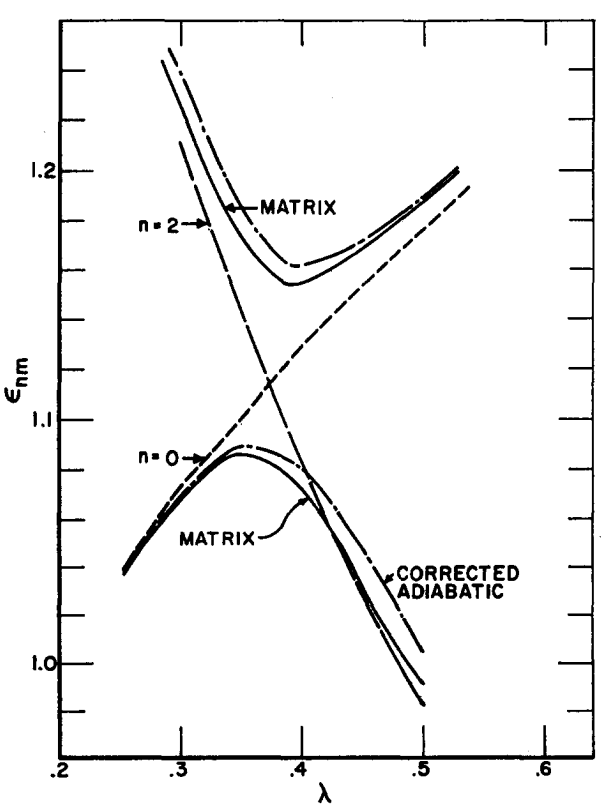


FIG. 2. Comparison of matrix, adiabatic, and corrected adiabatic solutions using Eq. (12) for a crossing near $\lambda = 0.4$.

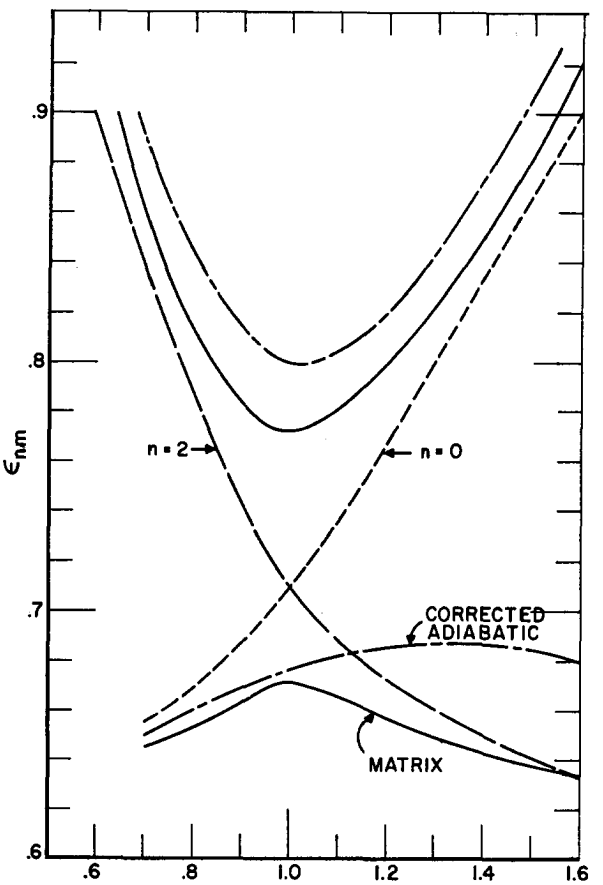


FIG. 3. Comparison of matrix, adiabatic, and corrected adiabatic solutions using Eq. (12) for a crossing near $\lambda = 1.0$.

The following observations can be made (see Fig. 1):

(i) The overall accuracy of the adiabatic solutions is uncannily good. Considering first the ground state; one sees that the error of the adiabatic solution builds up very slowly as λ increases and is still quite small even at $\lambda = 4$, where the condition $\lambda \ll 1$ clearly is totally violated. It is not evident why this is true; there seems to be no *a priori* reason for the term (5) to be small unless λ is small.

(ii) Considering the region $\lambda \ll 1$, e.g., $\lambda = 0.05$, one notices an even–odd effect: The matrix and the adiabatic eigenvalues differ more for the second, fourth, and sixth than for the first, third, and fifth states. Since the adiabatic approximation for such small λ should be exceedingly accurate, one has to conclude that this discrepancy results from inaccuracies of the matrix eigenvalues.

(iii) The adiabatic approximation allows a classification of the states according to the quantum numbers of the solutions of (6) and (8). In contrast to that, the

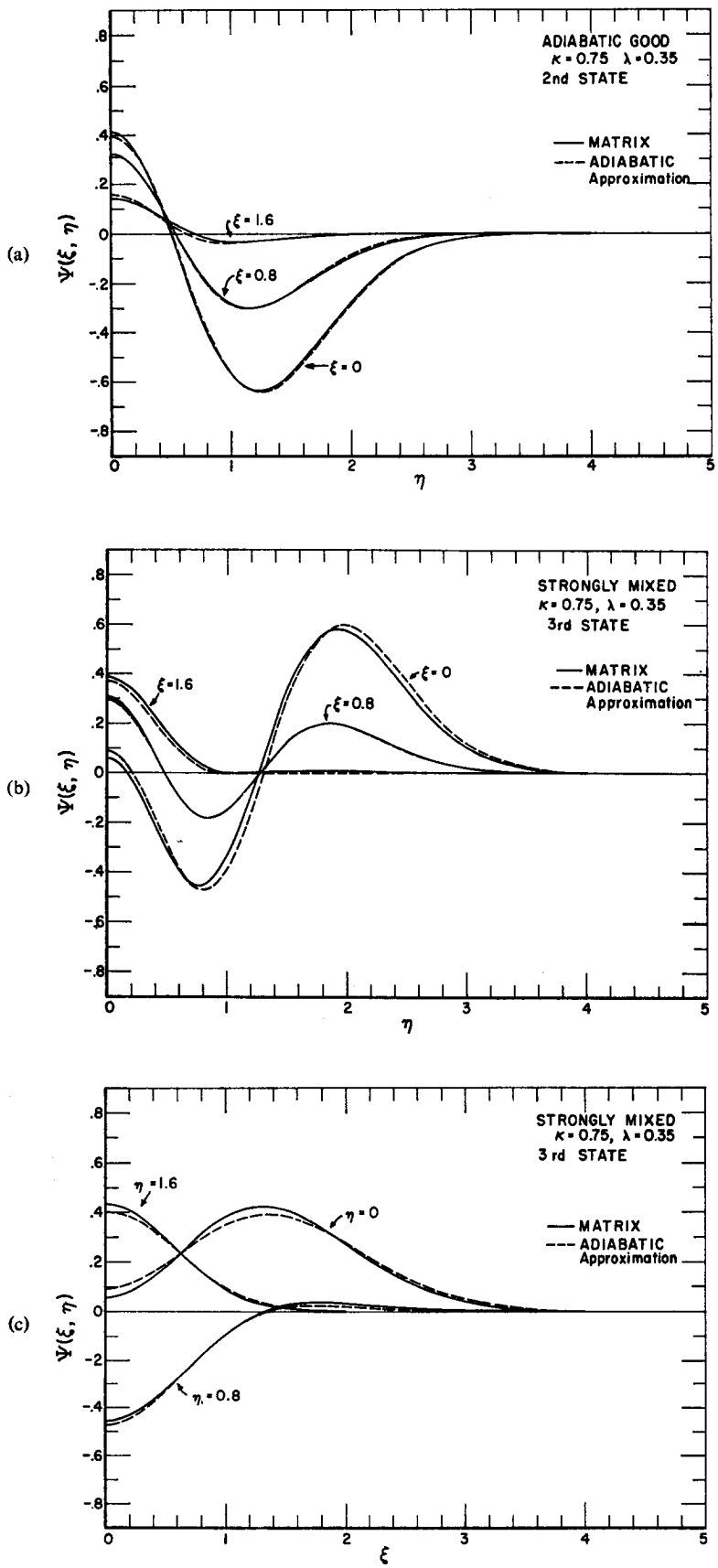


FIG. 4. Comparison of eigenfunctions from matrix and adiabatic approximation calculations. In (a) no special correction is applied to the adiabatic solution; in (b) and (c) the eigenfunction is constructed from the two adiabatic solutions closest to the crossing.

matrix eigenvalues just fall as they happen and, furthermore, as long as the coupling is strong, it is impossible to tell anything from the eigenvectors.

(iv) The only larger discrepancies appear to be associated with rather small regions around eigenvalue degeneracies. Since the equations (6) and (8) of the adiabatic approximation are dynamically decoupled, there is no reason for the eigenvalues associated with different quantum numbers to repel, while, naturally, no degeneracies occur in the solutions of the matrix.[5] The behavior of the eigenvalues in the crossing region suggests that the adiabatic solutions could be corrected by mixing the degenerate solutions via the term (5), i.e., by solving a 2×2 matrix.

(v) In the proper scale, i.e., in the scale of $\hbar\omega_2$, the residual discrepancies are independent of λ. In other words, the level repulsion at the different crossings is essentially a constant fraction of the level spacing.

In order to verify the conjecture of (iv) above, a check was performed for two crossing regions, viz., the lowest crossings around $\lambda = 1$ and $\lambda = 0.4$. Clearly, in neither of these cases is the condition $\lambda \ll 1$ fulfilled. We constructed the 2×2 Hamiltonian matrix by computing the matrix elements

$$h_{\alpha\beta} = -\lambda \int \varphi_\alpha \psi_\alpha \left(\varphi_\beta \frac{\partial^2 \psi_\beta}{\partial \eta^2} + 2 \frac{\partial \varphi_\beta}{\partial \eta} \frac{\partial \psi_\beta}{\partial \eta} \right) \partial\xi \partial\eta \quad (12)$$

of the neglected term (5) of the Hamiltonian (1), using the two crossing states. After diagonalization of this matrix, the only qualitative difference between the adiabatic and exact solution, viz., the degeneracy of the solutions, is eliminated. The resulting eigenvalues are shown in Figs. 2 and 3. The remaining differences indicate that some admixture from the more distant adiabatic solutions is contained in the exact solution. However, the amplitude of this admixture is very small. This is illustrated in Fig. 4 where both the exact and the two-state adiabatic eigenfunctions are plotted. The differences are of the order of percent, and any errors resulting from using the adiabatic eigenfunctions in the evaluation of matrix elements will be exceedingly small.

Some of the results obtained have a simple explanation, but we would like to re-emphasize that we have no explanation for the main result, i.e., for the large range of λ over which the adiabatic method works. In other words, it is not clear to us why the off-diagonal matrix elements $h_{\alpha\beta}$ of (12), both for $\alpha = \beta$ and $\alpha \neq \beta$, are so small: They are small only because of cancellations in the integration. The integrand itself is of reasonable magnitude. Perhaps even more significant is the fact that the quantity $h_{\alpha\beta}/\lambda$ turns out to be practically independent of λ. This shows that, indeed, a small λ is of help, but the quantity $h_{\alpha\beta}/\lambda$ itself happens to be small as compared with the diagonal energies, while it should have been expected to be of the same order of magnitude.

We do not think that the symmetry of the present model with respect to the exchange $\xi \leftrightarrow \eta$, $\lambda \leftrightarrow 1/\lambda$ is at the root of the puzzle. Namely, when making that exchange one is, in fact, solving a *different* problem, in which the adiabaticity condition may be satisfied. However, such an exchange is possible for every form of the interaction term: after the exchange one is led to a different equation which may then obey the adiabaticity condition.

In conclusion, it seems to us that one should seriously consider the use of the adiabatic separation method, even when, offhand, one would think it to be inapplicable. Although the results of this paper were obtained for a specific model, it is possible that they are quite generally valid, at least for systems with a discrete spectrum or for the discrete spectrum of more general systems. The adiabatic method could very well be substantially more economical than other possible solution methods.

[5] Note that the "crossing region" in the terminology of atomic collisions concerns crossings of different "potential energy" curves, i.e., of the quantities analogous to Eq. (7). (See, e.g., Ref. 1.) In the present model no such crossings occur.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Perturbation Theory for Large Coupling Constants Applied to the Gauss Potential

HARALD J. W. MÜLLER*
Physics Department, American University of Beirut, Beirut, Lebanon

(Received 7 May 1968; Revised Manuscript Received 2 June 1969)

As an example of a perturbation technique for large coupling constants g^2, we investigate the solutions and eigenvalues of the Schrödinger equation for a Gauss potential. In particular, we obtain the regular solution, valid for $r^2 < 1/|g|$, in terms of confluent hypergeometric functions by expanding the potential in the neighborhood of the origin. The Jost solution is obtained in an analogous manner in terms of a certain integral and is valid for $r^2 > 1/|4g|$. Both solutions are eigensolutions belonging to the same eigenenergy $E = k^2$. These eigenvalues are derived in the form of large-g asymptotic expansions which are useful and valid over a wide range of g. A noteworthy aspect of the investigation is the close analogy of the underlying mathematics with that of well-known periodic equations.

1. INTRODUCTION

Considerable insight into various aspects of scattering theory has recently been gained from an extensive study of potential theory. In particular, the interest in high-energy scattering for Yukawa potentials was revived with the hope of finding some hints on the behavior of the fully relativistic (field-theoretical) scattering amplitude. However, it has become clear also that Yukawa-like interactions, although they reproduce the Born approximation of a scattering amplitude remarkably well in most cases—even in elementary particle physics—do not generally suffice for higher-order approximations. Considerable doubt has been expressed on expecting the Regge trajectories in elementary particle physics to resemble those for Yukawa potentials. The need, therefore, arises for a better understanding of Regge trajectories (for instance) and of the scattering amplitude in general for other interactions besides the Yukawa potential.

In phenomenological theories, one of the most widely used interactions is the Gauss potential. This potential probably found its first application in the investigation of the ground state of the deuteron by Bethe (for a review see Bethe and Bacher[1]). However, despite its repeated application in many branches of theoretical physics and despite a large number of approximation and perturbation methods developed specifically for solving more complicated scattering problems, not a single attempt seems to have been made to examine the analytic properties of its ordinary, nonrelativistic scattering amplitude. Scattering and bound-state problems for the Gauss potential have therefore been studied exclusively by numerical methods.

We believe that analytic investigations as well as more effective mathematical techniques would help both the pure phenomenologist as well as the theoretician interested in a deeper understanding of the underlying scattering problem. Therefore, in the present paper, we examine as an example the solution of the Schrödinger equation for a Gauss potential $V(r) = -g^2e^{-a^2r^2}$ when the coupling constant g^2 is large. A field-theoretical model for large coupling constants has been considered, e.g., by Schiff.[2] In Sec. 2 we derive a complete large-g asymptotic solution of the S-wave radial wave equation in terms of parabolic cylinder functions. The solution, obtained by choosing the harmonic oscillator as the corresponding unperturbed eigenvalue problem, is valid near $r = 0$; therefore it is the regular solution. We also obtain a corresponding large-g asymptotic expansion for the eigenenergies. In Sec. 3 we obtain the corresponding results for all angular momenta, the solution in terms of confluent hypergeometric functions being valid for $r^2 < 1/|ga|$. (In view of the different nature of the solutions in the two cases, it is best to consider both separately.) In Sec. 4 we derive the large-g asymptotic expansion of the Jost solution valid for $r^2 > 1/|4g|$ in the region of bound-state energies. From a subsidiary condition we obtain again the same large-g asymptotic expansion for the eigenvalues as before. In Sec. 5 we examine the analytic continuation of these solutions and show that they are proportional to each other in their common range of validity.

Perhaps the most useful result is the expansion (3.30) for the eigenenergies—k^2 in terms of an odd-integer quantum number $q = 4n + 3$ $(n = 0, 1, 2, \cdots)$. We did not succeed in deriving a corresponding expansion for the phase shift (for the general method as applied to other potentials; see Ref. 3). We have

* Present address: Sektion Physik der Universität München, Lehrstuhl Prof. Bopp, München, Germany.

[1] H. A. Bethe and R. F. Bacher, Rev. Mod. Phys. **8**, 82 (1936).

[2] L. I. Schiff, Phys. Rev. **92**, 766 (1953).

[3] H. J. W. Müller, "Perturbation Approach for Regular Interactions," in *Beirut Lectures in Theoretical High-Energy Physics*, H. H. Aly, Ed. (Interscience Publishers, Inc., New York, 1968).

reason to believe that such an expansion may not exist: In the case of the Gauss potential, a stronger attraction of the particles is equivalent to an increase in the coupling constant g^2. One would therefore expect the behavior of the phase-shift for large g^2 to be similar to that for large k^2 (since $k^2 \sim -g^2$). However, an asymptotic expansion of the amplitude (and hence of the phase shift) in powers of $1/k^2$ does not exist[4] for a potential with vanishing odd-order derivatives at $r = 0$ [i.e., for a potential $V(r)$ which can be expanded in even powers of r around the origin]. This result is closely related to the observation of Bethe and Kinoshita,[5] that for such a potential a Regge pole and thus a bound-state wavefunction cannot be defined in the limit $|k^2| \to \infty$ in the ordinary way [since $\alpha(k) \to \infty$ for $|k^2| \to \infty$]. The behavior of Regge trajectories for the Gauss potential has been investigated by Aly, Müller, and Schilcher.[6] The present investigation, as also our previous paper,[7] reveals, however, an interesting mathematical feature of the Schrödinger equation for even-power potentials. This may be summarized by saying: For even-power potentials, the over-all analytic characteristics of the solutions are completely analogous to those of well-known periodic differential equations such as the Mathieu equation. Further investigations into the analytic behavior of the scattering amplitude for even-power potentials[8] may therefore be motivated also by solutions and properties of the Mathieu equation which have been studied in great detail.[9,10] Moreover, in both cases the mathematics is of the same degree of complexity (or simplicity); so that claims that the Schrödinger equation for a Gauss potential (for instance) be tractable only by numerical methods are disproved already by the vast literature on periodic differential equations. We discuss this analogy in slightly more detail in the conclusion.

2. S-WAVE SOLUTIONS NEAR THE ORIGIN

A. Approximate Behavior of the Eigenvalues in the Limit of Large Coupling Constants

We consider an S-wave radial Schrödinger equation ($\hbar = c = 1$, $m = \frac{1}{2}$) for the Gauss potential

$$V(r) = -g^2 e^{-a^2 r^2}, \tag{2.1}$$

g^2 and a^2 being both real and positive. Our first aim is to determine the approximate behavior of the eigenenergy k^2 under normal bound-state boundary conditions—however, in the limit of large positive values of the coupling constant g^2.

Since exponential x is an entire function of x, we may expand the potential (2.1) in ascending powers of r^2 and obtain

$$\left(\frac{d^2}{dr^2} + k^2 + g^2 - g^2 a^2 r^2\right)\psi = -g^2 \sum_{i=2}^{\infty} \frac{(-a^2)^i}{i!} r^{2i}\psi. \tag{2.2}$$

In this equation, we change the independent variable to

$$z = (2ga)^{\frac{1}{2}} r, \tag{2.3}$$

$$\left(\frac{d^2}{dz^2} + \frac{k^2 + g^2}{2ga} - \tfrac{1}{4}z^2\right)\psi = -\frac{1}{2}\sum_{i=2}^{\infty}\left(\frac{a}{g}\right)^{i-1} \frac{(-\frac{1}{2}z^2)^i}{i!}\psi. \tag{2.4}$$

In the limit $|g| \to \infty$, this equation becomes the parabolic cylinder equation

$$\left(\frac{d^2}{dz^2} + \tfrac{1}{2}q - \tfrac{1}{4}z^2\right)\psi_q = 0, \tag{2.5}$$

where

$$\psi_q(z) = D_{\frac{1}{2}(q-1)}(z) \tag{2.6}$$

is the parabolic cylinder function, and

$$q = (k^2 + g^2)/ga. \tag{2.7}$$

If, moreover, the boundary conditions—as those for bound states—require that ψ vanish at infinity, we have

$$q = 2n + 1, \quad n = 1, 2, 3, \cdots, \quad k^2 < 0. \tag{2.8}$$

The problem then reduces effectively to that of the harmonic oscillator. We have, therefore,

$$(k^2 + g^2)/ga = q - (2a/g)\Delta, \tag{2.9}$$

where Δ is an expansion of order zero in $1/g$. For $k^2 > 0$, of course, q is *not* given by (2.8); instead it is a function $q(k)$.

B. Solution of the S-Wave Schrödinger Equation

We now substitute (2.9) into (2.4) and multiply the equation by -2; then

$$\mathfrak{D}_q\psi = h\left(8\Delta - \sum_{i=0}^{\infty} h^i \frac{(2z^2)^{i+2}}{4(i+2)!}\right)\psi, \tag{2.10}$$

where

$$\mathfrak{D}_q \equiv -\left(2\frac{d^2}{dz^2} + q - \tfrac{1}{2}z^2\right) \tag{2.11}$$

[4] S. Rosendorff and S. Tani, Phys. Rev. **131**, 396 (1963).
[5] H. A. Bethe and T. Kinoshita, Phys. Rev. **128**, 1418 (1962).
[6] H. H. Aly, H. J. W. Müller, and K. Schilcher, Nucl. Phys. **B3**, 401 (1967).
[7] H. J. W. Müller, Ann. Physik (Leipzig) **21**, (1968).
[8] For a preliminary investigation see H. J. W. Müller, Z. Physik **205**, 149 (1967).
[9] J. Meixner and F. W. Schäfke, *Mathieusche Funktionen und Sphäroidfunktionen* (Springer-Verlag, Berlin, 1954).
[10] F. M. Arscott, *Periodic Differential Equations* (Pergamon Press Ltd., London, 1964).

and

$$h \equiv -a/4g. \tag{2.12}$$

As a first approximation to ψ, we have $\psi \simeq \psi^{(1)} = \psi_q$ and

$$\mathfrak{D}_q\psi_q(z) = 0. \tag{2.13}$$

It is now convenient to set

$$\psi_q(z) = D_{\frac{1}{2}(q-1)}(z)/[\tfrac{1}{4}(q-1)]!\,2^{\frac{1}{4}(q-1)}. \tag{2.14}$$

The factors introduced in this choice ensure that the expansion coefficients obtained below are particularly simple and symmetric.

Clearly the first approximation $\psi^{(1)}$ leaves uncompensated terms on the right-hand side of (2.10) amounting to

$$R^{(1)} = h\left(8\Delta - \sum_{i=0}^{\infty} h^i \frac{(2z^2)^{i+2}}{4(i+2)!}\right)\psi_q(z). \tag{2.15}$$

The recurrence relation for the functions ψ_q follows immediately from that for parabolic cylinder functions. Thus

$$2z^2\psi_q = (q+3)\psi_{q+4} + 2q\psi_q + (q-3)\psi_{q-4}$$

or, more generally,

$$(2z^2)^m\psi_q(z) = \sum_{i=-m}^{m} S_m(q, q+4i)\psi_{q+4i}(z), \tag{2.16}$$

where $S_1(q, q+4) = (q+3)$, etc.

Substituting now the relation (2.16) into (2.15), we have

$$R^{(1)} = \sum_{i=0}^{\infty} h^{i+1} \sum_{j=-(i+2)}^{i+2} [q, q+4j]_{i+1}\psi_{q+4j}, \tag{2.17}$$

where

$$[q, q]_1 = 8\Delta - (1/4\cdot 2!)S_2(q, q) \tag{2.18}$$

and—for i and j not simultaneously zero—

$$[q, q+4j]_{i+1} = -[1/4(i+2)!]S_{i+2}(q, q+4j), \tag{2.19}$$

with $|j| \leq i+1$.

Next we observe that

$$\mathfrak{D}_q\psi_{q+4j} = 4j\psi_{q+4j}. \tag{2.20}$$

Thus a term $\mu\psi_{q+4j}$ on the right-hand side of (2.11) or in (2.17) may be canceled out by adding to the previous approximation the contribution $(\mu/4j)\psi_{q+4j}$ except, of course, when $j = 0$. Hence the next contribution to $\psi^{(1)}$ becomes

$$\psi^{(2)} = \sum_{i=0}^{\infty} h^{i+1} \sum_{\substack{j=-(i+2)\\ j\neq 0}}^{i+2} \frac{[q, q+4j]_{i+1}}{4j}\psi_{q+4j}. \tag{2.21}$$

This contribution leaves uncompensated a sum of terms $R^{(2)}$, and these again lead to another contribution $\psi^{(3)}$. This process may be repeated indefinitely. Then, adding successive contributions and rearranging these in powers of h, we obtain

$$\psi = \psi^{(1)} + \psi^{(2)} + \psi^{(3)} + \cdots$$

and

$$\psi(z) = \psi_q(z) + \sum_{i=1}^{\infty} h^i \sum_{\substack{j=2i\\ j\neq 0}}^{-2i} P_i(q, 4j)\psi_{q+4j}(z). \tag{2.22}$$

The coefficients P_i in this expansion are found to be

$$P_1(q, \pm 8) = \frac{[q, q\pm 8]_1}{\pm 8}, \quad P_1(q, \pm 4) = \frac{[q, q\pm 4]_1}{\pm 4}, \quad \text{etc.} \tag{2.23}$$

For (2.22) to be a solution of (2.10), the sum of all terms in ψ_q—left uncompensated so far (e.g., in $R^{(1)}$)—must vanish. Thus

$$\begin{aligned} 0 = h[q, q]_1 + h^2\Big\{&[q, q]_2 + \frac{[q, q+4]_1}{4}[q+4, q]_1 \\ &+ \frac{[q, q-4]_1}{-4}[q-4, q]_1 \\ &+ \frac{[q, q+8]_1}{8}[q+8, q]_1 \\ &+ \frac{[q, q-8]_1}{-8}[q-8, q]_1\Big\} + \cdots. \end{aligned} \tag{2.24}$$

This is the equation from which Δ and hence the eigenenergy k^2 is determined. Evaluation of the terms up to $O(h^3)$ in (2.24) yields the expansion

$$\begin{aligned} k^2 + g^2 = qga &- \frac{3a^2(q^2+1)}{2^4} + \frac{q(11q^2+1)a^2}{2^6\cdot 3}h \\ &- \frac{(85q^4+2q^2-423)a^2}{2^9\cdot 3}h^2 + O(h^3). \end{aligned} \tag{2.25}$$

For small values of $|h|$ or large values of g, (2.25) is a proper asymptotic expansion with rapidly decreasing terms in magnitude.

3. GENERAL SOLUTIONS NEAR THE ORIGIN

A. Approximate Behavior of the Eigenvalues in the Limit of Large Coupling Constants

We consider the radial Schrödinger equation

$$\left[\frac{d^2}{dr^2} + k^2 - \frac{l(l+1)}{r^2} - V(r)\right]\psi = 0 \tag{3.1}$$

($\hbar = c = 1$, $m = \frac{1}{2}$) for the Gauss potential (2.1), g^2 and a^2 being again both real and positive. Our first aim is to determine the approximate behavior of the eigenenergy k^2 under normal bound-state boundary conditions—however, in the limit of large positive values of the coupling constant g^2.

Proceeding as before, we may rewrite (3.1) in the form

$$\frac{d^2\psi}{dr^2} + \left(k^2 - \frac{l(l+1)}{r^2} + g^2 - g^2a^2r^2\right)\psi = -g^2\sum_{i=2}^{\infty}\frac{(-a^2)^i}{i!}r^{2i}\psi. \quad (3.2)$$

In this equation we change the independent variable to

$$z = (2ga)^{\frac{1}{2}}r. \quad (3.3)$$

Then

$$\frac{d^2\psi}{dz^2} + \left(\frac{k^2+g^2}{2ga} - \frac{l(l+1)}{z^2} - \tfrac{1}{4}z^2\right)\psi = -\frac{1}{2}\sum_{i=2}^{\infty}\left(\frac{a}{g}\right)^{i-1}\frac{(-\frac{1}{2}z^2)^i}{i!}\psi. \quad (3.4)$$

In the limit $|g| \to \infty$, this equation may be approximated by

$$\frac{d^2\psi_0}{dz^2} + \left(\frac{k^2+g^2}{2ga} - \frac{l(l+1)}{z^2} - \tfrac{1}{4}z^2\right)\psi_0 = 0. \quad (3.5)$$

Setting

$$\psi_0(z) = z^{l+1}e^{-\frac{1}{4}z^2}X_0(z) \quad \text{and} \quad S = \tfrac{1}{2}z^2, \quad (3.6)$$

one finds that $X_0(z)$ satisfies the equation

$$S\frac{d^2X_0}{dS^2} + (b - S)\frac{dX_0}{dS} - aX_0 = 0, \quad (3.7)$$

where

$$a = \tfrac{1}{2}(l + \tfrac{3}{2}) - (k^2 + g^2)/4ga \quad \text{and} \quad b = l + \tfrac{3}{2}. \quad (3.8)$$

Equation (3.7) has the solution

$$X_0(z) = \Phi(a, b; z), \quad (3.9)$$

where Φ is a confluent hypergeometric function. The solution

$$\psi_0(z) = z^{l+1}e^{-\frac{1}{4}z^2}\Phi(a, b; \tfrac{1}{2}z^2) \quad (3.10)$$

is a normalizable function if

$$a = -n \quad \text{for} \quad n = 0, 1, 2, \cdots. \quad (3.11)$$

Setting $q = 4n + 3$ implies $k^2 + g^2 = ga(2l + q)$. Therefore, in the general case we may set

$$k^2 + g^2 = ga(2l + q) - 2a^2\Delta, \quad (3.12)$$

where Δ is of order zero in $1/g$. [Note: Only for $k^2 < 0$ is $q = 4n + 3$; for $k^2 > 0$, we consider q as a function of k: $q = q(k)$.]

B. Solution of the Radial Schrödinger Equation

We now substitute (3.12) into (3.4) and multiply the equation by one-half; then

$$\mathfrak{D}_q\psi = \left(-\tfrac{1}{2}\Delta h + \frac{1}{4}\sum_{i=2}^{\infty}\frac{h^{i-1}}{i!}(\tfrac{1}{2}z^2)^i\right)\psi, \quad (3.13)$$

where

$$\mathfrak{D}_q \equiv \frac{1}{2}\left(\frac{d^2}{dz^2} + l + \tfrac{1}{2}q - \frac{l(l+1)}{z^2} - \tfrac{1}{4}z^2\right) \quad (3.14)$$

and

$$h = -a/g. \quad (3.15)$$

As a first approximation to ψ we have

$$\psi \simeq \psi^{(1)} = \psi_q(z) = z^{l+1}e^{-\frac{1}{4}z^2}\Phi(a, b; \tfrac{1}{2}z^2). \quad (3.16)$$

Clearly, the first approximation leaves uncompensated terms on the right-hand side of (3.13) amounting to

$$R^{(1)} = \left(-\tfrac{1}{2}\Delta h + \frac{1}{4}\sum_{i=2}^{\infty}\frac{h^{i-1}}{i!}(\tfrac{1}{2}z^2)^i\right)\psi_q(z). \quad (3.17)$$

For convenience, we now write

$$\psi_q(z) \equiv \psi(a, b; z) \equiv \psi(a). \quad (3.18)$$

The recurrence relation for the functions $\psi(a)$ follows from that for confluent hypergeometric functions:

$$\tfrac{1}{2}z^2\psi(a) = (a, a+1)\psi(a+1) + (a, a)\psi(a) + (a, a-1)\psi(a-1), \quad (3.19)$$

where

$$\begin{aligned}(a, a+1) &= a = -\tfrac{1}{4}(q-3),\\ (a, a) &= b - 2a = l + \tfrac{1}{2}q,\\ (a, a-1) &= a - b = -\tfrac{1}{4}(q+3) - l.\end{aligned} \quad (3.20)$$

In general,

$$(\tfrac{1}{2}z^2)^m\psi(a) = \sum_{j=-m}^{m} S_m(a, a+j)\psi(a+m). \quad (3.21)$$

The coefficients $S_m(a, a+r)$ satisfy the following recurrence relation:

$$\begin{aligned}S_m(a, a+r) = {} & S_{m-1}(a, a+r-1)(a+r-1, a+r)\\ & + S_{m-1}(a, a+r)(a+r, a+r)\\ & + S_{m-1}(a, a+r+1)(a+r+1, a+r)\end{aligned} \quad (3.22)$$

with the boundary conditions

$$S_0(a, a) = 1; \quad S_0(a, a+i) = 0 \quad \text{for} \quad i \neq 0;$$
$$S_m(a, a+r) = 0 \quad \text{for} \quad |r| > m.$$

The expression $R^{(1)}$ may then be rewritten as

$$R^{(1)} = \sum_{i=0}^{\infty} h^{i+1}\sum_{j=-(i+2)}^{i+2}[a, a+j]_{i+1}\psi(a+j), \quad (3.23)$$

where

$$[a, a]_1 = -\tfrac{1}{2}\Delta + \frac{1}{4 \cdot 2!} S_2(a, a)$$

and, for i and j not zero simultaneously,

$$[a, a + j]_{i+1} = \frac{1}{4(i + 2)!} S_{i+2}(a, a + j). \quad (3.24)$$

Next we observe that

$$\mathfrak{D}_q \psi(a + j) = j\psi(a + j). \quad (3.25)$$

Thus a term $\mu\psi(a + j)$ on the right-hand side of (3.13) or in (3.23) may be canceled out by adding to the previous approximation the contribution $(\mu/j)\psi(a + j)$ except, of course, when $j = 0$. Hence the next contribution to $\psi^{(1)}$ becomes

$$\psi^{(2)} = \sum_{i=0}^{\infty} h^{i+1} \sum_{\substack{j=-(i+2) \\ j \neq 0}}^{i+2} \frac{[a, a + j]_{i+1}}{j} \psi(a + j). \quad (3.26)$$

This contribution leaves uncompensated a sum of terms $R^{(2)}$, and these again lead to another contribution $\psi^{(3)}$. This process may be repeated indefinitely. Then, adding successive contributions and rearranging these in powers of h, we obtain

$$\psi = \psi^{(1)} + \psi^{(2)} + \psi^{(3)} + \cdots$$

and

$$\psi(z) = \psi(a, b; z) + \sum_{i=1}^{\infty} h^i \sum_{\substack{j=-2i \\ j \neq 0}}^{2i} P_i(a, j)\psi(a + j, b; z). \quad (3.27)$$

The coefficients P_i in this expansion are found to be

$$P_1(a, \pm 2) = \frac{[a, a \pm 2]_1}{\pm 2}, \quad P_1(a, \pm 1) = \frac{[a, a \pm 1]_1}{\pm 1}, \quad (3.28)$$

etc. Others follow similarly.

For (3.27) to be a solution of (3.13), the sum of all terms in $\psi(a)$—left uncompensated so far (e.g., in $R^{(1)}$)—must vanish. Thus

$$0 = h[a, a]_1 + h^2\Big\{[a, a]_2 + \frac{[a, a + 2]_1}{2}[a + 2, a]_1 + \frac{[a, a + 1]_1}{1}[a + 1, a]_1 + \frac{[a, a - 1]_1}{-1}[a - 1, a]_1 + \frac{[a, a - 2]_1}{-2}[a - 2, a]_1\Big\} + \cdots. \quad (3.29)$$

This is the equation from which Δ and hence the eigenenergy k^2 is determined. Evaluation yields the expansion [calculated *independently* of (2.25)!]

$$\begin{aligned} k^2 + g^2 &= ga(2l + q) \\ &\quad - \frac{a^2}{2^4}[3(q^2 + 1) + 4(3q - 1)l + 8l^2] \\ &\quad - \frac{a^3}{3 \cdot 2^8 g}[q(11q^2 + 1) + 2(33q^2 - 6q + 1)l \\ &\quad + 24(5q - 1)l^2 + 64l^3] \\ &\quad - \frac{a^4}{3 \cdot 2^{15} g^2}[4(85q^4 + 2q^2 - 423) \\ &\quad + l(2720q^3 - 71q^2 + 32q + 2976) \\ &\quad + 32l^2(252q^2 - 12q + 64) \\ &\quad + 256l^3(41q - 9) + 4096l^4] + O(1/g^3). \end{aligned} \quad (3.30)$$

For small values of $|h|$ or large values of $|g|$, (3.30) is a proper asymptotic expansion with rapidly decreasing terms.

To ensure that the solution (3.29) is properly defined, we obviously require $|h| < 1$ and $|\frac{1}{2}z^2| < 1$, the latter expressing the convergence of the Kummer expansion of the confluent hypergeometric function inside the unit circle of its argument. Our solution (3.27) is therefore valid for

$$|a/g| < 1 \quad \text{and} \quad |r^2| < |1/ga|. \quad (3.31)$$

4. SOLUTIONS FOR r NEAR INFINITY

A. Method of Solution

In order to obtain the solutions for large r, we next derive the Jost solutions which behave like $e^{\pm ikr}$ for $r \to \infty$. So we define a solution $\psi^{(+)}$ of the radial wave equation by the asymptotic condition

$$k^2 > 0: \psi^{(+)}(l, k; r) \simeq e^{ikr} \quad \text{for} \quad |r| \to \infty. \quad (4.1)$$

Again we start from the equation [considering q, Δ as functions of k defined by (3.12)]

$$\psi'' + \{-[l(l + 1)/r^2] + ga(2l + q) - 2a^2\Delta - g^2 + g^2e^{-a^2r^2}\}\psi = 0. \quad (4.2)$$

In the limit of large coupling constants $g^2 \to \infty$, this equation is approximated by

$$\psi'' + \{-g^2 + g^2e^{-a^2r^2}\}\psi = 0. \quad (4.3)$$

An *approximate* solution of this equation is

$$\psi \sim \exp\left(\pm \int^r (g^2 - g^2e^{-a^2r^2})^{\frac{1}{2}}\, dr\right).$$

The terms in g^2 in (4.2) may therefore be removed by setting (no matter how q behaves!)

$$\psi = A(r) \exp\left(\pm g \int^r (1 - e^{-a^2r^2})^{\frac{1}{2}}\, dr\right). \quad (4.4)$$

The resulting equation for A may be written

$$\pm 2(1-e^{-a^2r^2})^{\frac{1}{2}}\frac{dA}{dr}+\left(a(2l+q)\pm\frac{a^2re^{-a^2r^2}}{(1-e^{-a^2r^2})^{\frac{1}{2}}}\right)A$$
$$=\frac{1}{g}\left(-A''+\frac{l(l+1)}{r^2}A+2a^2\Delta A\right). \quad (4.5)$$

It is now convenient to define functions $N(r)$ and $M(r)$ by

$$N(r)\equiv \ln M(r)=\int^{ar}\frac{dx}{(1-e^{-x^2})^{\frac{1}{2}}}. \quad (4.6)$$

Further, defining $A_q(r)$ as the solution of the first-order differential equation

$$\pm 2(1-e^{-a^2r^2})^{\frac{1}{2}}\frac{dA_q}{dr}$$
$$+\left(a(2l+q)\pm\frac{a^2re^{-a^2r^2}}{(1-e^{-a^2r^2})^{\frac{1}{2}}}\right)A_q=0 \quad (4.7)$$

and integrating, we find

$$A_q(r)=(1-e^{-a^2r^2})^{-\frac{1}{4}}[M(r)]^{\mp(l+\frac{1}{2}q)}. \quad (4.8)$$

Since the right-hand side of (4.5) is of order $1/g$, the function A_q represents the first approximation of A for large g in an as yet unlimited region of r provided $r\neq 0$.

Replacing A on the right of (4.5) by A_q, we obtain

$$R^{(1)}=[A_q(r)/4g(1-e^{-a^2r^2})^2]$$
$$\times\{-a^2(a^2r^2+2)e^{-2a^2r^2}+2a^2(1-2a^2r^2)e^{-a^2r^2}$$
$$-a^2(2l+q)^2(1-e^{-a^2r^2})\mp 4a^3(2l+q)r$$
$$\times e^{-a^2r^2}(1-e^{-a^2r^2})^{\frac{1}{2}}+4(1-e^{-a^2r^2})^2$$
$$\times[2a\Delta+l(l+1)/r^2]\}. \quad (4.9)$$

Our aim is to rewrite this expression in the form

$$R^{(1)}=\frac{1}{g}\sum_i d_iA_{q+4i}(r), \quad (4.10)$$

where the coefficients d_i are independent of g and r, and depend only on 1, Δ, and q.

Taking the above solution with the approximation A_q for A (for large g), we see that it has the following asymptotic behavior:

$$\psi\simeq e^{\pm gr}\cdot e^{\mp(l+\frac{1}{2}q)ar},\quad r\to\infty. \quad (4.11)$$

But from expansion (3.30), we know that in the same limit [for $k^2>0$, q is the solution of (3.30); for $k^2<0$, it is $4n+3$]

$$k^2\simeq -g^2[1-a(2l+q)/g]$$

or

$$\mp ik\simeq\pm g[1-a(2l+q)/2g]. \quad (4.12)$$

The solution (4.11) therefore has the behavior

$$\psi\simeq e^{\mp ikr},\quad r\to\infty. \quad (4.13)$$

The solution corresponding to the lower sign is thus the required Jost solution $\psi^{(+)}$ [provided, of course, the square root of k^2 is chosen as in (4.12)]. Here we considered $k^2>0$ only to fix the choice of signs in (4.4).

B. Some Properties of the Functions $M(r)$ and $A_q(r)$

To proceed further with the solution of the radial wave equation, we next examine some properties of the functions $M(r)$ and $A_q(r)$.

We have found no simple closed expression for the integral (4.6). It is therefore necessary to expand the integrand (in some region of r) and to integrate term by term. An expansion valid for r near infinity is not of immediate interest. Knowing the asymptotic behavior (4.13) of our solutions, our next step is to continue analytically the Jost solution to the region $r\simeq 0$. This means that we have to find the analytic continuation of the solution of this section to that in Sec. 3. The region of common validity of both types of solutions is, therefore, near r small but nonzero.

We observe first some important properties of the function A_q. By (4.8) we have

$$\frac{A_{q+4m}}{A_q}=[M(r)]^{\mp 2m}=\left(\frac{A_{q+4}}{A_q}\right)^m \quad (4.14)$$

and

$$\frac{A_{q+4m}}{A_q}=\frac{A_q}{A_{q-4m}},\quad \frac{A_{q+4m}}{A_q}\frac{A_{q+4n}}{A_q}=\frac{A_{q+4m+4n}}{A_q}. \quad (4.15)$$

Next we integrate (4.6) in the region of small r. We have

$$N(r)=\int^{ar}\frac{dx}{(1-e^{-x^2})^{\frac{1}{2}}}=\int^{ar}dx\,\frac{1}{x}\sum_{j=0}^{\infty}E_j(-x^2)^j, \quad (4.16)$$

where

$$E_0=1,\quad E_1=-\tfrac{1}{4},\quad E_2=-\tfrac{7}{90},$$
$$E_3=-\tfrac{1}{140},\quad E_4=\tfrac{7}{3240},\quad \text{etc.} \quad (4.17)$$

Hence

$$\frac{A_{q+4}}{A_q}=[M(r)]^{\mp 2}=e^{\mp 2N(r)}$$
$$=(ar)^{\mp 2}\exp\left(\mp\sum_{j=1}^{\infty}E_j\frac{(-a^2r^2)^j}{j}\right). \quad (4.18)$$

We rewrite this expression as

$$\frac{A_{q+4}}{A_q}=(ar)^{\mp 2}\left[\sum_{j=0}^{\infty}F_j^{\mp}(-a^2r^2)^j\right], \quad (4.19)$$

where

$$F_0^{\mp} = 1, \quad F_1^{\mp} = \pm\tfrac{1}{4}, \quad F_2^{-} = \tfrac{101}{1440}, \quad F_2^{+} = -\tfrac{11}{1440},$$
$$F_3^{-} = \tfrac{593}{40320}, \quad F_3^{+} = \tfrac{191}{40320}, \quad \text{etc.} \tag{4.20}$$

We now restrict ourselves to the lower signs. Then, reversing (4.19) and using (4.14) and (4.15), we have

$$a^2r^2 = \sum_{i=1}^{\infty} e_i \frac{A_{q+4i}}{A_q}, \tag{4.21}$$

where

$$e_1 = 1, \quad e_2 = F_1, \quad e_3 = 2F_1^2 - F_2,$$
$$e_4 = F_3 + 5F_1^3 - 5F_1F_2, \quad \text{etc.} \tag{4.22}$$

The inverse of (4.21) is readily found to be

$$\frac{1}{a^2r^2} = \sum_{i=-1}^{\infty} f_i \frac{A_{q+4i}}{A_q}, \tag{4.23}$$

where

$$f_{-1} = 1, \quad f_0 = -e_2, \quad f_1 = e_2^2 - e_3,$$
$$f_2 = 2e_2e_3 - e_2^3 - e_4, \quad \text{etc.} \tag{4.24}$$

Next we expand $R^{(1)}$ around $r = 0$ in ascending powers of r. Then, by (4.21) and (4.23),

$$\begin{aligned} R^{(1)} = \frac{a^2}{4g}\{&\mp[8l + 4q \pm 3 \pm (2l + q)^2 \\ &\mp 4l(l + 1)]A_{q-4} + [(2l + q)(1 \pm 1) \\ &+ (-\tfrac{1}{2} \pm \tfrac{1}{4})(q + 2l)^2 \pm \tfrac{3}{4} \mp l(l + 1) + 8\Delta]A_q \\ &+ [-\tfrac{1}{12}(2l + q)^2 + \tfrac{1}{12} \pm \tfrac{1}{8}(2l + q) \\ &\mp f_1\langle 8l + 4q \pm 3 \pm (2l + q)^2 \\ &\mp 4l(l + 1)\rangle]A_{q+4} + \cdots\}. \end{aligned} \tag{4.25}$$

We rewrite this expression in the form

$$R^{(1)}(q) = \frac{a^2}{4g} \sum_{j=-1}^{\infty} (q, q + 4j)A_{q+4j}, \tag{4.26}$$

where the coefficients $(q, q + 4j)$ follow by comparison with (4.25).

Perhaps the reader will wonder if these expansions and re-expansions, together with their manipulations, have any region of validity. However, for small r (unequal to zero), each individual expansion is valid, as is readily checked. So the real problem arises with Eq. (4.26): Does this expansion converge? There are indications—derived mainly from experience with other applications of this perturbation approach[11]—that the series can even be summed. Here, of course, it is precisely the expansion (4.26) in which we are interested, so in the following we assume its validity outside a small region around $r = 0$.

[11] H. J. W. Müller, J. Reine Angew. Math. **212**, 26 (1963); Math. Nachr. **31**, 89 (1966); **32**, 49, 157 (1966).

C. Solutions

We return to Eq. (4.5), which we want to solve by a large-g perturbation method. We observe first that the first of these equations may be obtained from the second (and vice versa) by changing the sign of r. For convenience—and in view of the asymptotic condition (4.13)—we therefore consider only the equation with lower signs. Dividing this equation by $-a$, we have

$$\mathfrak{D}_q A = -(1/ag)\{-A'' + [l(l + 1)/r^2]A + 2a^2\Delta A\}, \tag{4.27}$$

where

$$\mathfrak{D}_q \equiv \frac{2}{a}(1 - e^{-a^2r^2})^{\frac{1}{2}} \frac{d}{dr} - \left(q + 2l - \frac{are^{-a^2r^2}}{(1 - e^{-a^2r^2})^{\frac{1}{2}}}\right). \tag{4.28}$$

Then, by (4.7),

$$\mathfrak{D}_q A_q(r) = 0, \quad \mathfrak{D}_q A_{q+4j}(r) = 4jA_{q+4j}(r). \tag{4.29}$$

Thus a term μA_{q+4j} on the right of (4.5) may be removed by adding to A_q a term $\mu A_{q+4j}/4j$ except, of course, for $j = 0$.

We have seen that the first approximation to A (for large g and outside a small region around $r = 0$) is

$$A^{(1)} = A_q, \tag{4.30}$$

and that this approximation leaves uncompensated on the right-hand side of (4.5) the terms of expression (4.9) and so of (4.26). Here the terms in A_{q+4j} for $j \neq 0$ may be removed by adding to $A^{(1)}$ the new contribution

$$A^{(2)} = \frac{a^2}{4g} \sum_{\substack{j=-1 \\ j\neq 0}}^{\infty} \frac{(q, q + 4j)}{4j} A_{q+4j}. \tag{4.31}$$

In its turn, this contribution leaves uncompensated on the right of (4.5) an expression obtained by replacing A_{q+4j} in (4.31) by

$$R^{(1)}(q + 4j) = \frac{a^2}{4g} \sum_{j'=-1}^{\infty} (q + 4j, q + 4j + 4j')A_{q+4j+4j'}. \tag{4.32}$$

Proceeding as before, we obtain a new contribution

$$A^{(3)} = \left(\frac{a^2}{4g}\right)^2 \sum_{\substack{j=-2 \\ j\neq 0}}^{\infty} P_3(q, q + 4j)A_{q+4j}, \tag{4.33}$$

where

$$P_3(q, q-8) = \frac{(q, q-4)}{-4}\frac{(q-4, q-8)}{-8},$$

$$P_3(q, q-4) = \frac{(q, q-4)}{-4}\frac{(q-4, q-4)}{-4},$$

$$P_3(q, q+4) = \frac{(q, q-4)}{-4}\frac{(q-4, q+4)}{4} + \frac{(q, q+4)}{4}\frac{(q+4, q+4)}{4} + \frac{(q, q+8)}{8}\frac{(q+8, q+4)}{4},$$

$$P_3(q, q+8) = \frac{(q, q-4)}{-4}\frac{(q-4, q+8)}{8} + \frac{(q, q+4)}{4}\frac{(q+4, q+8)}{8} + \frac{(q, q+8)}{8}\frac{(q+8, q+8)}{8} + \frac{(q, q+12)}{12}\frac{(q+12, q+8)}{8}, \tag{4.34}$$

and so on.

It is obvious that higher approximations follow in exactly the same manner. Then successive contributions $A^{(1)}$, $A^{(2)}$, $A^{(3)}, \cdots$ to A form a rapidly decreasing sequence provided that

$$|(a^2/4g)A_{q+4j}| \ll |A_q|, \tag{4.35}$$

for $j = -1, 1, 2, \cdots$. Thus, in the region of moderately small r,

$$|a^2/4g| \ll |A_q/A_{q+4j}| = (ar)^{-2j}, \tag{4.36}$$

so that for $j > 0$,

$$(ar)^{2j} \ll 4g/a^2, \tag{4.37}$$

and for $j = -1$,

$$(ar)^2 \gg a^2/4g. \tag{4.38}$$

The inequality (4.37) allows unrestricted values of r; (4.38), however, excludes a definite region around the origin. The solution therefore requires $r^2 > 1/|4g|$.

So far we have ignored completely the term A_q in expansion (4.26), as also in corresponding expressions $R^{(2)}$, etc., left uncompensated by contributions $A^{(2)}$, etc. Hence, for

$$A = A^{(1)} + A^{(2)} + A^{(3)} + \cdots \tag{4.39}$$

to satisfy (4.5), the sum of these terms must vanish. Thus

$$0 = \frac{a^2}{4g}(q, q) + \left(\frac{a^2}{4g}\right)^2 \times \left(\frac{(q, q-4)}{-4}(q-4, q) + \frac{(q, q+4)}{4}(q+4, q)\right) + \cdots. \tag{4.40}$$

This is the equation from which Δ and thereby the eigenvalue is determined. The complexity of the coefficients $(q + 4j, q + 4k)$ [cf. (4.25)] seems to indicate that the expansion for Δ is best obtained by the method of Sec. 3. However, it is important to note that the eigenvalue (for $k^2 < 0$) obtained from (4.40) is identical with that obtained previously. This is easily verified to first order. Setting $(q, q) = 0$, we have [cf. (4.25)]

$$32\Delta = 3(q^2 + 1) + 4(3q - 1)l + 8l^2,$$

and this expression substituted into (3.12) leads again to the expansion (3.30).

We have now two types of large-g solutions of the radial wave equation belonging to one and the same eigenvalue: one type in terms of parabolic cylinder or confluent hypergeometric functions, the other in terms of the integral $N(r)$—the former being valid in the region of small r, the latter in regions of moderate and large r. Thus the behavior of the solutions over all regions of the independent variable can be investigated.

5. ANALYTIC CONTINUATION

In the derivation of the S matrix, we require the value of the (analytically continued) Jost solution at the origin. This is the question we examine in this section.

We recall our large-g asymptotic expansion of the Jost solution:

$$\psi^{(+)}(l, k; r) = A(r) \exp\left(-g\int^r (1 - e^{-a^2r^2})^{\frac{1}{2}}\, dr\right), \tag{5.1}$$

where

$$A(r) = A_q(r) + \frac{a^2}{4g}\sum_{\substack{j=-1\\ j\neq 0}}^{\infty}\frac{(q, q+4j)}{4j}A_{q+4j} + \left(\frac{a^2}{4g}\right)^2\sum_{\substack{j=-1\\ j\neq 0}}^{\infty}\sum_{\substack{j''=-1\\ j''\neq -j}}^{\infty}\frac{(q, q+4j)}{4j} \times \frac{(q+4j, q+4j+4j'')}{4j+4j''}A_{q+4j+4j''} + \cdots \tag{5.2}$$

for $r^2 > 1/|4g|$.

In the neighborhood of the origin, we found the solution

$$\psi_0(z) = \psi_q(z) + \sum_{i=1}^{\infty} \left(-\frac{a}{g}\right)^i \sum_{\substack{j=-2i \\ j\neq 0}}^{2i} P_i(a,j)\psi_{q+4j}(z), \quad (5.3)$$

where $z^2 = 2gar^2$, this expansion being valid for $r^2 < 1/|ga|$.

We now want to show that these two solutions are proportional to each other in their common region of validity.

We start with the Jost solution. Expanding the square root in the integrand of the exponential in the region of small r, we obtain

$$\exp\left[-g\int^r (1-e^{-a^2r^2})^{\frac{1}{2}}\,dr\right]$$
$$= \exp(-\tfrac{1}{2}gar^2)\exp\left(\frac{ga^3r^4}{16} - \frac{5ga^5r^6}{576} + \cdots\right)$$
$$= \exp(-\tfrac{1}{2}gar^2)\left(1 + \frac{ga^3r^4}{16} - \frac{5ga^5r^6}{576} + \cdots\right). \quad (5.4)$$

Further, by (4.16),

$$N(r) = \ln ar + \sum_{j=1}^{\infty} E_j \frac{(-a^2r^2)^j}{2j}, \quad (5.5)$$

with the E_j given in (4.17). Hence

$$M(r) = ar\exp\left(\sum_{j=1}^{\infty} E_j \frac{(-a^2r^2)^j}{2j}\right)$$
$$= ar(1 + \tfrac{1}{8}a^2r^2 - \tfrac{67}{5760}a^4r^4 + \cdots). \quad (5.6)$$

By (4.8) we then have

$$A_q(r) = (ar)^{l+\frac{1}{2}(q-1)}\left\{1 + \frac{q+2l+2}{16}a^2r^2 + \frac{a^4r^4}{512}[(q+2l)^2 - \tfrac{44}{45}(q+2l) - \tfrac{4}{3}] + \cdots\right\}. \quad (5.7)$$

This expression may now be substituted in the solution (5.2). The coefficient of the dominant factor

$$\exp(-\tfrac{1}{2}agr^2)(ar)^{l+\frac{1}{2}(q-1)}$$

in the solution $\psi^{(+)}$ is seen to be

$$S = 1 + \frac{a^2}{4g}\frac{(q,q-4)}{-4}\frac{(q+2l-2)}{16} + O\left(\left(\frac{a}{4g}\right)^2\right)$$
$$= 1 + \frac{(q-3)(q+4l-1)(q+2l-2)}{64}\frac{a^2}{4g} + \cdots. \quad (5.8)$$

We now consider the solution (5.3). The function ψ_q was earlier [cf. (3.16)] defined as

$$\psi_q(z) \equiv \psi(a,b;z) = z^{l+1}e^{-\frac{1}{4}z^2}\Phi(a,b;\tfrac{1}{2}z^2), \quad (5.9)$$

where $z^2 = 2gar^2$. The confluent hypergeometric function on the right is related to the generalized Laguerre function L by

$$\Phi(-\nu,\alpha+1;x) = \Gamma(\nu+1)L_\nu^{(\alpha)}(x), \quad (5.10)$$

so that

$$\Phi(a,b;\tfrac{1}{2}z^2) = \Gamma(\tfrac{1}{4}(q+1))L_{\frac{1}{4}(q-3)}^{l+\frac{1}{2}}(\tfrac{1}{2}z^2). \quad (5.11)$$

For the Laguerre function we have the following asymptotic expansion:

$$L_{\frac{1}{4}(q-3)}^{l+\frac{1}{2}}(\tfrac{1}{2}z^2)$$
$$= \frac{(-\frac{1}{2}z^2)^{\frac{1}{4}(q-3)}}{\Gamma(\frac{1}{4}(q+1))}\left[1 - \frac{(q-3)(q+4l-1)}{1!}\frac{1}{2^3z^2} + \frac{(q-3)(q-7)(q+4l-1)(q+4l-5)}{2!} \times \frac{1}{2^6z^4} - \cdots\right]. \quad (5.12)$$

Substituting this expansion in the solution (5.3), the coefficient of the dominant factor $\exp(-\frac{1}{2}agr^2) \times (ar)^{l+\frac{1}{2}(q-1)}$ in the result is found to be

$$t = (-1)^{\frac{1}{4}(q-3)}2^{\frac{1}{2}(l+1)}(g/a)^{\frac{1}{2}l+\frac{1}{4}(q-1)} \times \{1 - (a/g)[P_1(a,2)(1/2!\,2^8)(q-3)(q-7) \times (q+4l-1)(q+4l-5) + P_1(a,1)(1/1!\,2^4)(q-3)(q+4l-1)] + \cdots\}$$
$$= (-1)^{\frac{1}{4}(q-3)}2^{\frac{1}{2}(l+1)}(g/a)^{\frac{1}{2}l+\frac{1}{4}(q-1)} \times \{1 - (1/2^{17})(a/g)[(q-3)^2(q-7)^2 \times (q+4l-1)(q+4l-5) - 2^8(q-3)^2(q+4l-1)(q+2l-2)] + \cdots\}. \quad (5.13)$$

It follows that in their common range of validity

$$\psi^{(+)} = \gamma\psi_0, \quad (5.14)$$

with

$$\gamma = \frac{S}{t} = \frac{(a/g)^{\frac{1}{2}l+\frac{1}{4}(q-1)}}{(-1)^{\frac{1}{4}(q-3)}2^{\frac{1}{2}(l+1)}} \times \left\{1 + \frac{a}{2^{17}g}[(q-3)^2(q-7)^2 \times (q+4l-1)(q+4l-5) - 2^8(q-3)^2(q+4l-1)(q+2l-2) + 2^9a(q-3)(q+4l-1)(q+2l-2)] + \cdots\right\}. \quad (5.15)$$

6. CONCLUSION

In the foregoing, we have investigated the radial Schrödinger equation for a Gauss potential with large or moderately large values of the coupling constant g^2. In particular, we obtained the binding energies for all angular momenta in the form of an asymptotic expansion in g. This expansion is also useful for smaller g when treated by the method of Padé approximants.

The scattering problem, however, requires further investigation.

The analogy of the present problem with that of the Mathieu equation

$$y'' + \{\lambda - 2h^2 \cos 2x\}y = 0 \qquad (6.1)$$

may be seen by comparing (3.30) with the large-h asymptotic expansion[12] of the eigenvalue λ in (6.1):

$$\lambda + 2h^2 = 2hq - \tfrac{1}{8}(q^2 + 1) + O(1/h), \qquad (6.2)$$

where q again corresponds to an odd integer. Also, the small-h expansion[3,9]

$$\lambda = \nu^2 + \frac{h^4}{2(\nu^2 - 1)} + \frac{(5\nu^2 + 7)h^8}{32(\nu^2 - 1)^3(\nu^2 - 4)} - \cdots \qquad (6.3)$$

[12] R. B. Dingle and H. J. W. Müller, J. Reine Angew. Math. **211**, 11 (1962).

(for ν an integer) is similar to the corresponding expansion for k^2 in the case of a Gauss potential[7]:

$$k^2 = -g^2 + g^2a^2 + \tfrac{2}{5}l(l + 1) - (\nu + \tfrac{1}{2})^2 - \frac{l(l + 1) - 15g^2a^2}{30(2\nu - 1)(2\nu + 3)}(4\nu^2 + \nu - 2l - 3) + \cdots. \qquad (6.4)$$

The corresponding behavior persists in the solutions.

ACKNOWLEDGMENTS

The author is indebted to Professors H. H. Aly and A. B. Zahlan of the American University of Beirut for a 1967 Visiting Assistant Professorship in their department during which this work was done.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Parastochastics*

U. FRISCH†
Institut d'Astrophysique, 98 bis boulevard Arago, Paris, France
AND
R. BOURRET
Behavioural Science Foundations, Simon Fraser University, Burnaby, B.C., Canada

(Received 15 February 1969)

The generalized commutation relations $A_iB_j - \lambda B_jA_i = \Gamma_{ij}I$ are introduced, where A_i and B_i are adjoint of each other, I is the identity, and Γ_{ij} is a real covariance. For $\lambda = +1$ (-1, or 0, respectively) the *parastochastic* function $A_i + B_i$ is given an interpretation in terms of Gaussian random functions (a two-valued stationary Markov process, or infinite symmetric random matrices of the type considered by Wigner in connection with energy levels of heavy nuclei, respectively). In the Fock-space realization, A_i and B_i appear as destruction and creation operators for bosons (fermions or boltzmannons, i.e., distinguishable particles, respectively). A few purely algebraic theorems are proved, which are applied to *linear stochastic equations* (equations with random coefficients). Existence and uniqueness being presupposed, mean Green's functions are shown to satisfy closed *master equations*. A linear functional differential master equation is obtained for equations with Gaussian coefficients. It is shown that the often-used first cumulant-discard closure assumption, which leads to a very simple master equation, is exact for differential equations with a two-valued stationary Markovian coefficient. A parastochastic reformulation of the theory of Kraichnan is given, and his nonlinear master equations are, for the first time, rigorously derived without any recourse to perturbation theory or to diagrams. Kraichnan's *random-coupling model* is obtained by replacing scalar stochastic quantities by Wigner matrices or, equivalently, bosons by boltzmannons (i.e., changing λ from $+1$ to 0). Finally, the nonlinear Kraichnan equation,

$$dy(t)/dt = -\int_0^t y(t - t')\Gamma(t - t')y(t')\,dt',$$

is reduced to a linear parastochastic equation in the Fock space; existence, uniqueness, boundedness, and asymptotic behavior are obtained.

1. INTRODUCTION

Let A and B be boson destruction and creation operators in a Hilbert space $\mathcal{H}$, adjoint of each other and satisfying

$$AB - BA = \sigma^2 I, \quad \text{and} \quad A\Psi_0 = 0, \qquad (1.1)$$

where I is the identity operator and Ψ_0 is a unit vector (the "vacuum state" of quantum-field theory). It has been noticed[1,2] that the vacuum expectations

$$(\Psi_0, (A + B)^n\Psi_0)$$

* The research reported in this paper was sponsored by the Air Force Cambridge Research Laboratories, Office of Aerospace Research under Contract No. AF19(628)3868.

† Part of this work was done while at the Courant Institute of Mathematical Sciences, New York.

[1] K. Furutsu, "On The Statistical Theory of Electromagnetic Waves in a Fluctuating Medium," *Natl. Bur. Std. Monograph* Vol. 79: (National Bureau of Standards, Washington, D.C., 1964).

[2] I. E. Segal, *Lectures in Applied Mathematics*, Vol. 2: "Mathematical Problems of Relativistic Physics" (Am. Math. Soc., 1963).

The scattering problem, however, requires further investigation.

The analogy of the present problem with that of the Mathieu equation

$$y'' + \{\lambda - 2h^2 \cos 2x\}y = 0 \qquad (6.1)$$

may be seen by comparing (3.30) with the large-h asymptotic expansion[12] of the eigenvalue λ in (6.1):

$$\lambda + 2h^2 = 2hq - \tfrac{1}{8}(q^2 + 1) + O(1/h), \qquad (6.2)$$

where q again corresponds to an odd integer. Also, the small-h expansion[3,9]

$$\lambda = \nu^2 + \frac{h^4}{2(\nu^2 - 1)} + \frac{(5\nu^2 + 7)h^8}{32(\nu^2 - 1)^3(\nu^2 - 4)} - \cdots \qquad (6.3)$$

[12] R. B. Dingle and H. J. W. Müller, J. Reine Angew. Math. **211**, 11 (1962).

(for ν an integer) is similar to the corresponding expansion for k^2 in the case of a Gauss potential[7]:

$$k^2 = -g^2 + g^2a^2 + \tfrac{2}{5}l(l + 1) - (\nu + \tfrac{1}{2})^2 - \frac{l(l + 1) - 15g^2a^2}{30(2\nu - 1)(2\nu + 3)}(4\nu^2 + \nu - 2l - 3) + \cdots. \qquad (6.4)$$

The corresponding behavior persists in the solutions.

ACKNOWLEDGMENTS

The author is indebted to Professors H. H. Aly and A. B. Zahlan of the American University of Beirut for a 1967 Visiting Assistant Professorship in their department during which this work was done.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Parastochastics*

U. FRISCH†
Institut d'Astrophysique, 98 bis boulevard Arago, Paris, France
AND
R. BOURRET
Behavioural Science Foundations, Simon Fraser University, Burnaby, B.C., Canada

(Received 15 February 1969)

The generalized commutation relations $A_iB_j - \lambda B_jA_i = \Gamma_{ij}I$ are introduced, where A_i and B_i are adjoint of each other, I is the identity, and Γ_{ij} is a real covariance. For $\lambda = +1$ (-1, or 0, respectively) the *parastochastic* function $A_i + B_i$ is given an interpretation in terms of Gaussian random functions (a two-valued stationary Markov process, or infinite symmetric random matrices of the type considered by Wigner in connection with energy levels of heavy nuclei, respectively). In the Fock-space realization, A_i and B_i appear as destruction and creation operators for bosons (fermions or boltzmannons, i.e., distinguishable particles, respectively). A few purely algebraic theorems are proved, which are applied to *linear stochastic equations* (equations with random coefficients). Existence and uniqueness being presupposed, mean Green's functions are shown to satisfy closed *master equations*. A linear functional differential master equation is obtained for equations with Gaussian coefficients. It is shown that the often-used first cumulant-discard closure assumption, which leads to a very simple master equation, is exact for differential equations with a two-valued stationary Markovian coefficient. A parastochastic reformulation of the theory of Kraichnan is given, and his nonlinear master equations are, for the first time, rigorously derived without any recourse to perturbation theory or to diagrams. Kraichnan's *random-coupling model* is obtained by replacing scalar stochastic quantities by Wigner matrices or, equivalently, bosons by boltzmannons (i.e., changing λ from $+1$ to 0). Finally, the nonlinear Kraichnan equation,

$$dy(t)/dt = -\int_0^t y(t - t')\Gamma(t - t')y(t')\,dt',$$

is reduced to a linear parastochastic equation in the Fock space; existence, uniqueness, boundedness, and asymptotic behavior are obtained.

1. INTRODUCTION

Let A and B be boson destruction and creation operators in a Hilbert space $\mathcal{H}$, adjoint of each other and satisfying

$$AB - BA = \sigma^2 I, \quad \text{and} \quad A\Psi_0 = 0, \qquad (1.1)$$

where I is the identity operator and Ψ_0 is a unit vector (the "vacuum state" of quantum-field theory). It has been noticed[1,2] that the vacuum expectations

$$(\Psi_0, (A + B)^n\Psi_0)$$

* The research reported in this paper was sponsored by the Air Force Cambridge Research Laboratories, Office of Aerospace Research under Contract No. AF19(628)3868.

† Part of this work was done while at the Courant Institute of Mathematical Sciences, New York.

[1] K. Furutsu, "On The Statistical Theory of Electromagnetic Waves in a Fluctuating Medium," *Natl. Bur. Std. Monograph* Vol. 79: (National Bureau of Standards, Washington, D.C., 1964).

[2] I. E. Segal, *Lectures in Applied Mathematics*, Vol. 2: "Mathematical Problems of Relativistic Physics" (Am. Math. Soc., 1963).

are equal to the corresponding moments of a Gaussian random variable $m(\omega)$ with zero mean value and variance σ^2. This follows from the formula

$$e^{(A+B)} = e^B e^A e^{\frac{1}{2}(AB-BA)}, \tag{1.2}$$

valid whenever the commutator $AB - BA$ is a scalar. This identity, applied to the calculation of the vacuum expectation of $e^{iz(A+B)}$, yields

$$(\Psi_0, e^{iz(A+B)}\Psi_0) = e^{-\frac{1}{2}\sigma^2 z^2}, \tag{1.3}$$

which is precisely the characteristic function of $m(\omega)$.

The purpose of the first part of this paper (Sec. 2) is to extend this result to the time-dependent generalized commutation relations introduced by Bourret [3,4]:

$$A(t)B(t') - \lambda B(t')A(t) = \Gamma(t,t')I, \tag{1.4}$$

where $\Gamma(t,t')$ is a real covariance and λ a scalar ($\lambda \geq -1$). Preliminary results on this subject have already been reported.[3-5] The present treatment will be self-contained.

Other generalizations of the standard quantum-mechanical commutation relations have been proposed, leading to *parastatistics*.[6] These are chosen so as to preserve the Heisenberg equations of motion for a free field; the present generalization forfeits this property. However, our purpose is not to give an extension of quantum mechanics, but rather to investigate the probabilistic content of (1.4).

The *parastochastic function* $M(t) = A(t) + B(t)$, which is a one-parameter family of self-adjoint operators in a Hilbert space, will be shown to behave in certain respects like a random function. Moments of $M(t)$, i.e., vacuum expectations of products, are given by a formula that appears as a generalization of a well-known formula for Gaussian random functions. Except for $\lambda = +1$, moments of $M(t)$ do not coincide with moments of a scalar-valued random function. It will for example be shown that for $\lambda = 0$, the parastochastic function is related to a time-dependent generalization of the infinite symmetric random matrices with independent entries introduced by Wigner.[7]

The second part of this paper (Sec. 3) is concerned with the calculation of mean Green's functions for linear stochastic equations, such as arise in the study of turbulent convection, wave propagation in random media, impurity scattering, stochastic acceleration, random networks, etc. It is known that this calculation involves a difficult closure problem (see Ref. 8 and Sec. 3A). The problem was usually tackled by means of all-order perturbation expansions, together with diagrammatic representations of moments and formal resummations of infinite classes of diagrams selected by dimensional arguments.[5,9-11] Although such methods were inherited from quantum-field theory, the analogies were rather superficial (see, however, Refs. 1 and 12). Our approach is basically different: What might be called a second quantization of stochastic equations is performed from the beginning by changing stochastic (random) quantities into parastochastic ones. From the resulting *parastochastic equations*, closed master equations for the mean Green's functions may be derived in a purely algebraic way, without recourse to perturbation theory.

In a celebrated paper on linear and nonlinear stochastic equations, Kraichnan[13] has shown that, given a linear stochastic equation, a closed nonlinear master equation may be obtained for the mean Green's function of a judiciously modified version of the original problem, called the random-coupling model. This equation has proved to give reasonable predictions for the original problem and to be well-suited for numerical integration (Such questions are discussed in Secs. 3C, D). Unfortunately, Kraichnan's derivation, at least part of it, was based on the above-mentioned perturbation expansions and diagrams. This has somewhat obscured the fact that his method does not involve any approximation. It will be shown here that Kraichnan's random-coupling model corresponds to replacing the random coefficients of the original equation by Wigner matrices or, assuming that the original coefficients were Gaussian, to changing λ from $+1$ to 0 in the parastochastic formulation of the problem. As a consequence, it will be possible both to give a rigorous derivation of the nonlinear Kraichnan equation and to prove for it a few general nontrivial results, such as existence in the large.

Notations: Ordinary stochastic (random) quantities are denoted by lower case letters, parastochastic quantities are denoted by capitals, and Wigner matrices by script capitals. A boldface letter indicates that the corresponding quantity is, at the same time, an integral operator in an accessory function space.

[3] R. Bourret, Phys. Letters **12**, 323 (1964).
[4] R. Bourret, Can. J. Phys. **44**, 2519 (1966).
[5] U. Frisch, in *Probabilistic Methods in Applied Mathematics*, A. T. Bharucha-Reid, Ed. (Academic Press, Inc., New York, 1968).
[6] M. Dresden, *Brandesis Lectures in Theoretical Physics* (W. A. Benjamin, Inc., New York, 1963), Vol. 2, p. 377.
[7] E. P. Wigner, Ann. Math. **62**, 548 (1955).
[8] R. H. Kraichnan, Proc. Symp. Appl. Math. **13**, 199 (1962).
[9] R. Bourret, Nuovo Cimento **26**, 1 (1962).
[10] K. Furutsu, J. Res. Natl. Bur. Std. **67D**, 303 (1963).
[11] V. I. Tatarski, Zh. Eksp. Teor. Fiz. **46**, 1399 (1964) [Sov. Phys.—JETP, **19**, 946 (1964)].
[12] K. Furutsu, J. Radio Res. Lab. (Tokyo) **14**, 99 (1967).
[13] R. H. Kraichnan, J. Math. Phys. **2**, 124 (1961).

2. GENERALIZED COMMUTATION RELATIONS

A. Formulation

Let Ψ_0 be a unit vector in a complex Hilbert space H; let λ be a real number ($\lambda \geq -1$); let $\Gamma(t, t')$ be a real symmetric kernel of positive type (see definition below), defined on $T \times T$, where T is a nonempty set whose generic element t is called the time; T will usually be a real finite-dimensional vector space; let $A(t)$ and $B(t)$ be time-dependent linear operators in H, bounded or unbounded, which for any t, t' in T satisfy the following conditions:

$$A(t)B(t') - \lambda B(t')A(t) = \Gamma(t, t')I, \tag{2.1}$$

$$A(t) = B^*(t), \tag{2.2}$$

$$A(t)\Psi_0 = 0, \tag{2.3}$$

where I denotes the identity operator in H, and B^* the adjoint of B. Equation (21) will be called a "generalized commutation relation" for the "destruction" and "creation" operators $A(t)$ and $B(t)$. Notice that for $\Gamma(t, t') = \delta_{t,t'}$ and $\lambda = +1$ ($\lambda = -1$, respectively) Eq. (2.1) reduces to the standard commutation relation for bosons (fermions, respectively).[14]

We recall that the real symmetric function $\Gamma(t, t')$ is said to be of "positive type" (or sometimes positive-definite) if for any choice of $t_1, \cdots, T_n$ in T and of any real numbers $\alpha_1, \cdots, \alpha_n$,

$$\sum_{i,j=1}^{n} \alpha_i \alpha_j \Gamma(t_i, t_j) \geq 0. \tag{2.4}$$

An equivalent condition is the existence of a real random function [stochastic process] $m(\omega; t)$ such that[15,16]

$$\Gamma(t, t') = E\{m(\omega; t)m(\omega; t')\}, \tag{2.5}$$

where ω is an element in a probability space Ω, and $E\{\ \}$ denotes the probabilistic expectation value. If $T = \mathbf{R}$ and $\Gamma(t, t')$ is a function only of the difference of its arguments,

$$\Gamma(t, t') = \Gamma(t - t'), \tag{2.6}$$

then Γ is said to be "stationary."

In this paper, we make use of the following:

Bochner's Theorem[15]: A function is a real stationary covariance if and only if it is the Fourier transform of a positive finite even measure.

[14] D. Kastler, *Introduction à l'electrodynamique quantique* (Dunod Cie., Paris, 1961).

[15] J. L. Doob, *Stochastic Processes* (John Wiley & Sons, Inc., New York, 1953).

[16] P. Lévy, *Processus stochastiques et mouvement brownien* (Gauthier-Villars, Paris, 1965), 2nd ed.

Notations and definitions: The one-parameter family of self-adjoint operators

$$M(t) = A(t) + B(t) \tag{2.7}$$

is called a "parastochastic function"; $\Gamma(t, t')$ is called the "covariance" of $M(t)$. The inner product in H is denoted by $(\cdot, \cdot)$. According to the quantum-field theoretic terminology, the unit vector Ψ_0 is called the "vacuum state." For any linear operator L and H,

$$E\{L\} = (\Psi_0, L\Psi_0) \tag{2.8}$$

is called the "vacuum expectation" of L. Vacuum expectations such as

$$E\{M(t)\},\quad E\{M(t_1)M(t_2)\},\quad E\{M(t_1)M(t_2)M(t_3)\}, \cdots$$

are called "moments" of the parastochastic function $M(t)$.

A few problems that arise in connection with the generalized commutation relations are now stated.

Realization problem: Is it possible to find a Hilbert space H, a vacuum state Ψ_0, and operators $A(t)$ and $B(t)$ satisfying Eqs. (2.1)–(2.3)? See Sec. 2D.

Uniqueness problem: If it is assumed that the generalized commutation relations are realizable, are the moments of $M(t)$ uniquely defined in terms of λ and Γ? See Sec. 2C.

Random representation problem: From the Introduction it is known that, for fixed t and $\lambda = +1$, $\{E\{M^n(t)\}\}_{n\geq 0}$ is the sequence of moments of a Gaussian random variable with zero mean value and variance $\Gamma(t, t)$. More generally, it may be asked: Is there a real random function $m(\omega, t)$ having the same moments as the parastochastic function $M(t)$, i.e.,

$$E\{M(t_1) \cdots M(t_n)\} = E\{m(\omega; t_1) \cdots m(\omega; t_n)\}? \tag{2.9}$$

The rhs of (2.9) is a symmetric function of its arguments; the lhs, however, is usually not, for the $M(t)$'s need not commute. Hence, it may be necessary to look for more general (nonscalar-valued) random representations. See Sec. 2E.

B. The Time-Independent Case

If the set T has a single element, Eqs. (2.1)–(2.3) read

$$AB - \lambda BA = \sigma^2 I, \tag{2.10}$$

$$A = B^*, \tag{2.11}$$

$$A\Psi_0 = 0. \tag{2.12}$$

It may as well be assumed that $\sigma^2 = 1$. (If not, change A and B into $\sigma^{-1}A$ and $\sigma^{-1}B$.) An explicit realization will be found in the Hilbert space l_2 of real sequences $x = \{x_n\}_{n\geq 0}$ such that

$$\|x\|^2 = \sum_{n\geq 0} x_n^2 < +\infty.$$

Let

$$\Psi_0 = (1, 0, 0, \cdots), \tag{2.13}$$

and let us look for a realization of the following form:

$$A: (x_0, x_1, \cdots) \to (f_0 x_1, f_1 x_2, \cdots),$$
$$B: (x_0, x_1, \cdots) \to (0, f_0 x_0, \quad f_1 x_1, \cdots), \tag{2.14}$$

where the sequence f_n is to be determined. Equations (2.11) and (2.12) are satisfied; inserting the proposed operators into (2.10), we obtain

$$f_0^2 = 1 \quad \text{and} \quad f_n^2 - \lambda f_{n-1}^2 = 1, \quad n \geq 1. \tag{2.15}$$

The unique solution of this difference equation is

$$f_n^2 = 1 + \lambda + \cdots + \lambda^n = (1 - \lambda^{n+1})/(1 - \lambda), \quad \lambda \neq +1,$$
$$= n + 1, \quad \lambda = +1. \tag{2.16}$$

For $\lambda = 0$, the operators A and B are simply left- and right-shift operators. Notice also that A, B, and therefore the parastochastic operator $M = A + B$ are bounded for $-1 \leq \lambda \leq +1$ and unbounded for $\lambda \geq 1$.

Anticipating uniqueness of moments, which will be proved in the more general time-dependent frame of the next section, we obtain a random representation of M as follows: M, being self-adjoint, has a spectral decomposition,[17]

$$M = \int_{-\infty}^{+\infty} m\, dE(m), \tag{2.17}$$

such that

$$M^n = \int_{-\infty}^{+\infty} m^n\, dE(m), \tag{2.18}$$

where $dE(m)$ is a projector-valued measure on the real line. Taking the vacuum expectation of both members of (2.18), we obtain

$$E\{M^n\} = \int_{-\infty}^{+\infty} m^n (\Psi_0, dE(m)\Psi_0) = \int_{-\infty}^{+\infty} m^n\, dP(m), \tag{2.19}$$

where $dP(m)$ is now an ordinary probability measure on the real line. Any random variable $m(\omega)$ with the distribution $dP(m)$ is therefore a random representation of M. Moreover, the distribution $dP(m)$ will be obtained explicitly for $-1 \leq \lambda \leq +1$ in terms of a continued fraction which can be evaluated

[17] F. Riesz and B. Sz.-Nagy, *Leçons d'analyse fonctionnelle* (Akadémiai Kiado, Budapest, 1953).

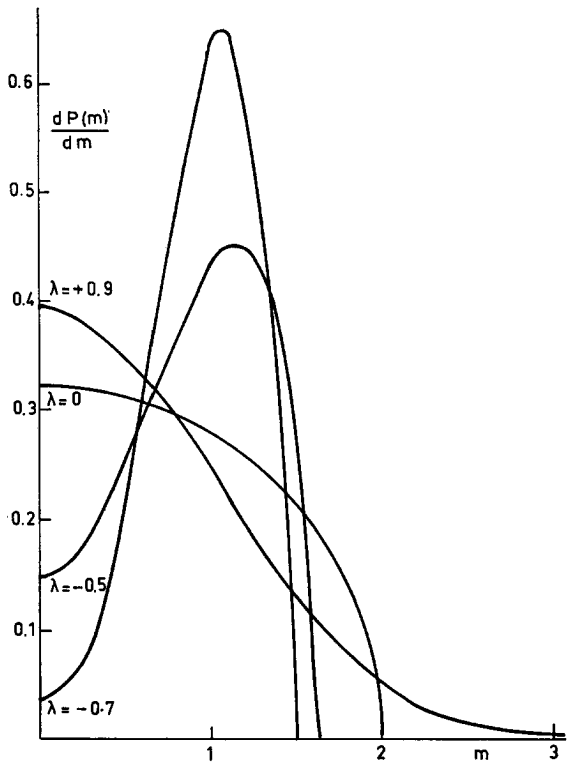


FIG. 1. Probability density $dP_\lambda(m)/dm$ for the time-independent generalized commutation relation (2.10) for various values of λ.

numerically (see Appendix). The results are as follows: The probability distribution $dP(m)$ is symmetric, has compact support in $[-2(1-\lambda)^{-\frac{1}{2}}, +2(1-\lambda)^{-\frac{1}{2}}]$ for $-1 \leq \lambda < 1$, and extends to infinity for $\lambda \geq 1$ (this could have been anticipated from the corresponding boundedness and unboundedness of M). For $\lambda = -1, 0$, and $+1$, the distribution $dP(m)$ is known:

$$\text{For } \lambda = -1, \quad dP(m) = \tfrac{1}{2}\delta(m-1) + \tfrac{1}{2}\delta(m+1) \tag{2.20}$$

is a two-valued Bernouilli distribution. For $\lambda = 0$,

$$dP(m) = (1/\pi)[1 - \tfrac{1}{4}(m)^2]^{\frac{1}{2}}\, dm \quad \text{if} \quad m^2 \leq 4,$$
$$= 0 \qquad \text{otherwise,} \tag{2.21}$$

is the so-called "semicircle distribution" introduced by Wigner.[7] For $\lambda = +1$,

$$dP(m) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}m^2}\, dm \tag{2.22}$$

is the standard Gaussian distribution, already obtained in the Introduction. Graphs of the probability density $dP(m)/dm$ are shown on Fig. 1 for various values of λ. The probability density being even, only positive values of m have been plotted.

C. Calculation of Moments

Henceforward we shall be concerned only with the general time-dependent case, which requires that T has more than one element.

In this section, anticipating the existence of realizations of the generalized commutation relations (2.1)–(2.3), we calculate in terms of λ and Γ the various moments of the parastochastic function,

$$E\{M_1M_2\cdots M_n\} = (\Psi_0, (A_1 + B_1)(A_2 + B_2)\cdots \times (A_n + B_n)\Psi_0). \quad (2.23)$$

Here $M_1, M_2, \cdots, M_n$ stand for $M(t_1), M(t_2), \cdots, M(t_n)$; similarly, Γ_{ij} will stand for $\Gamma(t_i, t_j)$.

Expanding the rhs of (2.23), we obtain a polynomial in A's and B's, which is the sum of all possible products of n factors A or B. Let us first calculate the vacuum expectation of such products.

Lemma 1: The vacuum expectation of a product of A's and B's vanishes if this product ends with an A or begins with a B.

Proof: It follows from Eqs. (2.2) and (2.3) that $(\Psi_0, \cdots, A\Psi_0) = 0$ and $(\Psi_0, B\cdots) = (B^*\Psi_0, \cdots) = (A\Psi_0, \cdots) = 0$, which proves the lemma.

Definition 1: A polynomial in A's and B's is said to be in "normal form" if, in each monomial, all the A's are to the right of the B's.

Using the generalized commutation relation (2.1) in the form

$$A_iB_j = \Gamma_{ij}I + \lambda B_jA_i, \quad (2.24)$$

we can write any product of A's and B's as a polynomial in normal form. We have, for example,

$$\begin{aligned} A_1A_2B_3B_4 &= A_1(\Gamma_{23}I + \lambda B_3A_2)B_4 \\ &= \Gamma_{23}(\Gamma_{14}I + \lambda B_4A_1) \\ &\quad + \lambda(\Gamma_{13}I + \lambda B_3A_1)(\Gamma_{24}I + \lambda B_4A_2) \\ &= \Gamma_{23}\Gamma_{14}I + \lambda\Gamma_{13}\Gamma_{24}I + \lambda\Gamma_{23}B_4A_1 \\ &\quad + \lambda^2\Gamma_{13}B_4A_2 + \lambda^2\Gamma_{24}B_3A_1 + \lambda^3\Gamma_{13}B_3A_2 \\ &\quad + \lambda^4B_3B_4A_1A_2, \end{aligned} \quad (2.25)$$

which is in normal form. By Lemma 1, the vacuum expectation of a product in normal form is zero, except if it is a scalar involving no A's and B's at all [such are the two first terms in the last member of (2.25)], in which case it is already equal to its vacuum expectation (recall that Ψ_0 is a unit vector). In the process of reduction of a product to normal form, repeated use is made of (2.24), which produces every time two new products; in the first one, A_i and B_j are said to be "contracted," while in the second one they are said to be "commuted." In order for a (nonnormal) product to yield, after reduction to normal form, one or several scalar terms, it must be possible to contract every A with a B. Hence, a product with nonvanishing vacuum expectation must have an equal number of A's and B's; this implies an even total number of A's or B's, and it may be concluded that odd-order moments of $M(t)$ vanish.

Now, we consider exclusively products with an equal number of A's and B's, and for such products we define an "operator pairing" as a partition of the set of A's and B's into pairs of one A and one B, the B being always to the right of the A (this ensures that they can be contracted). A convenient graphical representation is obtained by denoting pairs with overbars. For example, for the product $A_1A_2B_3B_4$, the two possible pairings are

$$\overline{A_1\overline{A_2B_3}B_4} \quad \text{and} \quad \overline{A_1\overline{A_2B_3B_4}}.$$

In a given pairing of A's and B's, two pairs A, B and A', B' can have either of the three following topological configurations:

(i) If

$$\overline{AB}\,\overline{A'B'} \quad \text{or} \quad \overline{A'B'}\,\overline{AB},$$

the pairs are "disjoint."

(ii) If

$$\overline{A\overline{A'B'}B} \quad \text{or} \quad \overline{A'\overline{ABB'}},$$

the pairs are "nested."

(iii) If

$$\overline{A\overline{A'B}B'} \quad \text{or} \quad \overline{A'\overline{AB'}B},$$

the pairs are "crossing."

A product that cannot be paired (e.g., $ABBAAB$) has a vanishing vacuum expectation, because it is impossible to contract all the operators. For products that can be paired there is the following lemma.

Lemma 2: To every pairing of a product of n factors A and n factors B corresponds a single scalar term in normal form, which is equal to $\lambda^s\Gamma_{i_1j_1}\Gamma_{i_2j_2}\cdots\Gamma_{i_nj_n}$, where $(i_1, j_1), (i_2, j_2), \cdots, (i_n, j_n)$ are the arguments of the various pairs, and s is the number of couples of crossing pairs.

Proof: In order to be contracted, A_i and B_j must be brought into direct contact by successive commutations; two disjoint or nested pairs can be contracted without any commutation *among* the two pairs, whereas two crossing pairs require one such commutation. The contraction of the pair (A_i, B_j) gives a factor Γ_{ij}, and each commutation gives a factor λ. Q.E.D.

We finally turn back to the evaluation of the moments (2.23), and define an "index pairing" as a partition of the set $\{1, 2, \cdots, n\}$ into pairs. If we expand $M_1M_2\cdots M_n$ in terms of products of A's and

B's, and if we write all possible operator pairings of these products, we obtain, once and only once, every possible index pairing. From this and lemma 2, we infer the following theorem.

Theorem 1 (Generalized Wick's Theorem): The moments of the parastochastic function $M(t)$ are uniquely defined in terms of λ and Γ. Odd-order moments vanish, and even-order moments are given by

$$E\{M_1M_2\cdots M_{2n}\} = \sum \lambda^s\Gamma_{i_1j_1}\Gamma_{i_2j_2}\cdots\Gamma_{i_nj_n}, \quad (2.26)$$

where the sum extends over all possible partitions of the set $\{1, 2, \cdots, 2n\}$ into pairs $(i_1, j_1), \cdots, (i_n, j_n)$, and s is the number of couples of crossing pairs.

A graphical representation of the various terms of (2.26) may be obtained as follows: Put $2n$ successive dots on a horizontal line and write the indices $1, 2, \cdots, 2n$ below; every partition is then represented by connecting the dots in pairs with straight lines running above the main line, and s is the number of intersections of such lines. We illustrate this on a fourth-order moment,

$$E\{M_1M_2M_3M_4\} = \overline{1\,2}\;\overline{3\,4} + \overline{1\,\overline{2\,3}\,4} + \overline{1\,\overline{2\,3}\,4}$$
$$= \Gamma_{12}\Gamma_{34} + \lambda\Gamma_{13}\Gamma_{24} + \Gamma_{14}\Gamma_{23}. \quad (2.27)$$

In Sec. 2D it is shown that the straight lines have an interpretation in terms of particles, called "fictons." However, little use is made of this graphical representation, and it is mainly given for comparison with other works on stochastic equations, based on formal diagram expansions.[5,9–11,13]

Remark 1: For $\lambda = 0$ only noncrossing diagrams survive; this may be considered as the basic reason of the simplifications occurring in perturbation-theoretical approaches to the Kraichnan theory of stochastic equations.[5,13]

We conclude this section with two theorems which are fundamental for the applications to stochastic equations. We need the following definition.

Definition 2: Two parastochastic functions $M(t)$ and $M'(t)$, defined on the same set T and operating in the same space H, are said to be "uncorrelated versions" of the same function if

$$M(t) = A(t) + B(t), \quad M'(t) = A'(t) + B'(t), \quad (2.28)$$

where $\{A(t), B(t)\}$ and $\{A'(t), B'(t)\}$ satisfy Eqs. (2.1–2.3) separately, and

$$A(t)B'(t') - \lambda B'(t')A(t) = 0. \quad (2.29)$$

Taking the adjoint of (2.29), we find that the same equation is still satisfied if we interchange prime and unprimed operators. Two uncorrelated versions of a parastochastic function can, for example, be constructed as follows: Let $t \to i(t)$ be a one-to-one mapping of T onto a set T_1; define a parastochastic function $\tilde{M}(s)$ on the set $T \cup T_1$ with covariance

$$\tilde{\Gamma}(s, s') = \Gamma(s, s'), \quad \text{if } s \in T \text{ and } s' \in T,$$
$$= \Gamma(i^1(s), i^1(s')), \quad \text{if } s \in T_1 \text{ and } s' \in T_1,$$
$$= 0, \quad \text{otherwise;} \quad (2.30)$$

finally put

$$M(t) = \tilde{M}(t), \quad (2.31)$$

$$M'(t) = \tilde{M}(i(t)). \quad (2.32)$$

The existence of uncorrelated versions will thus follow from the general realizability result of Sec. 1D. We now are in position to formulate the following theorem:

Theorem 2 (Pairing Theorem): Let $M(t)$ and $M'(t)$ be two uncorrelated versions of the same parastochastic function. If $p(M)$ is a polynomial depending on a finite number of $M(t)$'s, then

$$E\{M(t)p(M)\} = \frac{d}{d\epsilon} E\{M'(t)p(M + \epsilon M')\}_{\epsilon=0}. \quad (2.33)$$

Proof: Equation (2.33), being linear in $p(M)$, it is enough to prove it for monomials. We write

$$M(t) = M_1 \quad \text{and} \quad p(M) = M_2M_3\cdots M_{2n}. \quad (2.34)$$

By Theorem 1, we have

$$E\{M(t)p(M)\} = \sum \lambda^s\Gamma_{1k}\Gamma_{i_2j_2}\cdots\Gamma_{i_nj_n}; \quad (2.35)$$

in each partition of $\{1, 2, \cdots, 2n\}$ into pairs, k is the index which is paired to 1. The vacuum expectation (2.35) can also be written

$$E\{M(t)p(M)\} = \sum_{k=2}^{2n} \Gamma_{1k} \sum{}' \lambda^s\Gamma_{i_2j_2}\cdots\Gamma_{i_nj_n}; \quad (2.36)$$

$\sum'$ extends over all partitions of $\{2, 3, \cdots, \hat{k}, \cdots, 2n\}$ into pairs (the carat means that k has to be omitted). We assert that

$$\Gamma_{1k} \sum{}' \lambda^s\Gamma_{i_2j_2}\cdots\Gamma_{i_nj_n}$$
$$= E\{M_1'M_2M_3\cdots\hat{M}_kM_k'M_{k+1}\cdots M_{2n}\}. \quad (2.37)$$

Indeed, if we apply Theorem 1 to the rhs of (2.37) and use the covariance $\tilde{\Gamma}$ defined by (2.30), we find that the only possible pairing of M_1' which gives a non-vanishing factor is with M_k'. Inserting (2.37) into

(2.36), we have

$$E\{M(t)p(M)\} = \sum_{k=2}^{2n} E\{M_1' M_2 \cdots \hat{M}_k M_k' M_{k+1} \cdots M_{2n}\} = \frac{d}{d\epsilon} E\{M_1'(M_2 + \epsilon M_2') \cdots (M_{2n} + \epsilon M_{2n}')\}_{\epsilon=0}, \tag{2.38}$$

which concludes the proof.

Remark 2: In this proof we have not used the full strength of Theorem 1 (the explicit value of s was not needed).

Remark 3: If the $M(t)$'s are bounded operators and if there is a measure on the set T, then Theorem 2 can be extended from polynomials to analytic functionals of $M(t)$.

Theorem 3 (Sharpness Theorem): Let $M(t)$ and $M'(t)$ be two uncorrelated versions of the same parastochastic function, and let $\lambda = 0$. If $p(M)$ and $q(M)$ are polynomials depending on a finite number of $M(t)$'s, then

$$E\{M'(t_1)p(M)M'(t_2)q(M)\} = E\{M'(t_1)M'(t_2)\}E\{p(M)\}E\{q(M)\}. \tag{2.39}$$

For the proof we shall need the following lemma:

Lemma 3: Let x be a vector in H such that, for any t in H, $A(t)x = 0$; then

$$(x, p(M)\Psi_0) = E\{p(M)\}(x, \Psi_0). \tag{2.40}$$

Proof: Call the lhs of (2.40) the (x, Ψ_0) expectation of $p(M)$. If $p(M)$ is reduced to normal form, only the scalar terms (the ones involving no A's and B's) have a nonvanishing vacuum or (x, Ψ_0) expectation. Such terms are already equal to their vacuum expectation, and Eq. (2.40) follows.

Proof of Theorem 3: Expanding the lhs of (2.39) and using Lemma 1, we have

$$E\{M'(t_1)p(M)M'(t_2)q(M)\} = E\{A'(t_1)p(M)A'(t_2)q(M)\} + E\{A'(t_1)p(M)B'(t_2)q(M)\}. \tag{2.41}$$

The first term is equal to

$$(\Psi_0, A'(t_1)p(M)A'(t_2)q(M)\Psi_0) = (B'(t_2)p^*(M)B'(t_1)\Psi_0, q(M)\Psi_0), \tag{2.42}$$

where $p^*(M)$ is the adjoint of $p(M)$. It follows from (2.29) and $\lambda = 0$ that, for any t_3 in T,

$$A(t_3)B'(t_2)p^*(M)B'(t_1)\Psi_0 = 0. \tag{2.43}$$

Therefore we can apply Lemma 3 to get

$$\begin{aligned}(B'(t_2)p^*(M)B'(t_1)\Psi_0, q(M)\Psi_0) &= E\{q(M)\}(B'(t_2)p^*(M)B'(t_1)\Psi_0, \Psi_0) \\ &= E\{q(M)\}(\Psi_0, A'(t_1)p(M)A'(t_2)\Psi_0) \\ &= 0.\end{aligned} \tag{2.44}$$

The first term on the rhs of (2.41) thus is zero; to calculate the second term, we notice that $p(M)$, if put in normal form, can be written as a sum of three kinds of terms:

(i) its vacuum expectation $E\{p(M)\}$;

(ii) terms ending with some $A(s)$, which we write $(\cdots)A(s)$;

(iii) terms beginning with some $B(s)$, which we write $B(s)(\cdots)$.

As Eq. (2.41) is linear in $p(M)$, we treat those different terms separately. For (i), we obtain

$$\begin{aligned}E\{A'(t_1)E\{p(M)\}B'(t_2)q(M)\} &= \Gamma(t_1, t_2)E\{p(M)\}E\{q(M)\} \\ &= E\{M'(t_1)M'(t_2)\}E\{p(M)\}E\{q(M)\},\end{aligned} \tag{2.45}$$

which is exactly the rhs of (2.39). For (ii), we obtain $E\{A'(t_1)(\cdots)A(s)B'(t_2)q(M)\}$, which vanishes, because $A(s)B'(t_2) = 0$. Finally, for (iii), we obtain

$$E\{A'(t_1)B(s)(\cdots)B'(t_2)q(M)\},$$

which vanishes again, because $A'(t_1)B(s) = 0$. Q.E.D.

Remark 4: Like Theorem 2, Theorem 3 can again be extended, under suitable conditions, from polynomials to analytic functionals of $M(t)$.

D. Fock-Space Realizations and Fictons

It is known that the standard commutation relations for bosons and fermions can be realized in symmetrical and antisymmetrical tensor algebras; these are the so-called "Fock representations." (Here we use the word "realization," leaving "representation" for representation by random functions; see Kastler[14] for an excellent account of the "Fock representation.") It is therefore natural to look for similar realizations of our generalized commutation relations. So far, we have been able to find such realizations only for $\lambda = +1, -1$, and 0; hence, the problem of realizability for arbitrary values of λ remains open.

The construction of realizations requires the following result: If $\Gamma(t, t')$ is a real symmetric kernel of positive type defined on $T \times T$, it is possible to find a Hilbert space $\mathcal{H}$ with inner product $((\cdot, \cdot))$ and a mapping $x(\cdot)$: $T \to \mathcal{H}$ such that, for any $t, t' \in T$,

$$((x(t), x(t'))) = \Gamma(t, t'). \tag{2.46}$$

Proof: (See exercise III.3.2 of Ref. 18.) Let $\mathcal{K}_0$ be the pre-Hilbert space of formal finite linear combinations of functions $\Gamma(t, \cdot)$, provided with the inner product

$$\left(\left(\sum_i c_i\Gamma(t_i, .), \sum_j c'_j\Gamma(t'_j, .)\right)\right)_0 = \sum_{i,j} c_i c'_j \Gamma(t_i, t'_j). \tag{2.47}$$

Then, take the mapping $x(\cdot)$: $t \to \Gamma(t, \cdot)$, and take for $\mathcal{K}$ the completed space of $\mathcal{K}_0$.

It can be shown that if $y(\cdot)$ is another mapping from T to $\mathcal{K}$ satisfying (2.46), there is a unitary transformation U of $\mathcal{K}$ such that $y(t) = Ux(t)$. In other words, the family of inner products (2.46) uniquely defines the shape of the curve $x(t)$.

In case $T = R$ and Γ is a continuous and *stationary* covariance, a more explicit construction will be given, which will be needed in Sec. 3D, but not in the sequel of this section. Bochner's Theorem asserts that

$$\Gamma(t) = \int_R e^{itv}\, dF(v), \tag{2.48}$$

where $dF(v)$ is a positive finite even measure. Now take for $\mathcal{K}$ the space $L_2(dF(v))$ of complex functions $h(v)$ such that

$$\|h(v)\|^2_{\mathcal{K}} = \int_R |h(v)|^2\, dF(v) < +\infty. \tag{2.49}$$

The mapping from T to $\mathcal{K}$,

$$x(\cdot)\colon t \to e^{itv}, \tag{2.50}$$

obviously fulfills (2.46).

The mapping $x(\cdot)$ is now used to construct explicit realizations of the generalized commutation relations (2.1)–(2.3). Let $\mathcal{K}^{\otimes p}$ be the pth-tensor power of $\mathcal{K}$, completed for the norm, with the usual convention that $\mathcal{K}^{\otimes 0}$ is the field C of complex numbers. The unrestricted, antisymmetric, and symmetric tensor algebra are, respectively, defined as

$$\mathfrak{T}(\mathcal{K}) = \bigoplus_{p=0}^{\infty} \mathcal{K}^{\otimes p}, \tag{2.51}$$

$$\mathfrak{G}(\mathcal{K}) = \bigoplus_{p=0}^{\infty} \mathcal{A}\mathcal{K}^{\otimes p}, \tag{2.52}$$

$$\mathfrak{G}(\mathcal{K}) = \bigoplus_{p=0}^{\infty} \mathcal{K}\mathcal{S}^{\otimes p}, \tag{2.53}$$

where $\oplus$ denotes a direct sum, and where $\mathcal{A}$ and $\mathcal{S}$ are the antisymmetrization and symmetrization projectors (see Appendix A of Ref. 14). The three cases $\lambda = 0$, -1, and $+1$ are now treated separately:

(i) $\lambda = 0$. Hilbert space: $H = \mathfrak{T}(\mathcal{K})$; vacuum state $\Psi_0 = 1$ = vector whose components are 1 in C and 0 in $\mathcal{K}^{\otimes p}$ for $p > 0$;

$$\text{creation operator: } B(t) = x(t)\otimes, \tag{2.54}$$

where $\otimes$ denotes the tensor product.

(ii) $\lambda = -1$. Hilbert space: $H = \mathfrak{G}(\mathcal{K})$; vacuum state: $\Psi_0 = 1$;

$$\text{creation operators: } B(t) = (\mathcal{N})^{\frac{1}{2}}\mathcal{A}x(t)\otimes; \tag{2.55}$$

the operator $(\mathcal{N})^{\frac{1}{2}}$ reduces to $(p)^{\frac{1}{2}}$ times the identity on $\mathcal{K}^{\otimes p}$.

(iii) $\lambda = +1$. Hilbert space: $H = \mathfrak{G}(\mathcal{K})$; vacuum state: $\Psi_0 = 1$;

$$\text{creation operator: } B(t) = (\mathcal{N})^{\frac{1}{2}}\mathcal{S}x(t)\otimes. \tag{2.56}$$

In each case $A(t)$ is defined as the adjoint of $B(t)$. For $\lambda = 0$ it is a pure routine calculation to check that Eqs. (2.1–2.3) are satisfied; for $\lambda = \pm 1$ the proof is a little bit more involved, but may be found in any textbook on quantum-field theory. The realizations given above of $A(t)$, $B(t)$, and $M(t) = A(t) + B(t)$ will be called the "canonical realizations."

We now indicate a few interesting properties of the parastochastic operators $M(t)$, some of which are independent of the realization.

(i) For $\lambda = 0$ and -1, the operators $M(t)$ are bounded.

Proof for $\lambda = 0$*:* For any Ψ in H,

$$\|B(t)\Psi\|^2 = (B(t)\Psi, B(t)\Psi) = (A(t)B(t)\Psi, \Psi) = \Gamma(t, t)\,\|\Psi\|^2; \tag{2.57}$$

hence

$$\|B(t)\| = \|A(t)\| = (\Gamma(t, t))^{\frac{1}{2}}, \tag{2.58}$$

and therefore,

$$\|M(t)\| \leq 2(\Gamma(t, t))^{\frac{1}{2}}. \tag{2.59}$$

Proof for $\lambda = -1$*:* For any Ψ in H,

$$\|A(t)\Psi\|^2 + \|B(t)\Psi\|^2 = ([B(t)A(t) + A(t)B(t)]\Psi, \Psi) = \Gamma(t, t)\,\|\Psi\|^2 \tag{2.60}$$

and the same estimate (2.5) holds as for $\lambda = 0$.

(ii) For $\lambda = +1$, the operators $M(t)$ are unbounded.

Proof: It is known from Sec. 2.B or the Introduction that for fixed t,

$$(\Psi_0, [A(t) + B(t)]^n\Psi_0) = E\{m^n(\omega)\}, \tag{2.61}$$

[18] J. Neveu, *Calcul des probabilités* (Mason Cie., Paris, 1964).

where $m(\omega)$ is a Gaussian random variable. Boundedness of $M(t)$ would imply

$$E\{m^n(\omega)\} < (\text{const})^n, \tag{2.62}$$

which is impossible for a Gaussian random variable. It can nevertheless be shown that the $M(t)$'s have dense domains of definition.

(iii) For $\lambda = +1$, the operators $M(t)$ commute among themselves.

Proof: It follows from the presence of the symmetrization operator in (2.56) that the $B(t)$'s, and therefore the $A(t)$'s, commute among themselves; then, using (2.1) and the symmetry of $\Gamma(t, t')$, we have

$$M(t)M(t') = M(t')M(t), \quad (\lambda = +1). \tag{2.63}$$

F. Fictons

In quantum-field theory and the many-body problem, the spaces $C, \mathcal{H}, \cdots, \mathcal{H}^{\otimes p}$ are known as the zero-, one-, $\cdots$, n-particle space. The operator $B(t)$, acting on the n-particle space, has values in the $(n+1)$-particle space; thus, it "creates" a particle. Similarly, the operator $A(t)$ "destroys" a particle. The problems that are dealt with in this paper are mathematical ones, with no real particles present; nevertheless, it is convenient to speak of fictive particles, which will be called "fictons." For example, in the diagrams of Eq. (2.27), any curved line connecting t_i and t_j may be thought of as a ficton, created at time t_i and destroyed at time t_j.

At present, physics knows only two kinds of particles, bosons and fermions. Wavefunctions for n particles, which must be either completely symmetrical (bosons) or antisymmetrical (fermions), belong to the symmetrized or antisymmetrized nth-tensor power of the one-particle space. Noticing that for $\lambda = 0$ the creation and destruction operators have been realized in the *unrestricted* tensor algebra (2.51), and that there is a similar lack of symmetry in the Boltzmann statistic of distinguishable particles, it may be concluded that *for* $\lambda = +1$, -1, *and* 0, *fictons are, respectively, bosons, fermions, and "Boltzmannons."* To make up for the lack of real existence of Boltzmannons, it will be shown in the second part of this paper that linear stochastic equations which are intractable with bose-fictons become soluble with Boltzmann-fictons.

E. Random Representations of Parastochastic Functions

The representation of a family of self-adjoint operators by random variables is beset with a fundamental difficulty; as shown by Nelson,[19] such representations are to be found only for commuting operators. Yet, it is easily seen that if $\lambda \neq +1$, the parastochastic operators $M(t)$ do not commute. Indeed, Theorem 1 gives, for the fourth-order moment,

$$E\{M_1M_2M_3M_4\} = \Gamma_{12}\Gamma_{34} + \lambda\Gamma_{13}\Gamma_{24} + \Gamma_{14}\Gamma_{23}, \tag{2.64}$$

which, for $\lambda \neq 1$, differs from

$$E\{M_1M_3M_2M_4\} = \lambda\Gamma_{12}\Gamma_{34} + \Gamma_{13}\Gamma_{24} + \Gamma_{14}\Gamma_{23}. \tag{2.65}$$

A possible way of avoiding this difficulty is to look for representations by *random matrices* $\mathcal{M}(\omega; t)$ satisfying

$$E\{M(t_1)\cdots M(t_n)\} = \operatorname{Tr} E\{\mathcal{M}(\omega; t_1)\cdots\mathcal{M}(\omega; t_n)\}, \tag{2.66}$$

where Tr denotes the trace. Both members of (2.66) are then invariant under cyclic permutations (for the rhs it is a well-known property of the trace, and for the lhs it follows from Theorem 1). No further *a priori* invariance is required. If $T = R$, it may still be possible to find scalar-valued random representations by restricting Eq. (2.9) to *ordered moments*, for which $t_1 \leq t_2 \leq \cdots \leq t_n$.

Explicit random representations of the parastochastic function are now given for $\lambda = +1$, 0, and -1. The existence of random representations for other values of λ remains an open problem.

1. $\lambda = +1$. *Representation by Gaussian Random Functions*

In the canonical realization, the operators $M(t)$ commute among themselves [see Eq. (2.63)], and may therefore be represented by scalar random functions. Let us evaluate the multivariate characteristic function

$$F(z_1, \cdots, z_n) = E\left\{\exp\left(i\sum_{j=1}^{n} z_jM_j\right)\right\}. \tag{2.67}$$

It follows from (2.1) that the commutator of $\Sigma_j(z_jA_j)$ and $\Sigma_j(z_jB_j)$ is a scalar; we may therefore use (1.2) to obtain

$$F(z_1, \cdots, z_n) = E\Big\{\exp\left[i\Sigma_j(z_jB_j)\right] \times \exp\left(-\tfrac{1}{2}\sum_{i,j} z_iz_j\Gamma_{ij}\right) \times \exp\left[i\Sigma_j(z_jA_j)\right]\Big\}. \tag{2.68}$$

By Eq. (2.3), we have

$$\exp\left[i\Sigma_j(z_jA_j)\right]\Psi_0 = \Psi_0. \tag{2.69}$$

[19] E. Nelson, *Dynamical Theories of Brownian Motion* (*Mathematical Notes*) (Princeton University Press, Princeton, N.J., 1967), Theorem 14.1.

Hence,

$$F(z_1, \cdots, z_n) = \exp\left(-\tfrac{1}{2}\sum_{i,j=1}^{n} z_i z_j \Gamma_{ij}\right), \quad (2.70)$$

which is the characteristic function of a multivariate Gaussian process with zero mean value and covariance Γ.[15] To summarize: *For $\lambda = +1$, the parastochastic function $M(t)$ can be represented by a Gaussian random function with zero mean value and covariance Γ.*

Remark 5: This result could have been inferred from Theorem 1. Indeed, Eq. (2.26) reduces for $\lambda = +1$ to a well-known formula expressing even-order moments of a Gaussian random function in terms of its covariance. The most pleasant properties of Gaussian random functions are probably that (i) moments are expressible in terms of the sole covariance, and (ii) that they are invariant under linear transformations. The parastochastic functions retain these properties for arbitrary λ. The first one follows from Theorem 1. To prove the second one, assume that there is a measure $d\mu(t)$ on T; let $A(t)$ and $B(t)$ satisfy Eqs. (2.1–2.3), and let $K(t, t')$ be the kernel of a linear integral operator; define

$$\begin{aligned} A'(t) &= \int_T K(t, t')A(t')\, d\mu(t'), \\ B'(t) &= \int_T K(t, t')B(t')\, d\mu(t'), \\ M'(t) &= \int_T K(t, t')M(t')\, d\mu(t'); \end{aligned} \quad (2.71)$$

a simple calculation shows that $A'(t)$ and $B'(t)$ still satisfy Eqs. (2.1)–(2.3), with the new covariance

$$\Gamma'(t, t') = \int_{T\times T} K(t, t_1)K(t', t_2)\Gamma(t_1, t_2)\, d\mu(t_1)\, d\mu(t_2). \quad (2.72)$$

2. $\lambda = 0$. *Representation by Random Wigner Matrices*

Random matrices have been introduced by Wigner[7] in the study of energy levels of complex systems, especially of heavy nuclei. A considerable amount of material has been published on this subject, concerned mostly with the distribution of eigenvalues and of the spacings between eigenvalues; the reader is referred to Porter's book for a review and a collection of reprints.[20]

Definition 3: Let T and $\Gamma(t, t')$ be defined as before. An infinite double sequence of real random functions $\mathcal{M}_{\alpha\beta}(\omega; t)$, with $\alpha, \beta = 1, 2, \cdots, n, \cdots$, is called a "random Wigner matrix function" (or simply a "Wigner matrix") if

(i) the matrix is symmetric, i.e.,

$$\mathcal{M}_{\alpha\beta}(\omega; t) = \mathcal{M}_{\beta\alpha}(\omega; t); \quad (2.73)$$

(ii) the $\mathcal{M}_{\alpha\beta}(\omega; t)$'s with $\alpha < \beta$ ($\alpha = \beta$, respectively) are independent versions of a same Gaussian random function, defined on T, with zero mean value and covariance $\Gamma(t, t')$ [$2\Gamma(t, t')$, respectively].

We denote by $\mathcal{M}^{(N)}(\omega; t)$ the finite truncated matrices

$$\mathcal{M}^{(N)}(\omega; t) = \mathcal{M}^{(N)}_{\beta\alpha}(\omega; t), \quad \alpha, \beta = 1, \cdots, N. \quad (2.74)$$

It can easily be shown that $\mathcal{M}^{(N)}(\omega; t)$ is invariant under arbitrary orthogonal transformations.

The following theorem asserts that parastochastic functions with $\lambda = 0$ can be represented by Wigner matrices; the proof is not essential for the understanding of the sequel.

Theorem 4 (Equivalence Theorem): Let $M(t)$ be a parastochastic function defined on an arbitrary set T with $\lambda = 0$ and covariance $\Gamma(t, t')$, and let $\mathcal{M}_{\alpha\beta}(\omega; t)$ be a Wigner matrix with the same covariance. For any integers α, β and any $t_1, \cdots, t_n \in T$, the following mean square (m.s.) convergence takes place:

$$\text{m.s. lim}_{N\to\infty} \left[\frac{\mathcal{M}^{(N)}(\omega; t_1)}{(N)^{\frac{1}{2}}} \cdots \frac{\mathcal{M}^{(N)}(\omega; t_n)}{(N)^{\frac{1}{2}}}\right]_{\alpha\beta} = E\{M(t_1)\cdots M(t_n)\}\delta_{\alpha\beta}, \quad (2.75)$$

with

$$\begin{aligned} \delta_{\alpha\beta} &= 1 \quad \text{if} \quad \alpha = \beta, \\ &= 0 \quad \text{if} \quad \alpha \neq \beta. \end{aligned} \quad (2.76)$$

In particular, this theorem states that the random lhs of (2.75) becomes statistically sharp as $N \to \infty$. To show mean square convergence, it suffices to prove (i) convergence of the expectation value of the lhs to the rhs, (ii) convergence of the expectation value of the squared lhs to the squared rhs. Our proof of the first point will be somewhat similar to Wigner's original derivation of the semicircle distribution[7]; however, different notations will be used.

Notations and definitions: Matrix indices varying from 1 to N are denoted by Greek letters and subscripts to such indices by Latin letters. In calculating the expectation value of the lhs of (2.75) it will be necessary to consider sequences of matrix indices,

$$\alpha, \gamma_1, \gamma_2, \cdots, \gamma_{n-1}, \beta, \quad (2.77)$$

[20] C. E. Porter, "Statistical Theories of Spectra: Fluctuations," in *Perspectives in Physics* (Academic Press, Inc., New York, 1965).

called "step sequences"; α and β are "external" indices, and $\gamma_1, \cdots, \gamma_{n-1}$ are "internal" indices. The pair of successive matrix indices (γ_{i-1}, γ_i) is called the ith "step"; γ_{i-1} and γ_i are the "entrance" and the "exit" of the ith step. (Notice that the exit of a step is also the entrance of the next step.) If the entrance and the exit are equal, the step is "diagonal." Two steps (γ_{i-1}, γ_i) and (γ_{j-1}, γ_j) are "equivalent" ("opposite", respectively) if $\gamma_{i-1} = \gamma_{j-1}$ and $\gamma_i = \gamma_j$ ($\gamma_{i-1} = \gamma_j$ and $\gamma_i = \gamma_{j-1}$, respectively). A "pairing" is defined for the class of all step-sequences with a given number $n = 2p$ of steps as a partition of the set $\{1, 2, \cdots, 2p\}$ into pairs (i_1, j_1), $(i_2, j_2), \cdots, (i_p, j_p)$. For a given step-sequence, a pairing can also be considered as a partition into pairs of the set of steps. The first (last) step in a pair is called a "creative" ("destructive") step. A given step-sequence and pairing are "compatible" if paired steps are always equivalent or opposite. In a given pairing, two pairs are "crossing" if the creative step of one pair occurs between the creative and destructive step of the other pair. A pairing is "noncrossing" if there are no crossing pairs. We are now in position to prove the following:

Lemma 4: In the class of all step-sequences with $2p$ steps and given external indices α and β, the number $C_{2p}^{(N)}(\Pi)$ of step-sequences which are compatible with a given pairing Π of $\{1, 2, \cdots, 2p\}$ is given by

$$C_{2p}^{(N)}(\Pi) = N^p + O(N^{p-1}), \quad \text{if } \alpha = \beta \text{ and the pairing is noncrossing,}$$
$$= O(N^{p-1}), \quad \text{otherwise.} \tag{2.78}$$

Proof: Without the restriction to compatible sequences, the indices $\gamma_1, \gamma_2, \cdots, \gamma_{2p-1}$ would vary independently from 1 to N, and there would be N^{2p-1} step-sequences. Compatibility imposes constraints which reduce this number. Calling constraints "independent" if none of them can be deduced from the other ones, we obtain that the number of compatible step-sequences is N^{2p-1-q}, where q is the number of independent constraints on the internal indices. The condition that paired steps be equivalent or opposite implies that the exits of the p destructive steps must repeat the exits or the entrances of the corresponding creative steps. These p constraints will be called "exit-constraints." If $\alpha \neq \beta$, the p exit-constraints are easily proved to be independent; this results in at most N^{p-1} compatible sequences; therefore, as far as Lemma 4 is concerned, the case $\alpha \neq \beta$ is settled.

From now on it is assumed that $\alpha = \beta$; this equality may reduce the number of independent constraints to $p - 1$. Lemma 4 is now proved by recurrence on p, the number of pairs. For $p = 1$ the lemma is obvious. Assume that it holds for all pairings with $(p - 1)$ pairs. Let there be given a pairing of $2p$ ordered steps into p pairs, and let the first destructive step be the ith. There are three possible situations:

(i) The ith-step (γ_{i-1}, γ_i) and the $(i - 1)$th-step $(\gamma_{i-2}, \gamma_{i-1})$ are paired and opposite.

(ii) The ith and $(i - 1)$th-steps are paired and equivalent.

(iii) The ith- and $(i - 1)$th-steps are not paired.

First consider step-sequences of type (i); they are of the following form:

$$\alpha, \gamma_1, \cdots, \gamma_{i-3}, \gamma_i, \gamma_{i-1}, \gamma_i, \gamma_{i+1}, \cdots, \gamma_{2n-1}, \alpha.$$

Now delete the ith- and $(i - 1)$th-steps to obtain the following subsequence:

$$\alpha, \gamma_1, \cdots, \gamma_{i-3}, \gamma_i, \gamma_{i+1}, \cdots, \gamma_{2n-1}, \alpha.$$

The original step-sequence and the subsequence have the same number of crossing pairs because the deleted pair $(i - 1, i)$ does not cross any other one. From the recurrence hypothesis and the observation that the matrix index γ_{i-1}, which does not appear in the subsequence, can take any value from 1 to N, it is inferred that the number of step-sequences of type (i) is $N^p + O(N^{p-1})$ if the original pairing is noncrossing, and $O(N^{p-1})$ otherwise. Now consider sequences of type (ii). The constraint that the entrances of the ith- and $(i - 1)$th-steps be equal is independent of the $p - 1$ exit-constaints; for the entrance of the ith-step, which is the first destructive one, cannot be the exit of a destructive step; therefore there are $O(N^{p-1})$ step-sequences of type (ii); it does not matter whether or not they have any crossing pairs. Finally, consider sequences of type (iii). As the $(i - 1)$th-step is necessarily a creative one, there must be at least one crossing. The constraint that the entrance of the ith-step be equal to the entrance or the exit of the corresponding creative step [which is not the $(i - 1)$th] is, for the same reason as in the preceding case, independent of the $p - 1$ exit-constraints; therefore there are again $O(N^{p-1})$ step-sequences of type (iii). The proof of Lemma 4 is completed.

Proof that Theorem 4 Holds in the Mean: Writing out the matrix products in (2.75), we want to show that

$$\lim_{N\to\infty} N^{-\frac{1}{2}n} \sum_{\gamma_1,\cdots,\gamma_{n-1}=1}^{N} E\{\mathcal{M}_{\alpha\gamma_1}^{(N)}(\omega; t_1) \times \mathcal{M}_{\gamma_1\gamma_2}^{(N)}(\omega; t_2) \cdots \mathcal{M}_{\gamma_{n-1}\beta}^{(N)}(\omega; t_n)\}$$
$$= E\{M(t_1) \cdots M(t_n)\}\delta_{\alpha\beta}. \tag{2.79}$$

It will be convenient to associate the step-sequence (2.77) to the lhs of (2.79); then, to every matrix element $\mathcal{M}^{(N)}_{\gamma_{i-1}\gamma_i}(\omega; t_i)$ corresponds the step (γ_{i-1}, γ_i). As the matrix elements are Gaussian with zero mean value, odd-order moments vanish and even-order moments are given by the sum over all pairings of $\{1, 2, \cdots, n\}$ of the products of covariances of pairs (see Remark 5). Equation (2.79) is satisfied for odd n because both members vanish. Now assume $n = 2p$; the contribution to (2.79) arising from a given pairing

$$\Pi = \{(i_1, j_1), \cdots, (i_p, j_p)\} \tag{2.80}$$

of the set $\{1, 2, \cdots, 2p\}$ is

$$N^{-p} \sum_{\gamma_1, \cdots, \gamma_{2p-1}=\gamma}^{N} E\{\mathcal{M}^{(N)}_{\gamma_{i_1-1}\gamma_{i_1}}(\omega; t_{i_1})\mathcal{M}^{(N)}_{\gamma_{j_1-1}\gamma_{j_1}}(\omega; t_{j_1})\} \cdots E\{\mathcal{M}^{(N)}_{\gamma_{i_p-1}\gamma_i}(\omega; t_{i_p})\mathcal{M}^{(N)}_{\gamma_{j_p-1}\gamma_{j_p}}(\omega; t_{j_p})\}. \tag{2.81}$$

The very definition of the Wigner matrix gives

$$E\{\mathcal{M}^{(N)}_{\gamma_{i-1}\gamma_i}(\omega; t_i)\mathcal{M}^{(N)}_{\gamma_{j-1}\gamma_j}(\omega; t_j)\}$$

$= \Gamma(t, t')$, if the ith- and the $(i-1)$th-steps are equivalent or opposite, and not diagonal,

$= 2\Gamma(t, t')$, if they are equivalent and diagonal,

$= 0$, otherwise. (2.82)

Discarding for a while diagonal steps, we find that (2.81) is equal to

$$\Gamma(t_{i_1}, t_{j_1}) \cdots \Gamma(t_{i_p}, t_{j_p}) N^{-p} C^{(N)}_{2p}(\Pi),$$

where $C^{(N)}_{2p}(\Pi)$ is the number of compatible step-sequences. Letting $N \to \infty$, using Lemma 4, and summing over all partitions Π, we find that the lhs of (2.79) converges to

$$\sum_{\Pi} \epsilon \Gamma(i_1, j_1) \cdots \Gamma(i_p, j_p) \delta_{\alpha\beta}, \tag{2.83}$$

where ϵ is equal to one for noncrossing pairings and to zero otherwise. This is precisely the value assigned to the rhs of (2.79) by Theorem 3 for $\lambda = 0$. It was legitimate, indeed, to discard step-sequences with diagonal steps, for the presence of a diagonal step imposes an additional constraint, and the corresponding contribution then vanishes in the limit $N \to \infty$.

Proof that Theorem 4 Holds in the Mean Square: To evaluate the expectation value of the square of the lhs of (2.75) we proceed as above, and are led to consider "double step-sequences"

$$\alpha, \gamma_1, \cdots, \gamma_{n-1}, \beta; \alpha, \gamma'_1, \cdots, \gamma'_{n-1}, \beta.$$

There are "factored" pairings Π_{12} of this double step-sequence which are merely products of a pairing Π_1 of the left-hand step-sequence and a pairing Π_2 of the right-hand step-sequence. Symbolically, we write $\Pi_{12} = \Pi_1\Pi_2$. For such pairings the number of compatible double step-sequences is obviously the product of the numbers of compatible left- and right-hand sequences. Therefore, the contribution of $\Pi_1\Pi_2$ to the expectation value of the square of (2.75) is the product of the separate contributions of Π_1 and Π_2 to the expectation value of (2.75). Pairings of the double step-sequence which are not factored give, in the limit $N \to \infty$, a vanishing contribution to the expectation value of the square of 2.75, for they introduce at least one additional constraint.

Summing over factored pairings only, we obtain that the limit of the expectation value of the square of (2.75) is also equal to the square of the limit of its expectation value, which proves mean-square convergence. Q.E.D.

Remark 6: In the time-independent case, Theorem 5 is a well-known result.[7,21,22] Indeed, assuming $\Gamma(t, t') = 1$, taking the trace of (2.75), and using (2.19) and (2.21), we obtain

$$\text{m.s.}\lim_{N\to\infty} \frac{1}{N} \operatorname{Tr}\left[\frac{\mathcal{M}(\omega)}{(N)^{\frac{1}{2}}}\right]^n = \frac{1}{\pi}\int_{-2}^{2} m^n \left\{1 - \left(\frac{m}{2}\right)^2\right\}^{\frac{1}{2}} dm. \tag{2.84}$$

From this it is easily inferred (see Ref. 22) that if $W_N(\omega; m)$ denotes the proportion of eigenvalues of the matrix $\mathcal{M}(\omega)/(N)^{\frac{1}{2}}$ that are less than m, then

$$\text{m.s.}\lim_{N\to\infty} W_N(\omega; m) = W(m), \tag{2.85}$$

where $W(m)$ is the semicircle distribution, the density of which is given by (2.21).

Actually Theorem 4 has been proved, in the time-independent case, under far less stringent conditions; such as cannot be relaxed are the conditions of symmetry and independence of the entries of the random matrix $\mathcal{M}(\omega)$. Arnold[21] proved that if (i) the $\mathcal{M}_{\alpha\beta}$'s with $\alpha \leq \beta$ have the same distribution function F (non-Gaussian), satisfying $\int x^6\, dF < \infty$ and $\int x\, dF = 0$, (ii) the $\mathcal{M}_{\alpha\alpha}$'s have the same distribution function G, satisfying $\int x^4\, dG < \infty$, then Theorem 4 is still true; moreover, mean-square convergence can be replaced by *almost sure* convergence. It is not difficult to extend Arnold's proof to the time-dependent case.

[21] L. Arnold, "On the Asymptotic Distribution of the Eigenvalues of Random Matrices," M.R.C. Report 736, The University of Wisconsin, Madison, Wisconsin.

[22] U. Grenander, *Probabilities on Algebraic Structures* (John Wiley & Sons, Inc., New York, 1963).

3. $\lambda = -1$, *and Exponential Covariance. Representation by a Dichotomic Markov Process*

Under the assumptions that $T = \mathbf{R}$, that $\lambda = -1$, and that $\Gamma(t, t') = \exp\{-|t - t'|\}$ (the exponential covariance), it will be shown that, as far as only *ordered moments* are concerned, the parastochastic function $M(t)$ can be represented by a dichotomic (two-valued) Markov process.

Theorem 5 (Cancellation Theorem): If $M(t)$ is a parastochastic function of a real variable t with $\lambda = -1$ and exponential covariance, then

(i) ordered moments of even order are given by

$$E\{M(t_{2n}) \cdots M(t_2)M(t_1)\} = \exp\{-(t_{2n} - t_{2n-1})\} \cdots \exp\{-(t_4 - t_3)\} \times \exp\{-(t_2 - t_1)\}, \quad t_{2n} \geq \cdots \geq t_2 \geq t_1. \quad (2.86)$$

(ii) There is a stationary dichotomic Markov process $m(\omega; t)$ with values ± 1 and transition probabilities $(t > 0)$

$$P(+1 \mid +1; t) = P(-1 \mid -1; t) = \tfrac{1}{2}(1 + e^{-t}),$$
$$P(+1 \mid -1; t) = P(-1 \mid +1; t) = \tfrac{1}{2}(1 - e^{-t}) \quad (2.87)$$

such that for $t_{2n} \geq \cdots \geq t_2 \geq t_1$,

$$E\{m(\omega; t_{2n}) \cdots m(\omega; t_2)m(\omega; t_1)\} = \exp[-(t_{2n} - t_{2n-1})] \cdots \exp[-(t_4 - t_3)] \exp[-(t_2 - t_1)]. \quad (2.88)$$

Proof of (i): By Theorem 3, $E\{M(t_{2n}) \cdots M(t_2)M(t_1)\}$ is a sum of contributions arising from all possible pairings of the time-arguments $t_{2n}, \cdots, t_2, t_1$. The λ factor associated to every pairing is $+1$ or -1 according as the number of crossings is even or odd. Recall that two pairs can be either disjoint, crossing, or concentric (see Sec.2.C). Equation (2.86) asserts that only the pairing $(t_{2n}, t_{2n-1}), \cdots, (t_4, t_3), (t_2, t_1)$, where all pairs are disjoint, contributes to ordered moments of order $2n$. Indeed, it will be shown that contributions from all other pairings cancel out. If t_i and t_j are paired and $t_i \geq t_j$, we shall say that a pair begins at t_j and ends at t_i. Consider a given pairing, distinct from the above-mentioned pairing; let t_i be the first time such that a pair begins at t_i and does not end at t_{i+1}; then another pair begins at t_{i+1}. Let the pairs beginning at t_i and t_{i+1} end at t_k and t_1, respectively; if $t_1 \geq t_k$, we say that the pairing is of type A; if $t_k \geq t_1$, we say that it is of type B. There is a one-to-one correspondence between type A and type B pairings, obtained by interchanging the ends of the pairs beginning at t_i and t_{i+1} and leaving all other pairs unchanged. The products of covariances for corresponding pairings are the same.

Proof: The unchanged pairs give the same factors, and the products of the covariances of the two remaining pairs are identical because, for $t_1 \geq t_k \geq t_j \geq t_i$,

$$\exp(-|t_k - t_i|) \exp(-|t_1 - t_j|) = \exp(-|t_1 - t_i|) \exp(-|t_k - t_j|). \quad (2.89)$$

For corresponding pairings the parities of the number of crossings of the two special pairs with an arbitrary third one are the same. Proof: A third pair can have only ten topologically distinct positions with respect to t_i, t_{i+1}, t_k, and t_1; in each case the assertion is trivially true. As for type A, there is the additional crossing of (t_i, t_k) with (t_{i+1}, t_1); it follows that the parities of the total number of crossings for corresponding pairings are different. Hence, corresponding pairings cancel each other. Q.E.D.

Proof of (ii): The transition-probability matrix

$$P_{\alpha\beta}(t) = \frac{1}{2}\begin{bmatrix} 1 + e^{-t}; & 1 - e^{-t} \\ 1 - e^{-t}; & 1 + e^{-t} \end{bmatrix}, \quad \alpha, \beta = 1, 2, \quad (2.90)$$

satisfies the Chapman–Kolmogorov semigroup equation[15.23]

$$\sum_{\beta} P_{\alpha\beta}(t_1)P_{\beta\gamma}(t_2) = P_{\alpha\gamma}(t_1 + t_2), \quad t_1, t_2 \geq 0, \quad (2.91)$$

and therefore defines a Markov Process $m(\omega; t)$ with possible values $m_1 = +1$ and $m_2 = -1$. A well-known property of Markov processes is that the joint probability that $m(\omega; t_1) = m_{\alpha_1}, \cdots$, and $m(\omega; t_p) = m_{\alpha_p}$ is, for $t_p \geq \cdots \geq t_1$, given by

$$(\tfrac{1}{2})P_{\alpha_p\alpha_{p-1}}(t_p - t_{p-1}) \cdots P_{\alpha_2\alpha_1}(t_2 - t_1),$$

where the factor one-half is the unconditional probability that $m(\omega; t_1) = \pm 1$. [Equal initial probabilities are affected to m_1 and m_2 in order to make the process $m(\omega; t)$ stationary.]

To calculate the pth-order moment

$$E\{m(\omega; t_p) \cdots m(\omega; t_1)\} = \sum_{\alpha_1, \cdots, \alpha_p = 1,2} m_{\alpha_1} \cdots m_{\alpha_p} P_{\alpha_p\alpha_{p-1}}(t_p - t_{p-1}) \cdots P_{\alpha_2\alpha_1}(t_2 - t_1) \cdot \tfrac{1}{2}, \quad (2.92)$$

it is convenient to define $\Sigma_+(p)$ [$\Sigma_-(p)$, respectively] as equal to the rhs of (2.92) with α_p held fixed and equal to one (two, respectively). It follows from (2.90)

[23] A. T. Bharucha-Reid, *Elements of the Theory of Markov Processes and their Applications* (McGraw-Hill Book Co., New York, 1960).

that Σ_+ and Σ_- satisfy the recurrence relations

$$\Sigma_+(p+1) = \tfrac{1}{2}\{1 + \exp[-(t_{p+1} - t_p)]\}\Sigma_+(p) + \tfrac{1}{2}\{1 - \exp[-(t_{p+1} - t_p)]\}\Sigma_-(p), \tag{2.93}$$

$$\Sigma_-(p+1) = \tfrac{1}{2}\{\exp[-(t_{p+1} - t_p)] - 1\}\Sigma_+(p) - \tfrac{1}{2}\{1 + \exp[-(t_{p+1} - t_p)]\}\Sigma_-(p). \tag{2.94}$$

Adding and substracting (2.93) and (2.94), we obtain

$$\Sigma_+(p+1) + \Sigma_-(p+1) = \exp\{-(t_{p+1} - t_p)\}[\Sigma_+(p) - \Sigma_-(p)], \tag{2.95}$$

$$\Sigma_+(p+1) - \Sigma_-(p+1) = \Sigma_+(p) + \Sigma_-(p). \tag{2.96}$$

Changing p into $p+1$ and eliminating $\Sigma_+(p+1) - \Sigma_-(p+1)$, we have

$$\Sigma_+(p+2) + \Sigma_-(p+2) = \exp\{-(t_{p+2} - t_{p+1})\}[\Sigma_+(p) + \Sigma_-(p)]. \tag{2.97}$$

The initial conditions are $\Sigma_+(1) = -\Sigma_-(1) = \frac{1}{2}$. Iterating equation (2.97) n times, and noticing that the rhs of (2.92) is equal to $\Sigma_+(p) + \Sigma_-(p)$, we obtain the desired formula (2.88). Q.E.D.

Theorem 5 has an important Corollary which requires the following definition:

Definition 4: A time-ordered monomial is a product $M(t_n)\cdots M(t_1)$ with $t_n \geq \cdots \geq t_1$; a time-ordered polynomial is a sum of time-ordered monomials.

Corollary of Theorem 5: Let $M(t)$ be the same parastochastic function as in Theorem 5; if $t_1 \geq t_2$ and $p(M)$ is a time-ordered polynomial in the $M(t)$'s involving only times prior to t_2, then

$$E\{M(t_1)M(t_2)p(M)\} = E\{M(t_1)M(t_2)\}E\{p(M)\}. \tag{2.98}$$

Proof: For time-ordered monomials it is an immediate consequence of (2.86); for polynomials it follows by linearity.

3. LINEAR STOCHASTIC AND PARASTOCHASTIC EQUATIONS

A. Formulation

1. *Linear Stochastic Equations*

Linear differential, partial differential, and integral equations with random coefficients (or kernels), such as arise in the study of turbulent convection, wave propagation in random media, impurity scattering, stochastic acceleration, random networks, etc., are called here "linear stochastic equations" (see Refs. 1, 5, 9–13, 24–26).

For the kind of investigations we have in mind, it is useful to assign to every linear stochastic equation a random Green's function $\mathbf{g}(\omega)$ which satisfies a random operator equation, the general form of which is

$$\mathbf{l}(\omega)\mathbf{g}(\omega) = \mathbf{u}. \tag{3.1}$$

In Eq. (3.1), ω is an element of a probability space Ω and $\mathbf{l}(\omega)$ is a given random operator in a function space F [e.g., $L_2(\mathbf{R})$], i.e., a mapping $(f, \omega) \to f'$ from $F \times \Omega$ to F, which is linear in f for fixed ω and measurable in ω for fixed f; the Green's function $\mathbf{g}(\omega)$ is also a random operator in the same function space; $\mathbf{u}$ is the identity operator in F. To avoid unnecessary complications, it will be assumed that F is a space of scalar functions of a real variable x and that $\mathbf{l}(\omega)$ and $\mathbf{g}(\omega)$ can, at least formally, be written as random integral operators with kernels $l(\omega; x, x')$ and $g(\omega; x, x')$; Eq. (3.1) then becomes a random integral equation

$$\int l(\omega; x, x')g(\omega; x', x'')\,dx' = \delta(x - x''), \tag{3.2}$$

where δ is the Dirac distribution. It will always be assumed that $\mathbf{l}(\omega)$ has finite expectation value, and therefore can be split as follows:

$$\mathbf{l}(\omega) = \mathbf{k} + \mathbf{m}(\omega), \tag{3.3}$$

where $\mathbf{k} = k(x, x')$ is the expectation value of $\mathbf{l}(\omega)$ and $\mathbf{m}(\omega) = m(\omega; x, x')$ has zero mean value. Inserting (3.3) into (3.1), we obtain

$$[\mathbf{k} + \mathbf{m}(\omega)]\mathbf{g}(\omega) = \mathbf{u}. \tag{3.4}$$

This equation will be our starting point.

To illustrate the preceding definitions we give two examples, which will again be used in the sequel.

Example 1. A randomly frequency-modulated oscillator: The amplitude $q(\omega; t)$ of a linear oscillator with a time-dependent random frequency $m(\omega; t)$ satisfies the following stochastic differential equation

$$\frac{dq(\omega; t)}{dt} = \mathrm{Im}\,(\omega; t)q(\omega; t). \tag{3.5}$$

For later convenience, $m(\omega; t)$ will be taken Gaussian with zero mean value and covariance $\Gamma(t, t')$. The Green's function $g(\omega; t, t')$ is defined by the following

[24] A. T. Bharucha-Reid, Proc. Symp. Appl. Math. **16**, 40 (1964).
[25] A. T. Bharucha-Reid, *Random Equations* (Academic Press Inc., New York, 1968).
[26] J. B. Keller, Proc. Symp. Appl. Math. **16**, 145 (1964).

conditions:

$$\left[\frac{d}{dt} - \mathrm{Im}\,(\omega; t)\right] g(\omega; t, t') = 0, \quad \text{for } t > t',$$
$$g(\omega; t', t') = 1, \qquad (3.6)$$
$$g(\omega; t, t') = 0, \quad \text{for } t < t'.$$

With the aid of Dirac distributions, Eq. (3.6) can also be written as an integral equation similar to (3.2):

$$\int_{-\infty}^{+\infty} [\delta'(t - t_1) - \mathrm{Im}\,(\omega; t)\delta(t - t_1)] g(\omega; t_1, t')\, dt_1 = \delta(t - t'). \quad (3.7)$$

Example 2. Propagation of harmonic waves in a spatially random medium: The amplitude $\Psi(\omega; x)$ of a time-harmonic wave radiated by a source with density $j(x)$ into a random medium with refractive index $n(\omega; x)$ satisfies the following stochastic partial differential equation[5]:

$$\Delta\Psi(\omega; x) + k^2 n^2(\omega; x)\Psi(\omega; x) = j(x), \quad (3.8)$$

where x is a three-dimensional variable. The Green's function $g(\omega; x, x')$ is now defined as the amplitude at x corresponding to a source $\delta(x - x')$; it satisfies

$$[\Delta_x + k^2 n_0^2(x) + k^2 m(\omega; x)] g(\omega; x, x') = \delta(x - x'), \quad (3.9)$$

where $n_0^2(x)$ and $m(\omega; x)$ denote the mean and fluctuating parts of $n^2(\omega; x)$.

To solve Eq. (3.4), we first of all show existence, uniqueness, and measurability of the solution. Depending on the specific equation under consideration, this may be a pure routine matter or a formidable task (examples 1 and 2 are good illustrations of those extreme situations). It will be assumed here that such problems have already been solved (see, e.g., Refs. 25 and 27). The practical problem is in a sense more limited; usually only the first few moments of the Green's function are wanted, but if possible in explicit form, or at least as solutions of equations that can easily be solved numerically. Here only the two first moments are considered: the mean Green's function $E\{\mathbf{g}(\omega)\}$ and its covariance

$$E\{\mathbf{g}(\omega) \otimes \mathbf{g}(\omega)\} = E\{g(\omega; x, x')g(\omega; y, y')\}.$$

The symbol $\otimes$ denotes the tensor product of two vectors or two operators in F.

Our aim will be to obtain these moments as solutions of closed nonrandom equations, also called "master equations." ("Closed" means that it can be written explicitly with a finite number of terms.) Let us make a brief survey of the difficulties of this program (see Refs. 5 and 8 for details). The central difficulty arises from the fact that, in spite of the linearity of (3.4), the Green's function $\mathbf{g}(\omega)$ has a nonlinear dependence on the random operator $\mathbf{m}(\omega)$; this implies that, even if an explicit solution is known for almost every $\omega \in \Omega$, one still has to face the very difficult problem of averaging nonlinear functionals to obtain the moments of $\mathbf{g}(\omega)$. Another way to see the difficulty is the following: Take the expectation value of (3.4) to obtain an equation relating $E\{\mathbf{g}(\omega)\}$ and $E\{\mathbf{m}(\omega)\mathbf{g}(\omega)\}$; then multiply (3.4) by $\mathbf{m}(\omega)\mathbf{k}^{-1}$ before taking the expectation value, to obtain an equation relating $E\{\mathbf{m}(\omega)\mathbf{g}(\omega)\}$ and $E\{\mathbf{m}(\omega)\mathbf{k}^{-1}\mathbf{m}(\omega)\mathbf{g}(\omega)\}$, etc.; this leads to an infinite hierarchy of equations. Closed equations for the moments of $\mathbf{g}(\omega)$ can only be obtained by making some *closure assumption* (usually quite arbitrary) on this hierarchy. It is, e.g., often assumed that

$$E\{\mathbf{m}(\omega)\mathbf{k}^{-1}\mathbf{m}(\omega)\mathbf{g}(\omega)\} \approx E\{\mathbf{m}(\omega)\mathbf{k}^{-1}\mathbf{m}(\omega)\}E\{\mathbf{g}(\omega)\}. \quad (3.10)$$

This assumption, called the "first cumulant-discard approximation,"[8] leads to the following incorrect (but quite tractable) master equation:

$$\mathbf{k}E\{\mathbf{g}(\omega)\} - E\{\mathbf{m}(\omega)\mathbf{k}^{-1}\mathbf{m}(\omega)\}E\{\mathbf{g}(\omega)\} = \mathbf{u}. \quad (3.11)$$

Kraichnan[8] has shown that (3.10) and similar closure assumptions may lead to physically meaningless results, such as negative energy densities or negative probabilities. It should finally be mentioned that the direct numerical solution of linear stochastic equations is an almost hopelessly difficult task: First, discrete random coefficients must be expressed in terms of N independent random variables; then, each of these random variables must be simulated by means of M independent trials; hence, when passing from nonrandom to random equations, the computing time is multiplied by a factor MN; of course, both M and N must be large to obtain a good accuracy.

2. Linear Parastochastic Equations

Let us write Eq. (3.4) in integral form,

$$\int [k(x, x') + m(\omega; x, x')] g(\omega; x', x'')\, dx' = \delta(x' - x''), \quad (3.12)$$

and let us assume that $m(\omega; x, x')$ is a Gaussian random function of x and x', with zero mean value and covariance:

$$E\{m(\omega; x, x')m(\omega; y, y')\} = \Gamma(x, x'; y, y'). \quad (3.13)$$

[27] Vo-Khac Khoan, Compt. Rend. **265**, 63 (1967).

In view of the equivalence proved in Sec. 2E.1 between Gaussian random functions and parastochastic functions (with $\lambda = +1$), it appears natural to associate to (3.12) the following "linear parastochastic equation":

$$\int [k(x, x')I + M(x, x')]G(x', x'')\,dx' = \delta(x' - x'')I. \tag{3.14}$$

In Eq. (3.14), $M(x, x')$ is a parastochastic function of the two-dimensional variable (x, x'), acting in the Hilbert space H, with $\lambda = +1$, and covariance $\Gamma(x, x'; y, y')$. The operator I is the identity in H, and the "parastochastic Green's function" $G(x, x')$ is, like $M(x, x')$, a two-parameter family of operators in H.

Just as the stochastic equation (3.12), the parastochastic equation (3.14) can be written in pure operator form

$$(\mathbf{kI} + \mathbf{M})\mathbf{G} = \mathbf{I}. \tag{3.15}$$

If $F = L_2(\mathbf{R})$, it is useful to think of $\mathbf{M}$, of $\mathbf{G}$, and of $\mathbf{I}$ as linear operators in the Hilbert space $L_2(\mathbf{R}; H)$ of functions $\Psi(x) \in H$ such that $\|\Psi(x)\|_H^2$ is square integrable.

The "mean parastochastic Green's function" $E\{\mathbf{G}\} = E\{G(x, x')\}$ is defined as the vacuum expectation of $\mathbf{G}$; similarly, the covariance of the parastochastic Green's function is

$$E\left\{\mathbf{G} \underset{F}{\otimes} \mathbf{G}\right\} = E\{G(x, x')G(y, y')\}, \tag{3.16}$$

where $\mathbf{G} \otimes \mathbf{G}$ is a tensor product with respect to F, but an ordinary operator product with respect to H; specifically, if $F = L_2(\mathbf{R})$, then $F \otimes F = L_2(\mathbf{R}^2)$ and $\mathbf{G} \otimes \mathbf{G}$ is the linear operator in the space $L_2(\mathbf{R}^2; H)$ whose kernel is given by the rhs of (3.16).

At least formally, the fact that any moment of $\mathbf{m}(\omega)$ is equal to the corresponding moment of $\mathbf{M}$ implies that, similarly, any moment of $\mathbf{g}(\omega)$ is equal to the corresponding moment of $\mathbf{G}$; this can be seen on the iterative solutions of Eqs. (3.4) and (3.15),

$$\mathbf{g}(\omega) = \sum_{n=0}^{\infty} [-\mathbf{k}^{-1}\mathbf{m}(\omega)]^n \mathbf{k}^{-1}, \tag{3.17}$$

$$\mathbf{G} = \sum_{n=0}^{\infty} (-\mathbf{k}^{-1}\mathbf{M})^n \mathbf{k}^{-1}, \tag{3.17'}$$

which can be used to express any moment of $\mathbf{g}(\omega)$ ($\mathbf{G}$, respectively) in terms of moments of $\mathbf{m}(\omega)$ ($\mathbf{M}$, respectively). It is not our aim to give a rigorous proof of this plausible assertion, since this would involve us into the questions of existence, uniqueness, measurability, and finiteness of moments of solutions of stochastic equations, questions which we are deliberately trying to avoid. We shall rather study the linear parastochastic equation (3.15) for its own sake, and without restricting ourselves to $\lambda = +1$.

B. Linear and Nonlinear Master Equations

In this section, closed master equations are derived for mean parastochastic Green's functions. It is assumed that Eq. (3.15) has a unique solution,

$$\mathbf{G}(\mathbf{k}, \mathbf{M}) = (\mathbf{kI} + \mathbf{M})^{-1}, \tag{3.18}$$

which is both right- and left-hand inverse of $\mathbf{kI} + \mathbf{M}$. It is also assumed that the functional dependence of $\mathbf{G}$ on $\mathbf{M}$ is gentle enough to ensure that Theorems 2 and 3 are still valid when polynomials in the $M(x, x')$'s are replaced by $\mathbf{G}(\mathbf{k}, \mathbf{M})$. As noticed in Remark 3, this will be the case if the operators $M(x, x')$ are bounded ($\lambda = 0$ or -1) and if $\mathbf{G}$ is an analytic functional of $\mathbf{M}$. The question of gentleness, being intimately connected with the problems of existence and uniqueness, will not be tackled here, except in the special case of Example 1 of Sec. 3A.1, for which gentleness will be shown in Sec. 3D.

As a preliminary result, we establish the following "pre-master equation," valid for any value of λ,

$$\mathbf{k}E\{\mathbf{G}\} - E\{\mathbf{M}'\mathbf{G}\mathbf{M}'\mathbf{G}\} = \mathbf{I}, \tag{3.19}$$

where $\mathbf{M}$ and $\mathbf{M}'$ are uncorrelated versions of the same function in the sense of Definition 2.

Proof of the pre-master equation: Taking the vacuum expectation of (3.15) and using Theorem 2 and Remark 3, we obtain

$$\mathbf{k}E\{\mathbf{G}\} + E\left\{\mathbf{M}' \frac{d}{d\epsilon} \mathbf{G}(\mathbf{M} + \epsilon\mathbf{M}')\right\}_{\epsilon=0} = \mathbf{u}. \tag{3.20}$$

Here $\mathbf{G}(\mathbf{M} + \epsilon\mathbf{M}')$ is, by definition, the solution of Eq. (3.15) with $\mathbf{M}$ replaced by $\mathbf{M} + \epsilon\mathbf{M}'$, which reads

$$\{\mathbf{kI} + (\mathbf{M} + \epsilon\mathbf{M}')\}\mathbf{G}(\mathbf{M} + \epsilon\mathbf{M}') = \mathbf{I}. \tag{3.21}$$

Taking the derivative of this equation with respect to ϵ, for $\epsilon = 0$, we have

$$(\mathbf{kI} + \epsilon\mathbf{M}) \frac{d}{d\epsilon} \mathbf{G}(\mathbf{M} + \epsilon\mathbf{M}')_{\epsilon=0} = -\mathbf{M}'\mathbf{G}(\mathbf{M}). \tag{3.22}$$

Then, using (3.18),

$$\begin{aligned} \frac{d}{d\epsilon} \mathbf{G}(\mathbf{M} + \epsilon\mathbf{M}')_{\epsilon=0} &= -(\mathbf{kI} + \mathbf{M})^{-1}\mathbf{M}'\mathbf{G}(\mathbf{M}) \\ &= -\mathbf{G}(\mathbf{M})\mathbf{M}'\mathbf{G}(\mathbf{M}). \end{aligned} \tag{3.23}$$

Inserting (3.23) into (3.20), we obtain the pre-master equation (3.19).

The three special cases $\lambda = +1, 0$, and -1 give rise to quite different master equations, and will be examined in Secs. 3B.1–3B.3 which can be read independently.

1. $\lambda = +1$. *Equations with Gaussian Coefficients*

It is now shown that for $\lambda = +1$, the mean parastochastic Green's function satisfies a linear functional differential equation. In the parastochastic equation (3.15), $\mathbf{k}$ operates only on the function space F; it is convenient for a while to replace it by an operator which, like $\mathbf{M}$, acts also on H. Assuming that (3.18) still holds, we have

$$\mathbf{G}(\mathbf{k} + \epsilon\mathbf{M}', \mathbf{M}) = \mathbf{G}(k, \mathbf{M} + \epsilon\mathbf{M}'); \quad (3.24)$$

hence,

$$\frac{d}{d\epsilon}\mathbf{G}(\mathbf{k}, \mathbf{M} + \epsilon\mathbf{M}')_{\epsilon=0} = \frac{d}{d\epsilon}\mathbf{G}(\mathbf{k} + \epsilon\mathbf{M}', \mathbf{M})_{\epsilon=0}. \quad (3.25)$$

The rhs of (3.25), being linear in $\mathbf{M}'$, can be written in terms of the functional derivative $\delta\mathbf{G}/\delta\mathbf{k}$ as

$$\frac{d}{d\epsilon}\mathbf{G}(\mathbf{k} + \epsilon\mathbf{M}', \mathbf{M})_{\epsilon=0} = \left\langle \frac{\delta\mathbf{G}}{\delta\mathbf{k}}, \mathbf{M}' \right\rangle. \quad (3.26)$$

Inserting (3.25) into (3.20), and using (3.26), we have

$$\mathbf{k}E\{\mathbf{G}\} + E\left\{\mathbf{M}'\left\langle \frac{\delta\mathbf{G}}{\delta\mathbf{k}}, \mathbf{M}' \right\rangle\right\} = \mathbf{u}. \quad (3.27)$$

For $\lambda = +1$, the uncorrelated parastochastic functions $M(x, x')$ and $M'(x, x')$ can be represented by Gaussian random functions $m(\omega; x, x')$ and $m'(\omega; x, x')$ (see Sec. 2E.1), which are, of course, also uncorrelated in the usual sense.[6] From the preceding section, it is known that for Gaussian $\mathbf{m}(\omega)$, the mean stochastic Green's function $E\{\mathbf{g}(\omega)\}$ for Eq. (3.4) is equal to the mean parastochastic Green's function $E\{\mathbf{G}\}$ for Eq. (3.15). Equation (3.27) can therefore be written in terms of ordinary stochastic quantities; in explicit integral form (3.27) then becomes

$$\int k(x, 1)E\{g(\omega; 1, x')\}\, d1 + \int E\left\{m'(\omega; x, 1)\frac{\delta g(\omega; 1, x')}{\delta k(2, 3)} m'(\omega; 2, 3)\right\} d1\, d2\, d3 = \delta(x - x'), \quad (3.28)$$

where, for notational convenience, 1, 2, 3 stand for x_1, x_2, x_3. It is known that two uncorrelated Gaussian random functions are also independent.[15] As $\mathbf{g}(\omega)$ depends upon $\mathbf{m}(\omega)$ but not upon $\mathbf{m}'(\omega)$, the expectation values over $\mathbf{m}(\omega)$ and $\mathbf{m}'(\omega)$ can be taken independently; Eq. (3.28) then becomes

$$\int k(x, 1)E\{g(\omega; 1, x')\}\, d1 + \int \frac{\delta E\{g(\omega; 1, x')\}}{\delta k(2, 3)} \Gamma(x, 1; 2, 3)\, d1\, d2\, d3 = \delta(x - x'), \quad (3.29)$$

where Γ is defined by (3.13).

This linear functional differential master equation may have some interesting theoretical applications, but it is probably illadapted for numerical studies. It is unlikely that any more tractable exact master equation can be found in the Gaussian case ($\lambda = +1$).

Remark 7: A more standard "stochastic" derivation of the master equation (3.29) can be based on the following result, a special case of Theorem 2: If $\{m_i(\omega)\}$ is a set of correlated Gaussian random variables with zero mean value and covariance Γ_{ij}, and if the function $f(m_1, \cdots, m_n)$ is differentiable, then

$$E\{m_j(\omega)f(m_1(\omega), \cdots, m_n(\omega))\} = \sum_k \Gamma_{jk}E\left\{\frac{\partial f}{\partial m_k}\right\}. \quad (3.30)$$

Equation (3.30) is easily proved by writing the expectation values in terms of the n-variate Gaussian probability density.

2. $\lambda = 0$. *The Kraichnan Equation*

In integral form the pre-master equation (3.19) reads

$$\int k(x, 1)E\{G(1, x')\}\, d1 - \int E\{M'(x, 1)G(1, 2)M'(2, 3)G(3, x')\}\, d1\, d2\, d3 = \delta(x - x'). \quad (3.31)$$

As $\lambda = 0$, Theorem 3 and Remark 4 can be applied to the second term on the lhs to give

$$\int k(x, 1)E\{G(1, x')\}\, d1 - \int \Gamma(x, 1; 2, 3)E\{G(1, 2)\}E\{G(3, x')\}\, d1\, d2\, d3 = \delta(x - x'), \quad (3.32)$$

where Γ is defined by (3.13). In operator form, (3.32) reads

$$\mathbf{k}E\{\mathbf{G}\} - E\{\mathbf{M}E\{\mathbf{G}\}\mathbf{M}\}E\{\mathbf{G}\} = \mathbf{u}. \quad (3.32')$$

Using the fact that $\mathbf{G}$ is also the left-hand inverse of $\mathbf{kI} + \mathbf{M}$, it can be shown along the same lines that

$E\{\mathbf{G}\}$ satisfies also

$$E\{\mathbf{G}\}\mathbf{k} - E\{\mathbf{G}\}E\{\mathbf{M}E\{\mathbf{G}\}\mathbf{M}\} = \mathbf{u}. \quad (3.32'')$$

Equation (3.32), which was first obtained by Kraichnan,[13] will be called the Kraichnan equation. The relation of our approach to Kraichnan's original derivation and an interpretation of this equation in terms of Wigner matrices may be found in Sec. 3C. Notice that the Kraichnan equation has a quadratic nonlinearity; it may seem strange that the mean Green's function of a linear problem satisfies a nonlinear equation; it should, however, be noticed that if (3.15) is linear, it is not homogeneous.

Let us also derive a master equation for the covariance of the parastochastic Green's function. Multiplying (3.15) by $\otimes_F \mathbf{G}$, taking the vacuum expectation, and using successively Theorem 2, Eq. (3.23), and Theorem 3, we obtain

$$\begin{aligned} E\Big\{(\mathbf{k} + \mathbf{M})\mathbf{G} \underset{F}{\otimes} \mathbf{G}\Big\} &= \mathbf{k}E\Big\{\mathbf{G} \underset{F}{\otimes} \mathbf{G}\Big\} - E\Big\{\mathbf{M}'\mathbf{G}\mathbf{M}'\mathbf{G} \underset{F}{\otimes} \mathbf{G}\Big\} \\ &\quad - E\Big\{\mathbf{M}'\mathbf{G} \underset{F}{\otimes} \mathbf{G}\mathbf{M}'\mathbf{G}\Big\} \\ &= \mathbf{k}E\Big\{\mathbf{G} \underset{F}{\otimes} \mathbf{G}\Big\} - E\{\mathbf{M}'E\{\mathbf{G}\}\mathbf{M}'\}E\Big\{\mathbf{G} \underset{F}{\otimes} \mathbf{G}\Big\} \\ &\quad - E\Big\{\mathbf{M}'E\Big\{\mathbf{G} \underset{F}{\otimes} \mathbf{G}\Big\}\mathbf{M}'\Big\}E\{\mathbf{G}\} \\ &= \mathbf{u} \underset{F}{\otimes} E\{\mathbf{G}\}. \end{aligned} \quad (3.33)$$

Multiplying (3.33) by $E\{\mathbf{G}\}$ on the left, and using (3.32''), we finally have

$$E\Big\{\mathbf{G} \underset{F}{\otimes} \mathbf{G}\Big\} = \Big[E\{\mathbf{G}\} \underset{F}{\otimes} E\{\mathbf{G}\}\Big] + E\{\mathbf{G}\}E\Big\{\mathbf{M}E\Big\{\mathbf{G} \underset{F}{\otimes} \mathbf{G}\Big\}\mathbf{M}\Big\}E\{\mathbf{G}\}. \quad (3.34)$$

The corresponding integral form is

$$\begin{aligned} &E\{G(x, x')G(y, y')\} \\ &\quad = E\{G(x, x')\}E\{G(y, y')\} \\ &\qquad + \int E\{G(x, 1)\}\Gamma(1, 2; 3, 4)E\{G(2, x')G(y, 3)\} \\ &\qquad \times E\{G(4, y')\}\, d1\, d2\, d3\, d4. \end{aligned} \quad (3.34')$$

It should be noticed that, once the mean Green's function is known, (3.34) is a linear equation for the covariance; similar equations can be obtained for higher-order moments of $\mathbf{G}$.

Remark 8: An alternative derivation of Eqs. (3.32) and (3.34) has been given by Frisch,[5] in which Theorem 3 plays no part; instead Remark 1 was applied to the diagrammatic representation of the vacuum expectation of the iterative solution (3.17') of (3.15).

We finally illustrate the Kraichnan equation on Examples 1 and 2 of the preceding section.

Example 1: The parastochastic analog of (3.6) is

$$\begin{aligned} \left[\frac{d}{dt} - iM(t)\right]G(t, t') &= 0, \quad \text{for } t > t', \\ G(t, t) &= 1, \\ G(t, t') &= 0, \quad \text{for } t < t'. \end{aligned} \quad (3.35)$$

The Kraichnan equation then reads

$$\begin{aligned} &\frac{d}{dt}E\{G(t, t')\} \\ &+ \int_{t'}^{t} E\{G(t, t_1)\}\Gamma(t, t_1)E\{G(t_1, t')\}\, dt_1 = 0, \quad \text{for } t > t', \\ &E\{G(t, t)\} = 1, \\ &E\{G(t, t')\} = 0, \quad \text{for } t < t', \end{aligned} \quad (3.36)$$

with $\Gamma(t, t') = E\{M(t)M(t')\}$.

Example 2: The parastochastic analog of (3.9) is

$$[\Delta + k^2n_0^2(x) + k^2M(x)]G(x, x') = \delta(x - x'), \quad (3.37)$$

and the Kraichnan equation reads

$$\begin{aligned} &[\Delta + k^2n_0^2(x)]E\{G(x, x')\} \\ &\quad - k^4\int E\{G(x, 1)\}\Gamma(x, 1)E\{G(1, x')\}\, d1 \\ &\quad = \delta(x - x'). \end{aligned} \quad (3.38)$$

3. $\lambda = -1$. *Dichotomic Markov Process and First Cumulant-Discard Approximation*

A very simple master equation is now derived for the linear stochastic operational differential equation

$$\begin{aligned} \frac{d}{dt}y(\omega; t) &= Cy(\omega; t) + m(\omega; t)Dy(\omega; t), \\ y(\omega; t) &\in \mathfrak{B}, \end{aligned} \quad (3.39)$$

where C and D are bounded operators in a Banach space $\mathfrak{B}$, and where $m(\omega; t)$ is the dichotomic Markov process of Sec. 2.E.3. A simple calculation shows that the Green's function of Eq. (3.39) satisfies the stochastic integral equation

$$\begin{aligned} &g(\omega; t, t') \\ &\quad = \exp[C(t - t')] \\ &\qquad + \int_{t'}^{t} \exp[C(t - t'')]m(\omega; t'')Dg(\omega; t'', t')\, dt''. \end{aligned} \quad (3.40)$$

The boundedness of C, D, and $m(\omega; t)$ implies uniform normal convergence in any bounded time-interval of the iterative solution of Eq. (3.40). It follows that the expectation value of $g(\omega; t, t')$ can be taken term by term. In this way, a uniformly convergent series is obtained for $E\{g(\omega; t, t')\}$ which obviously involves integrals over *time-ordered* moments of $m(\omega; t)$ only. Therefore, by Theorem 5, Eq. (3.40) is equivalent to the parastochastic equation

$$G(t, t') = \exp[C(t - t')] + \int_{t'}^{t} \exp[C(t - t'')]M(t'')DG(t'', t')\,dt'', \tag{3.41}$$

where $M(t)$ is a parastochastic function with $\lambda = -1$ and exponential covariance. After elimination of $E\{M(t)G(t, t')\}$ between the vacuum expectation of (3.41) and the vacuum expectation of (3.41), which has first been multiplied by $M(t)$, the following equation is obtained:

$$E\{G(t, t_1)\} = \exp[C(t - t_1)] + \int_{t'}^{t}\int_{t_1}^{t'} \exp[C(t - t'')] \times D\exp[C(t'' - t')]DE\{M(t'')M(t')G(t', t_1)\}\,dt'\,dt'', \tag{3.42}$$

where use was made of the fact that $M(t)$ commutes with C and D, since it acts like a scalar on $\mathcal{B}$.

The crucial step in deriving the master equation is to apply the Corollary of Theorem 5 to (3.42) in the following form:

$$E\{M(t'')M(t')G(t', t_1)\} = E\{M(t'')M(t')\}E\{G(t', t_1)\} = \exp[-|t'' - t'|]E\{G(t', t_1)\}, \tag{3.43}$$

which is easily justified term by term on the iterative solution of (3.41). The resulting equation, being translation invariant, we put as

$$E\{G(t, t')\} = E\{G(t - t')\}. \tag{3.44}$$

Taking the time-derivative of (3.42), and using (3.43) and (3.44), we finally obtain the master equation:

$$\frac{d}{dt}E\{G(t)\} = CE\{G(t)\} + \int_0^t D\exp[C(t - t')]D \times \exp[-(t - t')]E\{G(t')\}\,dt', \tag{3.45}$$
$$E\{G(0)\} = U,$$

where U is the identity in $\mathcal{B}$. A direct, more traditional derivation of this equation may be based on the fact that (3.39) is an equation with Markov coefficients.[5,28]

[28] U. Frisch, Compt. Rend. **262**, 762 (1966).

The master equation (3.45) can be explicitly solved by Laplace transformation and will therefore be successfully applied to most one-parameter linear stochastic problems; the parameter will usually be the time (e.g., random electrical networks), but may also be a space-variable (e.g., vibration of rods with random mass densities or cross sections). Actually, Eq. (3.45) has already been proposed several times as an *approximation* under various names, such as one-fiction approximation or first-order smoothing approximation.[5,9,26] The reason for this is probably the rather curious observation that the first cumulant-discard approximation is *exact* for Eq. (3.39), as can be seen by comparing (3.43) and (3.10). It may be asked, in view of this result, if it appears reasonable to apply the first cumulant-discard approximation to more general linear stochastic equations than (3.39). This question has been discussed in Ref. 5, where arguments are given which suggest that the first cumulant-discard approximation constitutes a uniformly valid approximation when a certain "generalized Reynolds number," a measure of the importance of randomness, is small; however, no rigorous result of sufficient generality has yet been obtained.

C. Kraichnan's Random-Coupling Model and Wigner Matrices

From Theorem 4, which asserts that ($\lambda = 0$)-parastochastic functions can be represented by Wigner matrices, it may be inferred that for $\lambda = 0$, the parastochastic equation (3.15) is equivalent to the stochastic matrix equation

$$\sum_{\beta=1}^{N} [\mathbf{k}\delta_{\alpha\beta} + N^{-\frac{1}{2}}\mathcal{M}_{\alpha\beta}^{(N)}(\omega)]\mathfrak{G}_{\beta\gamma}^{(N)}(\omega) = \delta_{\alpha\gamma}\mathbf{u}, \tag{3.46}$$

where $\mathcal{M}_{\alpha\beta}(\omega)$ is a Wigner matrix (see Def. 3), whose elements are random functions $m(\omega; x, x')$ with zero mean value and covariance $\Gamma(x, x'; y, y')$; more specifically,

$$\text{m.s.}\lim_{N\to\infty} \mathfrak{G}_{\alpha\beta}^{(N)}(\omega) = \delta_{\alpha\beta}E\{\mathbf{G}\} \tag{3.47}$$

and

$$\text{m.s.}\lim_{N\to\infty} \sum_{\beta=1}^{N} \mathfrak{G}_{\alpha\beta}^{(N)}(\omega) \underset{F}{\otimes} \mathfrak{G}_{\beta\gamma}^{(N)}(\omega) = \delta_{\alpha\beta}E\left\{\mathbf{G} \underset{F}{\otimes} \mathbf{G}\right\}. \tag{3.48}$$

Despite this rigorous derivation, which is connected with existence and uniqueness problems, it will be enough to notice that Eqs. (3.47) and (3.48) can be justified, at least formally, by applying Theorem 4 term by term to the iterative solutions of (3.15) and (3.46). We shall say that Eq. (3.46) constitutes a stochastic model for the parastochastic equation (3.15); following Kraichnan,[13] we call it the *random-coupling model*. The word "coupling" refers to the

off-diagonal elements of the Wigner matrix; if this matrix were to become purely diagonal, Eq. (3.46) would reduce to N independent versions of the ordinary stochastic equation (3.4). It is rather surprising that the additional coupling produces actually a great deal of simplification.

Before discussing the implications of the random-coupling model, we make a few comments on Kraichnan's original derivation. The Kraichnan equation (3.32) was first obtained, not from the random-coupling model (3.46), but directly from the ordinary stochastic equation (3.4) on the basis of a closure assumption, called the *direct interaction approximation*[29]; later on it was found, and this is a most remarkable achievement, that the resulting equation is also the *exact* solution of a model.[13,30] Kraichnan's formulation of the random-coupling model is somewhat different from ours, since it involves a Fourier transformation on a discrete group; however, we believe that it is simpler to work directly with Wigner matrices. Kraichnan's derivation of (3.32) from the random-coupling model makes use of the following:

(i) a variational procedure, somewhat similar to our introduction of an uncorrelated parastochastic function $M'(t)$;

(ii) a sharpness result similar to Theorem 3;

(iii) an all-order perturbation expansion of the Green's function with a diagrammatic representation of the various terms.[13]

The reason why it was possible to give a proof of the Kraichnan equation without recourse to perturbation expansions may be seen in the existence of Theorem 2, which has no counterpart in the Kraichnan theory.

It may now be asked what is the significance of the random-coupling model. In mathematical physics, models are mostly used to get some qualitative insight into seemingly insoluble problems. To obtain a tractable problem, more or less realistic simplifying assumptions are made, involving usually a reduction of the number of dimensions (most exactly-soluble problems are one-dimensional; see, e.g., Ref. 31) and a special choice of the interaction laws (potentials, covariances, etc.). This picture does not at all fit the random-coupling model, for the model and the original underlying stochastic problem have the same dimensionality and the same covariances of the coefficients. The appearance of simple master equations is merely a consequence of the change from scalar stochastic quantities to Wigner matrices or, equivalently, from $(\lambda = 1)$-parastochastics to $(\lambda = 0)$-parastochastics. It is true that the original stochastic equation also yields a closed master equation (3.29); however, this functional differential equation is of little practical use, whereas the Kraichnan equation is well-suited for both theoretical and numerical studies (this will be illustrated in Sec. 3D). The price which has to be paid for this is that the random-coupling model is probably too artificial to describe any *natural* system; we did not write any *real* system, since, in principle, nothing prevents us from constructing such a system. Briefly stated, the random-coupling model is a mathematical rather than physical model. Yet, following Kraichnan,[8,13,32] it is now shown that there are important qualitative and quantitative contacts—between the ordinary stochastic equation (3.4) and the random-coupling model (3.46) or, equivalently, the parastochastic equation (3.15)—which are expressed by the following consistency and approximation properties.

1. *Consistency Properties*

Any integral of motion of the original problem (e.g., an energy integral) will usually survive in the random-coupling model. This is a consequence of the fact that parastochastic operators are self-adjoint and, hence, behave roughly as real numbers. For example, the randomly frequency-modulated oscillator (3.5) has the energy integral

$$\overline{q(\omega; t)}q(\omega; t) = \text{const.} \tag{3.49}$$

The parastochastic analog of (3.5) is

$$\frac{d}{dt}Q(t) = iM(t)Q(t), \tag{3.50}$$

and denoting by $Q^*(t)$ the adjoint of $Q(t)$, we have

$$\frac{d}{dt}(Q^*Q) = Q^*iMQ + (iMQ)^*Q = 0; \tag{3.51}$$

hence, Q^*Q is also a constant. For turbulent convection, wave propagation in random media, etc., *energy spectra* are of fundamental importance. An energy spectrum is defined, for a stationary problem, as the Fourier transform of the covariance of the solution. As a consequence of Bochner's Theorem, it is always a nonnegative function. To show that this property survives in the model, we prove, more

[29] R. H. Kraichnan, Phys. Rev. **109**, 1407 (1958); J. Fluid Mech. **5**, 497 (1959).

[30] R. H. Kraichnan, *Second Symposium on Naval Hydrodynamics*, R. Cooper, Ed. (U.S. Government Printing Office, Washington, D.C., 1960).

[31] E. H. Lieb and D. C. Mattis, "Mathematical Physics in One Dimension," *Perspectives in Physics* (Academic Press Inc., New York, 1966).

[32] R. H. Kraichnan, *Symposium on the Dynamics of Fluids and Plasmas*, S. I. Pai, Ed. (Academic Press Inc., New York, 1966).

generally, that the parastochastic covariance

$$E\{G(x, x')G^*(y, y')\}$$

is of positive type, i.e., that for any $x_1, \cdots, x_n$; $y_1, \cdots, y_n$ and any complex numbers $c_1, \cdots, c_n$,

$$\sum_{i,j=1}^{n} c_i \bar{c}_j (\Psi_0, G(x_i, y_i)G^*(x_j, y_j)\Psi_0) \geq 0. \quad (3.52)$$

The left-hand side can be written as

$$\left(\sum_{i=1}^{n} \bar{c}_i G^*(x_i, y_i)\Psi_0, \sum_{j=1}^{n} \bar{c}_j G^*(x_j, y_j)\Psi_0\right) \cdot = \left\| \sum_{i=1}^{n} \bar{c}_i G^*(x_i, y_i)\Psi_0 \right\|^2 \geq 0, \quad (3.53)$$

which proves the assertion.

Finally, it should be mentioned that the random-coupling model does not necessarily preserve realizability conditions which are of nonprobabilistic nature. In a recent paper, Orszag and Kraichnan[33] studied the problem of stochastic acceleration by means of the random Liouville equation

$$\frac{\partial f}{\partial t} + v_i \cdot \frac{\partial f}{\partial x_i} + \frac{q}{m} \cdot E_i \cdot \frac{\partial f}{\partial v_i} = 0, \quad i = 1, 2, 3, \quad (3.54)$$

where E_i is a given random electric field. For the case of a time- and space-independent electric field with three independent random components, they find that the random-coupling model predicts some negative values for $E\{f\}$, a fact which is inconsistent with the interpretation of f as a distribution function.

2. *Approximation Properties*

In Sec. 3D, the mean Green's function for the randomly frequency modulated oscillator (example 1 of Sec. 3A) and for its random-coupling model will be compared numerically. An excellent agreement will be found there, even in the asymptotic range, as long as the generalized Reynolds number R_0 (to be defined there) is small compared to one.

More generally, if the parastochastic equation (3.15) is written

$$(\mathbf{kI} + R_0\mathbf{M})\mathbf{G} = \mathbf{I}, \quad (3.55)$$

it is conjectured (by us and also by Kraichnan[8]) that, as R_0 goes to zero, the exact "Gaussian" mean Green's function ($\lambda = +1$) approaches (in a sense not yet specified) the mean Green's function for the random-coupling model ($\lambda = 0$). A more precise conjecture will be made in the concluding section. To support this conjecture, it can be shown that (i) the perturbation expansions of both mean Green's functions in powers of R_0 agree up to second order, and (ii) if Theorem 1 is applied to evaluate perturbation terms of arbitrary order, the random-coupling model mean Green's function retains only those contributions to the exact mean Green's function arising from non-crossing pairings. Since the first cumulant-discard approximation, which corresponds to the even more restrictive class of disjoint pairings, seems to behave well for small R_0 (see Ref. 5), the random-coupling model should, at least, be an improvement on this approximation.

A surprising feature of the random-coupling model is that, even for $R_0 \gg 1$, it seems to give correct predictions for such over-all quantities as relaxation times; for example, if in (3.5) $m(\omega; t)$ is assumed to be time-independent with variance σ^2, the exact and model mean Green's functions are $\exp(-\frac{1}{2}\sigma^2 t^2)$ and $J_1(2\sigma t)/t$, respectively; in both cases the relaxation time is of the order of σ^{-1}. In contrast to this, the first cumulant-discard approximation yields the mean Green's function $\cos \sigma t$, which does not show any damping at all (see Refs. 5 and 13).

D. Nonlinear Integral Equation

This section is devoted to the following special case of the Kraichnan equation,

$$\frac{dy(t)}{dt} = -\int_0^t \Gamma(t - t')y(t - t')y(t')\,dt', \quad y(0) = 1, \quad (3.56)$$

which is obtained from (3.36) by assuming that the real covariance Γ is stationary and by writing

$$E\{G(t, t')\} = y(t - t').$$

Moreover, it will be assumed that $\Gamma(t)$

(i) is continuous,

(ii) is absolutely integrable, i.e.,

$$\int_R |\Gamma(t)|\,dt < +\infty, \quad (3.57)$$

(iii) has a Fourier transform

$$\hat{\Gamma}(\nu) = \frac{1}{2\pi} \int_R e^{-it\nu} \cdot \Gamma(t)\,dt, \quad (3.58)$$

which is not merely nonnegative (Bochner's Theorem), but strictly positive for any real ν.

Examples of such covariances are e^{-t^2}, $e^{-|t|}$, and $(t^2 + 1)^{-1}$.

It will be shown that (3.56) has a solution which is unique, real, even, bounded by one, of positive type, and which goes to zero as t goes to infinity. Except uniqueness, we have not been able to obtain any of the

[33] S. A. Orszag and R. H. Kraichnan, Phys. Fluids **10**, 1720 (1967).

preceding results directly from the *nonlinear* equation (3.56). Our somewhat unusual proofs will be based on the important property of (3.56), namely, that it is satisfied by the vacuum expectation $E\{Q(t)\}$ of the solution of the *linear* parastochastic equation (see Example 1 of Sec. 3B):

$$\frac{dQ(t)}{dt} = iM(t)Q(t), \quad Q(0) = I, \tag{3.59}$$

where $M(t)$ is a ($\lambda = 0$) parastochastic function of the real variable t with stationary covariance $\Gamma(t - t')$.

It was shown in Sec. 2D that the operators $M(t)$ are bounded. In the stationary case, the estimate (2.59) reduces to

$$\|M(t)\| \leq 2\Gamma^{\frac{1}{2}}(0). \tag{3.60}$$

Furthermore, it follows from (i) that $B(t)$, and thus its adjoint $A(t)$, as well as $M(t)$, are continuous for the operator norm.

Proof: Continuity is implied by the following relation:

$$\begin{aligned}\|(B(t+h) - B(t))\Psi\|^2 &= ([B(t+h) - B(t)]\Psi, \\ &\quad \times [B(t+h) - B(t)]\Psi) \\ &= (A(t+h)B(t+h)\Psi, \Psi) \\ &\quad + (A(t)B(t)\Psi, \Psi) \\ &\quad - (A(t+h)B(t)\Psi, \Psi) \\ &\quad - (A(t)B(t+h)\Psi, \Psi) \\ &= 2[\Gamma(0) - \Gamma(0)]\,\|\Psi\|^2,\end{aligned} \tag{3.61}$$

where (2.1) has been used with $\lambda = 0$.

In deriving the Kraichnan equation in Sec. 3B, existence and uniqueness for linear parastochastic equations were presupposed. In the present case, uniqueness results from the existence of the *energy integral* [see (3.51)]

$$Q^*(t)Q(t) = Q^*(0)Q(0). \tag{3.62}$$

The continuity of $M(t)$ implies that (3.59) has a solution, which reads

$$\begin{aligned}Q(t) = I &+ i\int_0^t M(t_1)\,dt_1 \\ &+ i^2\int_0^t\int_0^{t_1} M(t_1)M(t_2)\,dt_1\,dt_2 + \cdots.\end{aligned} \tag{3.63}$$

The uniform normal convergence in any integral $(0, T)$ of this series is proved by applying (3.60) to estimate (3.63) term by term, i.e.,

$$\begin{aligned}\|Q(t) &\leq 1 + 2\Gamma^{\frac{1}{2}}(0)T + \cdots + [2\Gamma^{\frac{1}{2}}(0)T]^n/n! + \cdots \\ &= \exp\,[2\Gamma^{\frac{1}{2}}(0)T].\end{aligned} \tag{3.64}$$

This estimate then shows that $Q(t)$ is an analytic function of $M(s)$ for $0 \leq s \leq t$, and this justifies the extensions of Theorems 2 and 3 from polynomials in $M(s)$ to $Q(t)$—extensions used in Sec. 3B to derive the pre-master and master equations. After these preliminaries, we turn back to the nonlinear equation (3.56).

Uniqueness. Uniqueness for (3.56) does not follow from uniqueness for the parastochastic equation (3.59), but requires a separate proof. To show uniqueness in any interval $(0, T)$, let (3.56) be written in integral form

$$y(t) = 1 - \int_0^t\int_0^{t'} \Gamma(t' - t'')y(t' - t'')y(t'')\,dt'\,dt'', \tag{3.65}$$

and let $y_1(t)$ and $y_2(t)$ be two continuous solutions of (3.65); their difference $z(t) = y_1(t) - y_2(t)$ satisfies

$$\begin{aligned}z(t) &= -\tfrac{1}{2}\int_0^t\int_0^{t'} \Gamma(t' - t'')\{[y_1(t' - t'') \\ &\quad + y_2(t' - t'')]z(t'') \\ &\quad + [y_1(t'') + y_2(t'')]z(t' - t'')\}\,dt'\,dt'' \\ &= -\tfrac{1}{2}\int_0^t\int_0^{t'} [\Gamma(t' - t'') + \Gamma(t'')] \\ &\quad \times [y_1(t' - t'')y_2(t' - t'')]z(t'')\,dt'\,dt''.\end{aligned} \tag{3.66}$$

Now let

$$\begin{aligned}C = \sup_{0 \leq t'' \leq t' \leq T} &|[\Gamma(t' - t'') + \Gamma(t'')] \\ &\times [y_1(t' - t'') + y_2(t' - t'')]|, \\ D = \sup_{0 \leq t \leq T} &|z(t)|.\end{aligned} \tag{3.67}$$

Using (3.67) to estimate (3.66), we find

$$|z(t)| \leq \tfrac{1}{2}DCt^2/2!, \quad \text{for} \quad 0 \leq t \leq T. \tag{3.68}$$

Using (3.68), we obtain a second estimate for $\overline{z(t)}$ involving t^4; repeating this process n times, we have

$$\begin{aligned}|z(t)| \leq \tfrac{1}{2}D^nCt^{2n}/(2n)! \leq \tfrac{1}{2}D^nCT^{2n}/(2n)!, \\ \text{for} \quad 0 \leq t \leq T,\end{aligned} \tag{3.69}$$

and it follows that $z(t) = b$, which proves uniqueness. To show that $y(t)$ is real and even, it suffices to check that $\overline{y(t)}$ and $y(-t)$ also satisfy (3.56), and to use uniqueness.

Existence, boundedness, and positive type: Existence follows from

$$y(t) = E\{Q(t)\} = (\Psi_0, Q(t)\Psi_0), \tag{3.70}$$

where $Q(t)$ is the unique solution of the linear parastochastic equation (3.59). Then, using the Schwarz inequality and (3.62), we have

$$\begin{aligned} |y(t)| = |(\Psi_0, Q(t)\Psi_0)| &\leq \|Q(t)\Psi_0\| \\ &= (Q(t)\Psi_0, Q(t)\Psi_0)^{\frac{1}{2}} \\ &= (Q^*(t)Q(t)\Psi_0, \Psi_0)^{\frac{1}{2}} \\ &= (\Psi_0, \Psi_0)^{\frac{1}{2}} \\ &= 1. \end{aligned} \tag{3.71}$$

Therefore $|y(t)|$ is bounded by one.

At this point the explicit realization, given in Sec. 2D, of the parastochastic function $M(t)$ is needed. Recall that

$$M(t) = B(t) + B^*(t) = x(t) \otimes + (x(t)\otimes)^*, \tag{3.72}$$

where $x(t)$ is given by (2.5). Let N denote the self-adjoint operator

$$N\colon\ h(v) \to -vh(v) \tag{3.73}$$

acting in the Hilbert space $\mathcal{H} = L_2(\hat{\Gamma}(v)\,dv)$, and let H_0 denote the extension of N to the tensor algebra (2.51)—an extension which is also self-adjoint and reduces to zero on C and to $N \otimes N \otimes \cdots \otimes N$, (taken p times) on $\mathcal{H}^{\otimes p}$. Equation (2.50) may be written as

$$x(t) = e^{-itN}x(0). \tag{3.74}$$

Then an easy calculation, based on (3.72), shows that

$$M(t) = e^{-itH_0}M(0)e^{itH_0}. \tag{3.75}$$

Performing in (3.59) the change of variable

$$Q(t) = e^{-itH_0}S(t), \tag{3.76}$$

and using (3.75), we obtain for $S(t)$ the following equation:

$$\frac{dS(t)}{dt} = i[H_0 + M(0)]S(t), \quad S(0) = I. \tag{3.77}$$

In terms of $S(t)$, the solution of (3.56) reads

$$\begin{aligned} y(t) &= (\Psi_0, e^{-itH_0}S(t)\Psi_0) = (e^{itH_0}\Psi_0, S(t)\Psi_0) \\ &= (\Psi_0, S(t)\Psi_0) = E\{S(t)\}, \end{aligned} \tag{3.78}$$

where we used $H_0\Psi_0 = 0$.

Equation (3.77) is reminiscent of the Schrödinger equation for a quantum-mechanical system with total Hamiltonian $H_T = H_0 + M(0)$. Equation (3.59) then corresponds to the "interaction representation." The interest of (3.77) is that it is an equation with constant coefficients for which semigroup and spectral techniques are available. The operator iH_0 generates the strongly continuous semigroup e^{itH_0}, and the operator $iM(0)$ is bounded; therefore, by a well-known perturbation theorem,[34] iH_T also generates a strongly continuous semigroup

$$S(t) = e^{itH_T}. \tag{3.79}$$

With the operator H_T being self-adjoint, this semigroup is unitary.[34] The following relations, together with (3.78) and (3.79), imply that $y(t)$ is of positive type:

$$\begin{aligned} &\sum_{i,j=1}^{n} c_i\bar{c}_j E\{\exp[i(t_i - t_j)H_T]\} \\ &\quad = \left(\sum_{j=1}^{n} c_j \exp(it_jH_T)\Psi_0, \sum_{i=1}^{n} c_i \exp(it_iH_T)\Psi_0\right) \\ &\quad = \left\| \sum_{j=1}^{n} c_j \exp(it_jH_T)\Psi_0 \right\|^2 \geq 0. \end{aligned} \tag{3.80}$$

Using the spectral decomposition of H_T and the Riemann–Lebesgue lemma, it is easily shown that if H_T has a continuous spectrum, then $\lim_{t\to\pm\infty} y(t) = 0$. As H_T is of the type studied by Friedrichs,[35] namely, obtained from an operator H_0 with continuous spectrum by perturbation with creation and destruction operators, it may be that the asymptotic behavior of $y(t)$ is given by spectral perturbation theory. In the present case a direct method will be used.

Asymptotic behavior. In order to show that $\lim_{t\to\infty} y(t) = 0$, it is convenient to introduce the Laplace transforms

$$\tilde{y}(\zeta) = \int_0^\infty i^{i\zeta t}y(t)\,dt, \quad \zeta = \nu + i\mu, \tag{3.81}$$

and

$$\tilde{k}(\zeta) = \int_0^\infty e^{i\zeta t}\Gamma(t)y(t)\,dt. \tag{3.82}$$

Since $y(t)$ is bounded and $\Gamma(t)$ is absolutely integrable, the former converges in the open half-plane $\mu > 0$, and the latter in the closed half-plane $\mu \geq 0$.

A few preliminary inequalities are needed which involve

$$c_1 = \int_0^\infty |\Gamma(t)|\,dt. \tag{3.83}$$

(i) If $\mu \geq 0$, then $|\tilde{k}(\zeta)| \leq c_1$. This is a consequence of (3.71).

(ii) If $\mu \geq 0$ and $|\zeta| \geq 2c_1$, then $(\tilde{k}(\zeta) - i\zeta)^{-1} = (i\zeta)^{-1} + \zeta^{-2}F(\zeta)$, with $|F(\zeta)| \leq 2$. This inequality is implied by

$$|\tilde{k}(\zeta) - i\zeta| \geq \tfrac{1}{2}|\zeta|, \tag{3.84}$$

[34] E. Hille and R. Phillips, *Functional Analysis and Semigroups* (American Mathematical Society, Providence, R.I., 1957), 2nd ed.

[35] K. O. Friedrichs, *Lectures in Applied Mathematics*, Vol. 3: "Perturbation of Spectra in Hilbert Space" (American Mathematical Society, Providence, R.I., 1965).

which holds under the same conditions as (ii) and follows from (i).

(iii) $|dy(t)|/dt \leq c_1$ for any t. This follows from (3.56) and (3.71).

(iv) Re $[\tilde{k}(\nu)] > 0$ for any real ν. *Proof;* Putting $\mu = 0$ in (3.82), and noticing that $\Gamma(t)$ and $y(t)$ are even, we obtain

$$\tilde{k}(\nu) = \tfrac{1}{2}\int_{-\infty}^{+\infty} e^{i\nu t}\Gamma(t)y(t)\,dt. \tag{3.85}$$

Since the Fourier transformation exchanges ordinary products and convolution products, $\tilde{k}(\nu)$ is the convolution product of the Fourier transforms of $\Gamma(t)$ and $y(t)$. The former is, by Sec. 3.C(iii), an everywhere strictly positive continuous function, and the latter is a positive measure, since $y(t)$ is of positive type. Therefore, their convolution product is an *everywhere* positive function. Q.E.D.

For $\mu > 0$, the Laplace transform of (3.56) reads

$$-i\zeta\tilde{y}(\zeta) - 1 = -\tilde{k}(\zeta)\tilde{y}(\zeta). \tag{3.86}$$

Hence,

$$\tilde{y}(\zeta) = [\tilde{k}(\zeta) - i\zeta]^{-1}. \tag{3.87}$$

As $y(t)$ has a bounded derivative, it is of bounded variation, and the Laplace inversion formula[36] gives for $\mu > 0$

$$y(t) = (2\pi)^{-1} \times \lim_{X\to\infty}\int_{-X}^{+X} e^{-i(\nu+i\mu)t}[\tilde{k}(\nu + i\mu) + \mu - i\nu]^{-1}\,d\nu. \tag{3.88}$$

For fixed real ν the integrand of (3.88) converges as $\mu \to 0$ to

$$e^{-i\nu t}[\tilde{k}(\nu) - i\nu]^{-1}, \tag{3.89}$$

which is finite because of (iv). Furthermore, the convergence is uniform for ν in any bounded closed interval $(-Y, +Y)$ because the integrand, as a function of ν and μ, is continuous on the compact set $(-Y, +Y) \times (0, \mu_1)$. Because of (ii), the integral of the limit exists (as a Cauchy principal value). It is then easily shown, by splitting up the integration interval $(-X, +X)$ into $(-X, -Y)$, $(-Y, +Y)$, and $(+Y, +X)$, and by using uniform convergence in $(-Y, +Y)$ and inequality (ii) in the remaining intervals, that

$$y(t) = (2\pi)^{-1}\lim_{X\to\infty}\int_{-X}^{+X} e^{-i\nu t}[\tilde{k}(\nu) - i\nu]^{-1}\,d\nu. \tag{3.90}$$

Thus, $y(t)$ appears as the Fourier transform of $[\tilde{k}(\nu) - i\nu]^{-1}$. From (ii), it is inferred that the latter is square integrable; hence, the former is also square integrable; since it has a bounded derivative, it follows that

$$\lim_{t\to\pm\infty} y(t) = 0. \tag{3.91}$$

Explicit solutions. These are obtained for $\Gamma(t) = \epsilon^2 e^{-|t|}$. With the same notations as in the preceding section, we have

$$\tilde{k}(\zeta) = \int_0^\infty e^{i\zeta t}\epsilon^2 e^{-t}y(t)\,dt = \epsilon^2\tilde{y}(\zeta + i). \tag{3.92}$$

Equation (3.87) then becomes a difference equation

$$\tilde{y}(\zeta) = [\epsilon^2\tilde{y}(\zeta + i) - i\zeta]^{-1}. \tag{3.93}$$

Iterating (3.93), we obtain the continued-fraction expansion

$$\tilde{y}(\zeta) = \cfrac{1}{-i\zeta + \cfrac{\epsilon^2}{-i\zeta + 1 + \cfrac{\epsilon^2}{-i\zeta + 2 + \cdots + \cfrac{\epsilon^2}{-i\zeta + n} + \cdots}}}. \tag{3.94}$$

This well-known continued fraction is equal to the ratio of two Bessel functions,[37]

$$\tilde{y}(\zeta) = J_{(-i\zeta)}(2i\epsilon)/J_{(-i\zeta-1)}(2i\epsilon). \tag{3.95}$$

The original function $y(t)$, which is given by the Laplace inversion formula,[36] is, from a practical viewpoint, probably less useful than direct numerical solutions of (3.56).

Numerical solutions. The mean Green's function for the randomly frequency-modulated oscillator (3.5) with Gaussian $m(\omega; t)$ is known explicitly[5]; it reads

$$y_E(t) = \exp\left(-\tfrac{1}{2}\int_0^t\int_0^t \Gamma(t' - t'')\,dt'\,dt''\right). \tag{3.96}$$

It is interesting to compare this exact mean Green's function to the solution $y_K(t)$ of the Kraichnan equation (3.56) and to the solution $y_{C-D}(t)$ of the first cumulant-discard equation, which reads

$$\frac{dy(t)}{dt} = -\int_0^t \Gamma(t - t')y(t')\,dt', \quad y(0) = 1. \tag{3.97}$$

All three mean Green's functions $y_E(t)$, $y_K(t)$, and $y_{C-D}(t)$ have been calculated numerically by discretization of the time variable and for the following

[36] G. Doetsch, *Mathematische Reihe, BAND 24: Einführung in Theorie und Anwendung der Laplace-Transformation* (Birkhäuser Verlag, Basel, 1958).

[37] G. N. Watson, *Theore of Bessel Functions* (Cambridge University Press, New York, 1944), 2nd ed.

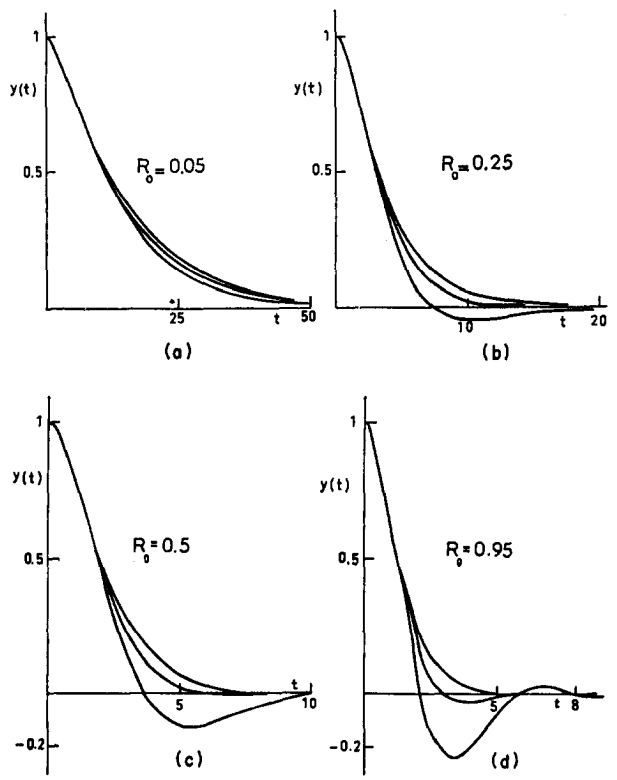


FIG. 2. Exact mean Green's function $y_E(t)$ for the randomly frequency-modulated oscillator (upper curve). Mean Green's function $y_K(t)$ for the corresponding random-coupling model (middle curve). Mean Green's function $y_{C-D}(t)$ for the corresponding first cumulant-discard approximation (lower curve). All three calculated with the covariance $\Gamma_1(t) = R_0^2(1 + t^2)^{-1}$. The parameter R_0 is the generalized Reynolds number.

choices of the covariance:

$$\Gamma_1(t) = \epsilon^2[1 + (t/T)^2]^{-1},$$

$$\Gamma_2(t) = \epsilon^2 e^{-t/T}, \quad \text{and} \quad \Gamma_3(t) = \epsilon^2 e^{-t^2/T^2}. \tag{3.98}$$

It is easily found that in each case, the mean Green's functions actually depend, except for a rescaling of the time, only on the dimensionless number

$$R_0 = \epsilon T, \tag{3.99}$$

called the "generalized Reynolds number." [5]

The results of the calculation with $\Gamma_1(t)$ have been plotted on Figs. 2a–2d for various values of R_0. They show that

(i) y_E and y_{C-D} agree within less than five per cent as long as R_0 is less than one-tenth;

(ii) y_E and y_K agree within less than five per cent as long as R_0 is less than one-half;

(iii) the relaxation times for y_E and y_K are approximately equal for arbitrary large R_0;

(iv) there is a good qualitative agreement between y_E and y_{C-D} as long as R_0 is less than one fourth; for larger values of R_0 the first cumulant-discard approximation gives insufficient damping.

Very similar results, in particular for the critical values of R_0, are obtained with the covariances Γ_2 and Γ_3.

E. Concluding Remarks

It is hoped that the parastochastic formulation has contributed to clarify the mathematical basis of the method of Kraichnan, which, for the present, is the only one that (i) involves no heuristic approximations, (ii) applies to all linear stochastic equations without any restriction on the covariances of the coefficients or on the number of independent variables, (iii) yields a tractable master equation.

The method described in Sec. 3B.3 for stochastic differential equations with a dichotomic Markov process (D.M.P.) as coefficient has two advantages over the method of Kraichnan: the original equation is solved exactly, and the resulting master equation (3.45) is more tractable than the Kraichnan equation, for it can be solved by integrals. On the other hand, the range of possible applications is of course limited to problems with one independent variable. Applications of the D.M.P. to the Goldstein diffusion model, proton magnetic resonance, and Brownian motion have already been published elsewhere.[4] In a forthcoming paper by one of the authors (Frisch), the distribution of electronic energy levels in a one-dimensional binary alloy will be calculated exactly, provided that the potential function can be represented by a D.M.P.

A few open problems are now indicated.

Realizations and random representations of the time-dependent generalized commutation relations have been obtained only for $\lambda = +1, 0, -1$. What happens for other values of λ?

It is of fundamental importance, we believe, to give a rigorous basis to the conjecture made in Sec. 3C that the solution of the Kraichnan equation approaches the exact mean Green's function as $R_0 \to 0$. To make a more precise statement, let us again consider the parastochastic equation (3.59) for the randomly frequency-modulated oscillator. Under the assumption of stationarity, this equation was converted into an evolution equation (3.77) involving the time-independent "Hamiltonian" $H_T = H_0 + M(0)$. Introducing a positive generalized Reynolds number R_0, we consider the two families of Hamiltonians,

$$H_T^{(\lambda)}(R_0) = H_0 + R_0 M(0), \quad \lambda = +1, 0, \tag{3.100}$$

where the superscripts one and zero refer to the Gaussian case and the random-coupling model, respectively. Denoting by $dE^{(\lambda)}(m, R_0)$ the spectral

measure[17] associated to $H_T^{(\lambda)}$, we conjecture that, as $R_0 \to 0$,

$$\sup_B \left| \int_B (\Psi_0, [dE^{(1)}(mR_0^2, R_0) - dE^{(0)}(mR_0^2, R_0)]\Psi_0) \right| = O(R_0^{-1}), \quad (3.101)$$

where B is an arbitrary Borel subset of the real line. If $\Gamma(t) = R_2^0 e^{-|t|}$, Eq. (3.101) can be checked (not too easily) on the explicit solutions (3.95) and (3.96). Anyhow, such a conjecture seems to account well for the observed numerical "closeness," even in the asymptotic range (see end of Sec. 3D). One should try to prove it directly on the operators (i.e., not on the master equations), e.g., by means of the perturbation method of Friedrichs.[35] The proof might then be extended to stochastic equations with coefficients that are stationary in time but depend also on other variables because the method which led us to introduce a time-independent Hamiltonian is easily generalized.

Finally, we recall that the method of Kraichnan was originally designed for turbulence dynamics[13,30] and the many-body problem[38] rather than for linear stochastic equations. It may be asked if those problems have a parastochastic reformulation, too. Although this question is still under investigation, we would like to mention that Kraichnan's master equations for turbulence in the random-coupling model have already been obtained without any recourse to perturbation theory. The derivation, based on the result stated in Remark 7, is quite similar to that of Eqs. (3.32) and (3.34).

ACKNOWLEDGMENTS

It is our pleasant duty to thank Dr. Robert Kraichnan for pointing out the analogy between his random-coupling model and Wigner matrices, Dr. Freeman Dyson for showing that the Kraichnan equation could, at least in a special case, be derived without recourse to perturbation expansions, Dr. Steven Orszag for many fruitful discussions, Dr. Joseph Keller, Mr. Louis Boutet de Monvel, and Mr. Gérard Schiffmann for mathematical assistance, and Mrs. Hélène Frisch for assistance in numerical calculations.

APPENDIX

The probability distribution $dP_\lambda(m)/dm$ is calculated for $-1 \leq \lambda < 1$; the notations are the same as in Sec. 2.B. In view of the boundedness of M the *moment generating function*

$$F_\lambda(z) = \sum_{n \geq 0} z^{-n-1}(\Psi_0, M^n \Psi_0) \quad (A1)$$

converges for $|z|$ large enough to the Stieltjes transform

$$F_\lambda(z) = \int_{-\infty}^{+\infty} \frac{dP_\lambda(m)}{z - m} \quad (A2)$$

of the probability distribution. Introducing the vector

$$\mathbf{x} = \sum_{n \geq 0} z^{-n-1}(A + B)^n \Psi_0, \quad (A3)$$

we notice that

$$F_\lambda(z) = (\Psi_0, \mathbf{x}). \quad (A4)$$

Obviously, $\mathbf{x}$ satisfies the equation

$$\mathbf{x} = z^{-1}(A + B)\mathbf{x} + z^{-1}\Psi_0. \quad (A5)$$

Using the representation (2.13) and (2.14), we obtain a second-order difference equation for the coordinates $\{x_n\}_{n \geq 0}$ of $\mathbf{x}$:

$$x_0 = z^{-1} f_0 x_1 + z^{-1},$$
$$x_n = z^{-1}(f_{n-1} x_{n-1} + f_n x_{n+1}), \quad n \geq 1. \quad (A6)$$

Dividing the last equation by x_n, we have

$$\frac{x_n}{x_{n-1}} = \frac{z^{-1} f_{n-1}}{1 - z^{-1} f_n x_{n+1}/x_n}. \quad (A7)$$

Iterating (A7), we obtain the continued-fraction expansion

$$F_\lambda(z) = x_0 = \cfrac{1}{z - \cfrac{f_0^2}{z - \cfrac{f_1^2}{z - \cdots - \cfrac{f_n^2}{z - \cdots}}}}, \quad (A8)$$

where the f_n's are given by (2.16).

It follows from the Poincaré–Perron Theorem[39] and the convergence of Σx_n^2 that the continued fraction picks out the right solution of (A6). Since (A8) is a positive-definite J-fraction,[40] it converges for $\operatorname{Im} z > 0$ to the Stieltjes' transform of a probability distribution, which is even because $F_\lambda(z)$ is an odd function of z. This probability distribution can be obtained by means of the Stieltjes' inversion formula[40]; for the case of an absolutely continuous even probability distribution, this formula reads

$$\frac{dP_\lambda(m)}{dm} = -\frac{1}{\pi} \lim_{\substack{\eta > 0 \\ \eta \to 0}} \operatorname{Im} F_\lambda(m + i\eta). \quad (A9)$$

If the probability distribution is not absolutely continuous, the same formula still holds provided that the limit is taken in the sense of distributions.

[38] R. H. Kraichnan, J. Math. Phys. **3**, 475 and 496 (1962).

[39] L. M. Milne-Thomson, *The Calculus of Finite Differences* (Macmillan and Co., Ltd., London, 1951), 2nd ed.

[40] H. S. Wall, *Analytic Theory of Continued-Fractions* (D. Van Nostrand, Inc., Princeton, N.J., 1948).

Let us first look at two special cases. For $\lambda = -1$, (A8) becomes

$$F_{-1}(z) = z/(z^2 - 1), \tag{A10}$$

which is the Stieltjes' transform of $\frac{1}{2}[\delta(m + 1) + \delta(m - 1)]$, the simplest Bernouilli distribution. For $\lambda = 0$, (A8) becomes

$$F_0(z) = \cfrac{1}{z - \cfrac{1}{-z - \cdots}} \cdots = \tfrac{1}{2}[z \pm (z^2 - 4)^{\frac{1}{2}}], \tag{A11}$$

where the analytic branch, which behaves like z^{-1} for large $|z|$ in the upper half-plane, must be taken. Applying the Stieltjes' inversion formula to (A11), we obtain the semicircle distribution (2.21).

In the general case, we cannot directly apply the inversion formula to (A8) for the continued fraction converges more and more slowly as $\operatorname{Im} z \to 0$. Instead, we use the following indirect procedure. For $-1 < \lambda < +1$, the sequence f_n^2 converges to $(1 - \lambda)^{-1}$ as $n \to \infty$. In (A8), we approximate f_n^2 by $(1 - \lambda)^{-1}$ for $n \geq N$, and notice that

$$\cfrac{(1-\lambda)^{-1}}{z - \cfrac{(1-\lambda)^{-1}}{-z - \cdots}} \cdots = (1 - \lambda)^{-\frac{1}{2}} F_0((1 - \lambda)^{\frac{1}{2}} z). \tag{A12}$$

Applying the inversion formula to the approximated continued fraction, we obtain

$$\frac{dP(m)}{dm} \approx -\frac{1}{\pi} \operatorname{Im}\left[\cfrac{1}{z - \cfrac{f_0^2}{-z - \cdots - \cfrac{f_{N-1}^2}{-z - (1-\lambda)^{-\frac{1}{2}} R_\lambda(m)}}}\right], \tag{A13}$$

with

$$R_\lambda(m) = \lim_{\substack{\eta > 0 \\ \eta \to 0}} F_0((1 - \lambda)^{\frac{1}{2}}(m + i\eta)). \tag{A14}$$

Using (A11), we obtain

$$\begin{aligned} R_\lambda(m) &= \tfrac{1}{2}\{(1-\lambda)^{\frac{1}{2}} m - i[4 - (1-\lambda)m^2]^{\frac{1}{2}}\}, \\ &\qquad \text{if } |m| < 2(1-\lambda)^{-\frac{1}{2}}, \\ &= \tfrac{1}{2}\{(1-\lambda)^{\frac{1}{2}} m - [(1-\lambda)m^2 - 4]^{\frac{1}{2}}\}, \\ &\qquad \text{otherwise.} \end{aligned} \tag{A15}$$

For $|m| \geq 2(1 - \lambda)^{-\frac{1}{2}}$, the probability density $dP_\lambda(m)/dm$ vanishes because the coefficients of (A13) and $R(m)$ are real. For $|m| < 2(1 - \lambda)^{-\frac{1}{2}}$, the finite fraction (A13) has been calculated numerically for 19 values of λ ranging from -0.9 to $+0.9$. N was successively given the values 10 and 40; no appreciable differences were found, even for $\lambda = +0.9$. The results are shown on Fig. 1; the probability distribution being even, only positive values are plotted.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Unique Hamiltonian Operators via Feynman Path Integrals

EDWARD H. KERNER AND WILLIAM G. SUTCLIFFE*
Sharp Physics Laboratory, University of Delaware, Newark, Delaware

(Received 12 May 1969)

The old problem of how to represent uniquely a prescribed classical Hamiltonian H as a well-defined quantal operator $\hat{H}$ is shown to have a clear answer within Feynman's path-integral scheme (as expanded by Garrod) for quantum mechanics. The computation of $\hat{H}$ involves the momentum Fourier transform of a coordinate average of H. A differential equation for a reduced form of the Feynman propagator giving $\hat{H}$ from H is found; and the example of polynomial H worked out to give the Born–Jordan ordering rule for $\hat{H}$ in this case.

INTRODUCTION

From its beginnings, quantum theory, as comprised fundamentally of commutation rules on coordinates and momenta, plus the dynamical law $\hat{H}\psi = i\hbar\dot{\psi}$, has contained a substantial logical gap: how to represent the Hamiltonian, say $H(x, p)$, uniquely as an operator.[1] When H has simple structure like $f(x) + g(p)$, the problem does not seem to arise, e.g., $f(x) + g(-i\hbar\partial/\partial x)$ will often do; but even here there is ambiguity owing to the ambiguity of the idemfactor $\varphi(x, p)/\varphi(x, p) = 1$, which multiplies f or g. Apart from this, when the Hamiltonian contains a term like x^2p^2, for instance, the question[2] of ordering x- and p-factors receives no clear answers under representations of x and p in accordance with the commutation rules. A logically distinct ordering rule becomes necessary. Many have been proposed (Born–Jordan, Weyl–McCoy, etc.). Each has a certain plausibility.[3] None has universality or basic theoretical standing, as their variety attests.

In the present note[4] it will be shown that Feynman's[5] path-integral formulation of quantum mechanics in the Hamiltonian statement of it, first indicated by Feynman and later developed by Garrod,[6] does provide a unique construction of the Hamiltonian operator. The reason is that the path-integral scheme moves directly to seize the statement of the dynamical law, based upon the primitive action principle, without so much as a glance at commutation rules; these come up only later, as a deduction, not as a first principle.

CALCULATION OF HAMILTONIAN OPERATOR

We confine discussion to a particle dynamics controlled by Hamilton's equations $\dot{\mathbf{x}} = \partial H/\partial \mathbf{p}$, $\dot{\mathbf{p}} = -\partial H/\partial \mathbf{x}$, which make stationary the action

$$A(\mathfrak{F}) = \int_{\mathfrak{F}} \mathbf{p} \cdot d\mathbf{x} - H\, dt \tag{1}$$

computed on a set of paths $\mathfrak{F}$ between fixed endpoints in $\mathbf{x}$, $\mathbf{p}$ space. Let us agree that $\mathbf{x}$ stands for the collection of all Cartesian coordinates, $\mathbf{p}$ their canonical mates. It is a separate issue, which will not be considered here, to allow general contact transformations or generalized coordinates other than Cartesians. Except as noted below, only a Cartesian coordinate representation of quantum mechanics is then in view.

Feynman's hypothesis is that quantal probability amplitudes $\psi(\mathbf{x}, t)$ evolve according to

$$\psi(\mathbf{x}'', t'') = \int K(\mathbf{x}'', \mathbf{x}', t'' - t')\psi(\mathbf{x}', t')\, d^n\mathbf{x}', \tag{2}$$

where the propagator K is computed from the "sum over all paths $\mathfrak{F}$":

$$K \sim \sum_{\mathfrak{F}} \exp(i/\hbar)A(\mathfrak{F}), \tag{3}$$

$A(\mathfrak{F})$ being the classical action equation (1). Rather than Feynman's action $\int L(\mathbf{x}, \dot{\mathbf{x}})\, dt$, which apparently limits things to Lagrangians that are quadratic in $\dot{\mathbf{x}}$, it is Garrod's use of $A(\mathfrak{F})$ that holds the power to advance to very general Hamiltonians. The proportionality constant in Eq. (3) is to be taken simply as $(2\pi\hbar)^{-n}$, where n is the number of degrees of freedom, in order to recover ordinary Schrödinger dynamics.

Now, following Garrod initially, we work out Eqs. (2) and (3) infinitesimally. Let $t'' - t'$ be very small. In accordance with Eq. (1), we reckon a family of paths $\mathfrak{F}$ with fixed endpoints for $t' \leq t \leq t''$ as

$$\mathfrak{F}_1\colon \mathbf{p} = \mathbf{p}_1, \quad \mathbf{x} = \mathbf{x}' + [(\mathbf{x}'' - \mathbf{x}')/(t'' - t')](t - t'),$$
$$\mathfrak{F}_2\colon \mathbf{p} = \mathbf{p}_2, \quad \mathbf{x} = \mathbf{x}' + [(\mathbf{x}'' - \mathbf{x}')/(t'' - t')](t - t'),$$

---

* Present address: Lawrence Radiation Laboratory, Livermore, California.

[1] See J. R. Shewell [Am. J. Phys. **27**, 16 (1957)] for bibliography.

[2] L. Cohen, J. Math. Phys. **7**, 781 (1966).

[3] For instance, the import of the Weyl–McCoy rule is shown by U. Uhlhorn, Ark. Fysik **11**, 87, (1956).

[4] A preliminary report was given in Bull. Am. Phys. Soc. **14**, 68 (1969).

[5] R. P. Feynman, Rev. Mod. Phys. **20**, 367 (1948).

[6] R. P. Feynman, Phys. Rev. **84**, 108 (1951) (Appendix B); C. Garrod, Rev. Mod. Phys. **38**, 483 (1966).

and so on, i.e., $\mathfrak{F}$ is simply straight-line motion in $\mathbf{x}$ running from $\mathbf{x}'$ to $\mathbf{x}''$ at an arbitrary but fixed momentum $\mathbf{p}$. The family $\mathfrak{F}$ is then parametrized by $\mathbf{p}$, and the $\sum_{\mathfrak{F}}$ is just an $\int d\mathbf{p}$. Equation (3) reads

$$K = (2\pi\hbar)^{-n}\int d^n\mathbf{p} \exp(i/\hbar)[\mathbf{p}\cdot(\mathbf{x}''-\mathbf{x}') - (t''-t')\bar{H}],$$

where $\bar{H}$ is the time average of H over $\mathbf{p}$ = fixed, $\mathbf{x} = \mathbf{x}(t)$; that is,

$$\bar{H} = \frac{1}{t''-t'}\int_{t'}^{t''} H\left(\mathbf{x}' + \frac{\mathbf{x}''-\mathbf{x}'}{t''-t'}(t-t'), \mathbf{p}\right) dt$$
$$= \int_0^1 H(\mathbf{x}' + (\mathbf{x}''-\mathbf{x}')u, \mathbf{p})\, du. \tag{4}$$

Expanding the exponential in K to first order in $t''-t'$ [7] (nothing more is needed in view of the limit $t''-t' \to 0$, to be taken in a moment), we have

$$K = \delta(\mathbf{x}''-\mathbf{x}') - (i/\hbar)k(\mathbf{x}'', \mathbf{x}')(t''-t'),$$

where k, the "reduced propagator," is defined as

$$k(\mathbf{x}'', \mathbf{x}') \equiv (2\pi\hbar)^{-n}\int d^n\mathbf{p}\bar{H} \exp(i/\hbar)\mathbf{p}\cdot(\mathbf{x}''-\mathbf{x}'). \tag{5}$$

Now the dynamical law equation (2) becomes

$$i\hbar\frac{\psi(\mathbf{x}'',t'') - \psi(\mathbf{x}'',t')}{t''-t'} = \int k(\mathbf{x}'',\mathbf{x}')\psi(\mathbf{x}',t')\, d^n\mathbf{x}',$$

and in the limit $t'' \to t'$ this is

$$i\hbar\frac{\partial\psi(\mathbf{x}'',t')}{\partial t'} = \int k(\mathbf{x}'',\mathbf{x}')\psi(\mathbf{x}',t')\, d^n\mathbf{x}'.$$

The essentially unambiguous meaning of the Hamiltonian operator in the sense $i\hbar\dot{\psi} = \hat{H}\psi$, and the connection of quantal $\hat{H}$ and classical H, is now evident:

$$\hat{H}\psi(\mathbf{x}'',t') \equiv \int k(\mathbf{x}'',\mathbf{x}')\psi(\mathbf{x}',t')\, d^n\mathbf{x}',$$
$$k \equiv \mathbf{p}\text{-Fourier transform of } \mathbf{x}\text{-time average of } H, \text{ averaged over an infinitesimal interval in which coordinates run in straight-line motion [Eqs. (4) and (5)].} \tag{6}$$

In a purely formal manner, Eq. (4) can be expressed as

$$\bar{H} = \int_0^1 [\exp u(\mathbf{x}''-\mathbf{x}')\cdot\nabla']H(\mathbf{x}',\mathbf{p})\, du$$
$$= \frac{[\exp(\mathbf{x}''-\mathbf{x}')\cdot\nabla'] - 1}{(\mathbf{x}''-\mathbf{x}')\cdot\nabla'} H(\mathbf{x}',\mathbf{p}),$$

where the exponential is understood to be written as a power series in $(\mathbf{x}''-\mathbf{x}')\cdot\nabla'$ (which sometimes can break off in a polynomial). Denoting the $\mathbf{p}$-Fourier transform of H as

$$\Phi(\boldsymbol{\xi})\cdot H(\mathbf{x}',\mathbf{p})$$
$$\equiv \mathcal{H}(\mathbf{x}', \hbar\boldsymbol{\xi}) = \frac{1}{(2\pi\hbar)^n}\int H(\mathbf{x}',\mathbf{p})\left(\exp\frac{i}{\hbar}\mathbf{p}\cdot\boldsymbol{\xi}\right) d^n\mathbf{p},$$

the reduced propagator is, formally,

$$k = \left\{\frac{[\exp(\mathbf{x}''-\mathbf{x}')\cdot\nabla'] - 1}{(\mathbf{x}''-\mathbf{x}')\cdot\nabla'}\mathcal{H}(\mathbf{x}', \hbar\boldsymbol{\xi})\right\}_{\boldsymbol{\xi}=\mathbf{x}''-\mathbf{x}'},$$

where in Eq. (5) the operators of Fourier transforming and the time averaging have been reversed.

The formal calculation suggests in passing that there is a simple differential equation for k which may show its makeup a bit more clearly. This can be obtained from Eq. (5) by performing

$$(\mathbf{x}''-\mathbf{x}')\cdot(\nabla'+\nabla'')k$$
$$= (\mathbf{x}''-\mathbf{x}')\cdot(\nabla'+\nabla'')\Phi(\mathbf{x}''-\mathbf{x}')$$
$$\cdot\int_0^1 H(\mathbf{x}' + u(\mathbf{x}''-\mathbf{x}'), \mathbf{p})\, du.$$

Call $\mathbf{w}$ the argument $\mathbf{x}' + u(\mathbf{x}''-\mathbf{x}')$. Then the right-hand side here is

$$0 + \Phi\cdot\int_0^1 (\mathbf{x}''-\mathbf{x}')\cdot[u\nabla_w + (1-u)\nabla_w]H(\mathbf{w},\mathbf{p})\, du$$
$$= \Phi\cdot\int_0^1 (\mathbf{x}''-\mathbf{x}')\cdot\nabla_w H(\mathbf{w},\mathbf{p})\, du$$
$$= \Phi\cdot\int_0^1 \frac{\partial}{\partial u} H(\mathbf{x}' + u(\mathbf{x}''-\mathbf{x}'),\mathbf{p})\, du$$
$$= \Phi\cdot[H(\mathbf{x}'',\mathbf{p}) - H(\mathbf{x}',\mathbf{p})],$$

so that k satisfies

$$(\mathbf{x}''-\mathbf{x}')\cdot(\nabla'+\nabla'')k = \Phi\cdot[H(\mathbf{x}'',\mathbf{p}) - H(\mathbf{x}',\mathbf{p})]$$

or

$$(\mathbf{x}''-\mathbf{x}')\cdot(\nabla'+\nabla'')k$$
$$= \mathcal{H}(\mathbf{x}'', \hbar(\mathbf{x}''-\mathbf{x}')) - \mathcal{H}(\mathbf{x}', \hbar(\mathbf{x}''-\mathbf{x}')).$$

In the whole discussion above, it will be clear how to go into a momentum representation by starting with an action

$$\int -\mathbf{x}\cdot d\mathbf{p} - H\, dt$$

instead of the "coordinate" action (1), to which it is canonically equivalent [the two action integrands differing by $d(\mathbf{x}\cdot\mathbf{p})$].

[7] R. P. Feynman, Ref. 5, 374.

EXAMPLES OF OPERATOR ORDERING

In one dimension let H be $f(x)p^k$ (or a sum of such terms). Then it follows from Eq. (6) that

$$\begin{aligned}\hat{H}\psi(x'',t) &= \frac{1}{2\pi\hbar}\iint\left(\exp\frac{i}{\hbar}p(x''-x')\right)p^k \\ &\quad\times\int_0^1 f(x'+(x''-x')u)\,du\;\psi(x',t)\,dp\,dx' \\ &= (-i\hbar)^k\int\delta^{(k)}(x''-x') \\ &\quad\times\int_0^1 f(x'+(x''-x')u)\,du\;\psi(x',t)\,dx',\end{aligned}$$

or, dropping primes,

$$\hat{H}\psi(x,t) = (-i\hbar)^k\sum_{l=0}^{k}\binom{k}{l}\frac{f^{(k)}(x)}{l+1}\frac{\partial^{k-l}}{\partial x^{k-l}}\psi(x,t).$$

For $f=1$ and $k=1$, the conventional $\hat{p}=-i\hbar\partial/\partial x$ follows, in terms of which the operator equation

$$\hat{H} = \sum(-i\hbar)^l\binom{k}{l}\frac{f^{(l)}(x)}{l+1}\hat{p}^{k-l}$$

is obtained for $\hat{H}$. This is, for $f=x^m$,

$$\hat{H} = \widehat{x^m p^k} = \sum'(-i\hbar)^l\binom{k}{l}\binom{m}{l}\frac{l!}{l+1}x^{m-l}\hat{p}^{k-l},$$

with $\sum'$ meaning a sum on l from 0 to min (k,m).

Finally, let us compare the latter with other well-known ordering rules.

Weyl–McCoy rule:

$$\begin{aligned}\widehat{x^m p^k} &= 2^{-m}\sum\binom{m}{l}x^{m-l}\hat{p}^k x^l \\ &= \sum'(-i\hbar)^l\binom{k}{l}\binom{m}{l}\frac{l!}{2^l}x^{m-l}\hat{p}^{k-l},\end{aligned}$$

or, for $m=2$ and $k=2$,

$$\widehat{x^2p^2} = x^2\hat{p}^2 - 2i\hbar x\hat{p} - \tfrac{1}{2}\hbar^2.$$

Symmetrization rule:

$$\begin{aligned}\widehat{x^m p^k} &= \tfrac{1}{2}(x^m\hat{p}^k + \hat{p}^k x^m) \\ &= \tfrac{1}{2}x^m\hat{p}^k + \tfrac{1}{2}\sum'(-i\hbar)^l\binom{k}{l}\binom{m}{l}l!\,x_m^{m-l}\hat{p}^{k-l}\end{aligned}$$

and

$$\widehat{x^2p^2} = x^2\hat{p}^2 - 2i\hbar xp - \hbar^2.$$

Born–Jordan rule:

$$\begin{aligned}\widehat{x^m p^k} &= (k+1)^{-1}\sum\hat{p}^{k-l}x^m\hat{p}^l = (m+1)^{-1}\sum x^{m-l}\hat{p}^k x^l \\ &= \sum'(-i\hbar)^l\binom{k}{l}\binom{m}{l}\frac{l!}{l+1}x^{m-l}\hat{p}^{k-l},\end{aligned}$$

with

$$\widehat{x^2p^2} = x^2\hat{p}^2 - 2i\hbar\times\hat{p} - \tfrac{2}{3}\hbar^2.$$

These rules can be written in the common form

$$\widehat{x^m p^k} = \sum'(-i\hbar)^l\binom{k}{l}\binom{m}{l}\frac{l!}{a_l}x^{m-l}\hat{p}^{k-l},$$

where

$$\begin{aligned}a_l &= 2^l && \text{(Weyl–McCoy rule)},\\ &= \frac{2}{\delta_{l0}+1} && \text{(symmetrization rule)},\\ &= l+1 && \text{(Born–Jordan rule)}.\end{aligned}$$

The coefficients a_0 and a_1 are the same for all rules, so that all give the same result for k or m unity. For k and $m \geq 2$, the rules differ from each other by terms of order $\hbar^2$. The ordering rules proposed by Von Neumann and Dirac have been shown by Shewell[1] to be nonunique and so have not been considered. As mentioned in the beginning, all such rules stand as independent assumptions in quantum theory when based on commutation rules, whereas in the present discussion, for the simple case of polynomials in x and p, it is exactly the old Born–Jordan[8] rule that comes out straightforwardly from Feynman's single basic postulate.

Quite generally, the path-integral calculus in the form of the final result, Eq. (6), entails simply Fourier transformation and one quadrature (for $\hat{H}$) for obtaining $\hat{H}$ substantially uniquely. The range and freedom from ambiguity of this calculus may argue strongly for its fundamental position in quantum theory.

ACKNOWLEDGMENTS

This work stems out of another program, for whose support we are indebted to the National Science Foundation. We have enjoyed discussion with Dr. R. N. Hill.

[8] M. Born and P. Jordan, Z. Physik, 34, 873 (1925).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Bethe–Salpeter Equation and the Goldstein Problem

R. F. KEAM
Department of Physics, University of Auckland, Auckland, New Zealand

(Received 17 February 1969)

The Goldstein problem for the ladder-approximation Bethe–Salpeter equation for a spin-$\frac{1}{2}$ fermion–antifermion system bound to zero total mass by particle exchange is re-examined. It is suggested that the problem arises because attention has previously been focused on the wrong Dirac-space sectors of the equation. Criteria limiting the acceptable behavior of solutions are investigated and it is shown that discrete spectra for the values of the coupling constant are allowed to exist for the T–A sector and S–V sector solutions. Continuous spectra are always excluded.

1. INTRODUCTION

When the first solution of the ladder-approximation Bethe–Salpeter equation[1,2] describing a bound system of two fermions was found by Goldstein,[3] it turned out to have an associated problem which remains to the present day. Contrary to the expectation of finding a discrete set of values for the coupling constant g for the interaction between the fermions and the exchange field quanta for which the bound system would have zero total mass, the equation appeared to have solutions for a continuous range of values. Later, however, it was shown[4] that none of these solutions would satisfy a normalization condition. If the simple equation that Goldstein discussed is to have any physical significance, it appears necessary, therefore, to introduce a cut-off procedure which reduces or removes the contribution of high relative momenta to the solutions. This is justified by assuming that some equivalent, but as yet unknown, form factor exists. Narayanaswamy and Pagnamenta[5] have published the results of numerical calculations using such a cut-off. The present author has considered zero-total-mass solutions using a relativistic square-well potential in configuration space.[6] This is similar to introducing a cut-off, but one which is less violent than that of Narayanaswamy and Pagnamenta.

There is, however, another possibility. In two previous papers,[7,6] hereafter referred to, respectively, as I and II, we have considered formal solutions of the ladder-approximation equation for a zero-total-mass spin-$\frac{1}{2}$ fermion–antifermion system, and in Table I of II we have classified these solutions according to their total angular momentum $\mathbf{J}^2$, parity $\bar{\mathcal{P}}$, and charge parity $\bar{\mathcal{C}}$ properties. In particular, to the angular momentum and parity assignments J and $(-1)^{J+1}$, there correspond solutions of type

$$f^P, \quad g^{TA^-}, \quad f^{TA^-} \tag{1}$$

and to assignments J and $(-1)^J$, there correspond

$$f^{TA^+}, \quad g^{TA^+}, \quad f^{SV}. \tag{2}$$

For each of the six types of solutions, charge parity of either sign may be selected by suitable choice of a quantum number j. The Goldstein problem is associated with the solutions of type f^P and f^{TA^+}. It seems possible therefore that we may be able to avoid the problem by considering solutions of the other four types listed in sets (1) and (2).

In the following sections we shall investigate in detail the existence of solutions of each class for a particular case which allows some simplification and for which some solutions are already known, namely, where the exchange quanta are vector or axial-vector massless particles. Indeed, for the vector interaction, the work of Tiktopoulos[8] already shows that a discrete spectrum exists for nonvanishing as well as vanishing total mass. The results of our investigation applied to exchange of massless scalar or pseudoscalar particles are also quoted.

The conventions and notation are those of Papers I and II.

2. CONSTRAINTS

In relative-momentum space, the ladder-approximation Bethe–Salpeter equation for a spin-$\frac{1}{2}$ fermion and spin-$\frac{1}{2}$ antifermion, each of mass m, in a bound state of total momentum $P = (P^0, \mathbf{P})$ is

$$\{\gamma\cdot(p+\tfrac{1}{2}P) - im\}f(p)\{\gamma\cdot(p-\tfrac{1}{2}P) - im\} = i\lambda\pi^{-2}\int d^4k\Gamma f(k)(p-k)^{-2}, \tag{3}$$

where, in the notation of I, $\Gamma = -\hat{\Gamma}_2$ for a vector exchange, $\Gamma = \hat{\Gamma}_4$ for an axial-vector exchange, and $\lambda = g^2/(4\pi)^2$.

[1] H. A. Bethe and E. E. Salpeter, Phys. Rev. **84**, 1232 (1951).
[2] M. Gell-Mann and F. E. Low, Phys. Rev. **84**, 350 (1951).
[3] J. S. Goldstein, Phys. Rev. **91**, 1516 (1953).
[4] S. Mandelstam, Proc. Roy. Soc. (London) **A233**, 248 (1955).
[5] P. Narayanaswamy and A. Pagnamenta, Nuovo Cimento **53A**, 875 (1968).
[6] R. F. Keam, J. Math. Phys. **10**, 594 (1969).
[7] R. F. Keam, J. Math. Phys. **9**, 1462 (1968).
[8] G. Tiktopoulos, J. Math. Phys. **6**, 573 (1965).

In the center-of-mass system the Wick-rotated[9] form of (3) becomes

$$(\gamma \cdot p - im)f(p)(\gamma \cdot p - im) = -\lambda\pi^{-2}\int d^4k\Gamma f(k)(p-k)^{-2}, \quad (4)$$

where the total mass E vanishes.

Alternatively, in configuration space[10] we have, in place of (3),

$$f(x)_{\alpha\beta} = -i\lambda\int d^4x'[G_D(x-x')]_{\alpha\rho,\sigma\beta}(4x'^{-2})[\Gamma f(x')]_{\rho\sigma}, \quad (5)$$

where the Green's function is

$$[G_D(y)]_{\alpha\rho,\sigma\beta} = \{\gamma \cdot (\partial_y + \tfrac{1}{2}iP) - m\}_{\alpha\rho} \times \{\gamma \cdot (\partial_y - \tfrac{1}{2}iP) - m\}_{\sigma\beta}G(y) \quad (6)$$

and[11]

$$G(y) = (2\pi)^{-4}\int d^4q e^{iq\cdot y}(q_1^2 + m^2)^{-1}(q_2^2 + m^2)^{-1},$$

where

$$q_1 = \tfrac{1}{2}P + q \quad \text{and} \quad q_2 = \tfrac{1}{2}P - q. \quad (7)$$

The Wick-rotated form of (5) with $P = 0$ may be written, for example, as

$$f(x) = \lambda(\gamma \cdot \partial_x - m) \times \int d^4x' G(x-x')(4x'^{-2})\Gamma f(x')(\gamma \cdot \overleftarrow{\partial}_x - m). \quad (8)$$

There are a number of constraints which any formal solution of the problem must satisfy.

A. Constraints Imposed by the Integral Form of the Equations

Instead of working directly with the integral Eqs. (3) or (5), it is often more convenient to study a differential equation derived from one of these. Clearly, however, any solution of such a differential equation has to satisfy the integral equation in order to be acceptable. After Wick rotation and when $E = 0$, angular and Dirac-space factors may be separated, and this has been done explicitly for configuration space in Paper I. For the integrals in the right members of (4) and (8) to exist, the radial functions are restricted in their behavior near the origin and in the asymptotic region. Suppose that K and R are four-dimensional radial coordinates in the Euclidean momentum space and Euclidean configuration space, respectively. For the radial factor $g(K)$ or $g(R)$ associated with some set of terms in the complete solution $f(k)$ or $f(x)$, if

$$g(K) \sim K^a \quad \text{as} \quad K \to 0,$$
$$g(K) \sim K^b \quad \text{as} \quad K \to \infty,$$
$$g(R) \sim R^c \quad \text{as} \quad R \to 0, \quad (9)$$

then, in general, for the inverse-square potential we have assumed

$$a > -4, \quad b < -2, \quad c > -2. \quad (10)$$

An exception occurs for the radial factor for any set of terms that the Γ operator causes to vanish. In our case, this occurs for tensor terms in f. The integral equation, therefore, places no direct restriction on the behavior of radial factors w, $\bar{w}$, t_1, and t_2 (in the notation of Paper I) either in momentum space or configuration space.

In configuration space, $g(R)$ behaves asymptotically not as a power but as an exponential. Clearly, it must be a negative exponential.

We note that the condition on b is sufficient to allow the Wick rotation.

B. Configuration- and Momentum-Space Interchangeability

We assume that the problem can be described both in momentum space and in configuration space. That is, the Fourier transforms of the wavefunctions in either space must exist. Referring to Papers I and II, we note that each separate radial function in a complete solution is associated in the bispherical basis with angular functions $Z_{lm_l^+m_l^-}$ of fixed l. It can then be shown[12] that the corresponding radial functions $g_l(K)$ and $g_l(R)$ are related by the reciprocal formulas:

$$Rg_l(R) = (2\pi)^{-2}i^{2l}\int_0^\infty dKKJ_{2l+1}(KR)Kg_l(K), \quad (11a)$$

$$Kg_l(K) = (2\pi)^2 i^{-2l}\int_0^\infty dRRJ_{2l+1}(KR)Rg_l(R), \quad (11b)$$

i.e., $(2\pi)^2Rg_l(R)$ and $Kg_l(K)$ are mutual Hankel transforms[13] with respect to the Bessel function of order $2l + 1$. The existence of these transforms implies

$$a > -2l - 4, \quad b < -\tfrac{3}{2}, \quad c > -2l - 4. \quad (12)$$

C. Configuration Space Normalization

Configuration space wavefunctions must satisfy the Mandelstam criterion[4] for normalization to be possible, i.e.,

$$c > -1. \quad (13)$$

[9] G. C. Wick, Phys. Rev. **96**, 1124 (1954).
[10] See Ref. 9, Eq. (38).
[11] The function $G(y)$ is treated in full in R. F. Keam, J. Math. Phys. **7**, 2196 (1966).
[12] For example, as is done in the Appendix.
[13] V. A. Ditkin and A. P. Prudnikov, *Integral Transforms and Operational Calculus* (Pergamon Press, Oxford, 1965), pp. 71–75.

D. Momentum-Space Normalization

Momentum-space wavefunctions must satisfy a normalization condition which we examine in detail in the next section. As is discussed there, we take this constraint to imply

$$a > -2, \quad b < -2. \tag{14}$$

Collecting all the constraints, we have, finally,

$$a > -2, \quad b < -2, \quad c > -1. \tag{15a}$$

From the considerations of the next section, it will appear that some question remains concerning the limit on b in (14). While it appears *necessary* for $b < -2$, this condition may not be restrictive enough. One would expect that the configuration-space condition (13) be equivalent to some condition in momentum space. The rigorous determination of this condition will require a much more careful treatment of the various possibilities. However, we shall now present two facts from which it seems reasonable to expect that the required momentum-space condition is probably

$$b < -3. \tag{16}$$

The first[14] is that mathematically (16) is a sufficient (but not necessary) condition to imply (13) and the second is that Goldstein's P-sector solutions,[15] whose transforms violate (13), can be excluded only by assuming (16). We notice also that (16) is equivalent to the square-integrability condition Tiktopoulos has assumed.

Taking account of this argument, we emend (15a) to

$$a > -2, \quad b < -3, \quad c > -1. \tag{15b}$$

When any of these quantities is complex, we understand that its *real* part satisfies (15b).

There is a quoted S–V sector solution[16,17] which has $b = -3$ and whose transform is stated to obey (13). We do not, however, confirm Eq. (3.22) of Ref. 16, and the correct Fourier transform has $c = -1$.

3. NORMALIZATION CONDITION

We consider the normalization condition in momentum space in the form given by Lurié, Macfarlane, and Takahashi[18]:

$$(2\pi)^{-4} i \iint d^4p\, d^4q \bar{f}(p)_{\beta\alpha} \times \{(\partial/\partial P^0)[I(pqP) + G(pqP)]_{\alpha\beta\gamma\delta}\} f(q)_{\gamma\delta} = 2P^0, \tag{17}$$

with $P^2 = -E^2$, where E is the total proper mass of the bound system. In our case the interaction function G is independent of P^0, and

$$I(pqP)_{\alpha\beta\gamma\delta} = \delta^{(4)}(p - q)\{\gamma \cdot (p + \tfrac{1}{2}P) - im\}_{\alpha\gamma} \times \{\gamma \cdot (p - \tfrac{1}{2}P) - im\}_{\delta\beta}. \tag{18}$$

Regarded as a matrix, the conjugate wavefunction $\bar{f}$ satisfies

$$\bar{f}(\mathbf{p}, p^{0*}) = [\gamma_4 f(\mathbf{p}, p^0)\gamma_4]^\dagger \tag{19}$$

at all points in the complex p^0 plane except on the real-axis cuts. Equation (17) now simplifies to

$$-\tfrac{1}{2}(2\pi)^{-4} \int d^4p \,\mathrm{Tr}\,[\{\gamma \cdot (p - \tfrac{1}{2}P) - im\}\bar{f}\gamma_4 f - \{\gamma \cdot (p + \tfrac{1}{2}P) - im\} f \gamma_4 \bar{f}] = 2P^0, \tag{20}$$

and in this result we must put $P^2 = -E^2$. The terms involving $\gamma_4 ip^0$ cancel upon taking the trace. Let us regard the integrand as a function in the complex p^0 plane and perform the p^0 integration first. Provided $|f| = O(|p^0|^d)$, where $d < -\frac{1}{2}$ as $|p^0| \to \infty$, we may perform a Wick rotation in the normalization integral. The only visible change in (20) is then the appearance of an extra factor i in the left member. In the center-of-mass system P becomes $(\mathbf{0}, iE)$. We are interested in the form that the normalization condition takes in the limit as $E \to 0$. Let us suppose that all functions involved may be expanded in powers of E for small positive E, and in particular,

$$f = f_0 + Ef_1 + O(E^2) \tag{21}$$

so that f_0 satisfies (4). Then, from the Wick-rotated form of (20), we have

$$-\tfrac{1}{2}(2\pi)^{-4} i \int d^4p \,\mathrm{Tr}\,[A + E(-i\gamma_4 \bar{f}_0 \gamma_4 f_0 + B) + O(E^2)] = 2E, \tag{22}$$

where, after cycling factors inside the trace,

$$A = (\boldsymbol{\gamma} \cdot \mathbf{p} - im)(\bar{f}_0 \gamma_4 f_0 - f_0 \gamma_4 \bar{f}_0), \tag{23}$$

$$B = (\boldsymbol{\gamma} \cdot \mathbf{p} - im)(\bar{f}_0 \gamma_4 f_1 + \bar{f}_1 \gamma_4 f_0 - f_0 \gamma_4 \bar{f}_1 - f_1 \gamma_4 \bar{f}_0). \tag{24}$$

The contribution of A vanishes on performing the angular integrations in (22). The zero-mass limit of the normalization condition therefore becomes

$$\int d^4p \,\mathrm{Tr}\,(\gamma_4 \bar{f}_0 \gamma_4 f_0 + iB) = -4(2\pi)^4. \tag{25}$$

[14] See Ref. 13, Eq. (3.5.14).
[15] See Sec. 4.
[16] A. Bastai, L. Bertocchi, G. Furlan, and M. Tonin, Nuovo Cimento **30**, 1532 (1963).
[17] See Eq. (48).
[18] D. Lurié, A. J. Macfarlane, and Y. Takahashi, Phys. Rev. **140**, B1091 (1965). See especially Eq. (3.24).

If $\lambda = \lambda_0 + E\lambda_1 + O(E^2)$, then, in the limit $E \to 0$, f_1 satisfies

$$(\gamma \cdot p - im)f_1(\gamma \cdot p - im) + \lambda_0\pi^{-2}\int d^4k\Gamma f_1(k)(p-k)^{-2}$$
$$= (\lambda_1/\lambda_0)(\gamma \cdot p - im)f_0(\gamma \cdot p - im)$$
$$- \tfrac{1}{2}i\{\gamma_4 f_0(\gamma \cdot p - im) - (\gamma \cdot p - im)f_0\gamma_4\}, \quad (26)$$

an equation which is invariant only under three-dimensional rotations. We have been able to show[19] in a specific case that $\lambda_1 = 0$, and it is clear from (26) that f_1 does not vanish, nor is it a multiple of f_0. It is not clear, however, what general conclusions can be drawn about the term B in (25). We assume that its contribution to the integral is not infinite. The finiteness of $\int d^4p \operatorname{Tr}(\gamma_4 f_0 \gamma_4 f_0)$ can then be ensured by assuming (14).

4. SOLUTIONS OF TYPE f^P, f^{TA^+}

Since the substitution of any linear combination of matrices $\sigma_{\mu\nu}$ for $f(k)$ in the right member of (4) gives zero, there is no solution of type f^{TA^+}.

Putting $f^P = X(p)\gamma_5$ into (4), we obtain

$$(p^2 + m^2)X(p) = -4\epsilon\lambda\pi^{-2}\int d^4kX(k)(p-k)^{-2}, \quad (27)$$

where $\epsilon = \mp 1$ accordingly as the interaction is of V or A type. Equation (27) is equivalent to the equation originally discussed by Goldstein,[3] and its solutions were treated in detail by Kummer.[20] The different radial functions are distinguished by a parameter $n\ (= 2j)$ having integral values. The solutions, which arise for a continuous range of λ, have an acceptable radial behavior $\sim K^j$ as $K \to 0$. As $K \to \infty$, they behave as a linear combination of $K^{-3\pm\zeta}$, where

$$\zeta = [(2j+1)^2 + 16\epsilon\lambda]^{\frac{1}{2}}.$$

This satisfies (15a) for all cases $(2j+1)^2 < -16\epsilon\lambda$, where ζ is pure imaginary, and for a small range where ζ is real and less than 1. The corresponding configuration space radial functions can be obtained with Eq. (11a) and are $(mR)^{-1}K_\zeta(mR)$. These are, of course, obtained much more easily by considering directly the differential form of the Bethe–Salpeter equation in configuration space. They are unacceptable since they behave like $R^{-1-\zeta}$ as $R \to 0$ and so do not satisfy the Mandelstam criterion. Thus in this case it would appear that the criterion for momentum-space acceptability is (15b).

[19] A perturbation treatment of the ladder approximation Bethe–Salpeter equation for a bound fermion–antifermion system is currently in preparation.

[20] W. Kummer, Nuovo Cimento 31, 219 (1964).

5. SOLUTIONS OF TYPE $g^{TA^\pm}$

The configuration-space solutions $g^{TA^\pm} \equiv g^{TA^\pm}_{j\,j+1\,Jm}$ are defined in Eq. (12) of Paper II and contain three radial functions a, t_1, and t_2, associated in the notation of Paper I with bispherical angular functions $Z_{lm^+m^-}$ having, respectively, $l = j + \frac{1}{2}$, $l = j+1$, and $l = j$. These radial functions satisfy a coupled set of equations which are written explicitly in differential form in Eqs. (49) of I.

If we take $\rho = mR$ as the independent radial variable and write

$$d^\pm_\alpha = \frac{d}{d\rho} \pm \frac{2(j+\alpha)}{\rho} = \frac{1}{m}D^\pm_\alpha, \quad (28)$$

$$\Delta_j = d^+_{\frac{1}{2}}d^-_0 = d^-_{-\frac{1}{2}}d^+_1, \quad (29)$$

Eqs. (49) of I become

$$\Delta_{j+\frac{1}{2}}a + d^+_2 t_1 + d^-_0 t_2 = -(1 + 8\lambda\rho^{-2})a, \quad (30a)$$
$$d^-_{\frac{1}{2}}(d^-_0 t_2 + 2a) = -t_1, \quad (30b)$$
$$d^+_{\frac{1}{2}}(d^+_2 t_1 + 2a) = -t_2 \quad (30c)$$

for both V and A interactions with the potential we are considering. Using (11a), these can be converted into momentum-space equations[21] and, in terms of the dimensionless radial variable $\sigma = Km^{-1}$, they are

$$\Delta_{j+\frac{1}{2}}\{(1-\sigma^2)a + i\sigma(t_1 + t_2)\} = 8\lambda a,$$
$$\sigma^2 t_2 - t_1 = 2i\sigma a, \quad (31)$$
$$\sigma^2 t_1 - t_2 = 2i\sigma a.$$

[For momentum-space operators $d^\pm_\alpha$ and Δ_j, we replace ρ by σ in the definitions (28) and (29) above.] Immediately, we have

$$t_1 = t_2 = 2i\sigma a/(\sigma^2 - 1) \quad (32)$$

and

$$\Delta_{j+\frac{1}{2}}\left(\frac{(\sigma^2+1)^2}{\sigma^2-1}a\right) = -8\lambda a. \quad (33)$$

It can be verified for any interaction that if we expand $g^{TA^\pm}$ as

$$g^{TA^\pm} = \tfrac{1}{2}T^\pm_{\mu\nu}\sigma_{\mu\nu} + A^\pm_\mu i\gamma_5\gamma_\mu, \quad (34)$$

then

$$(\partial/\partial x^\mu)A^\pm_\mu = 0. \quad (35)$$

Comparing (33) and (35) with Kummer's equations (2.13) and (2.11), we see that $g^{TA^\pm}$ are just his "transversal" solutions. Since λ in (33) above is positive, we have the case where Kummer finds a continuous range of values that permit solution of the integral

[21] For a rigorous derivation one should go via the momentum-space integral equation.

equation.[22] He shows, however, that none of these is acceptable.

Near the origin the solution which behaves as σ^{2j+1} is acceptable. Asymptotically, it is always a linear combination of terms $\sim \sigma^{-3\pm\zeta_1}$, where $\zeta_1 = [(2j+2)^2 - 8\lambda]^{\frac{1}{2}}$. In order to exclude the solutions for $2\lambda \geq (j+1)^2$, it would thus appear that in this case also we must take the criterion for momentum-space acceptability as (15b).

6. SOLUTIONS OF TYPE f^{TA^-}

When $j = 0$, as a differential equation in configuration space we have the radial equation [cf. (52) of I]:

$$d_2^+ d_{\frac{1}{2}}^- a_1 = (1 + 8\lambda\rho^{-2})a_1. \tag{36}$$

Comparing this with the corresponding P-sector equation (53) of I,

$$d_{\frac{3}{2}}^+ d_0^- p = (1 \mp 16\lambda\rho^{-2})p, \tag{37}$$

we see that the solutions of (36) are those of (37), provided we put $j = \frac{1}{2}$ in the latter and also redefine λ. As discussed earlier, there is no acceptable P-sector solution, and hence also no acceptable solution of (36). It is just Kummer's "longitudinal" solution.

When $j \neq 0$, we have the coupled set of radial equations [cf. Eq. (46) of I]:

$$\begin{aligned}
&\Delta_j w + 2(d_{\frac{3}{2}}^+ a_1 + d_{-\frac{1}{2}}^- a_2) = -w,\\
&d_0^-[-d_{\frac{3}{2}}^+ a_1 + 2j(d_{-\frac{1}{2}}^- a_2 + w)]\\
&\qquad = -(2j+1)(1 + 8\lambda\rho^{-2})a_1, \qquad (38)\\
&d_1^+[d_{-\frac{1}{2}}^- a_2 + 2(j+1)(d_{\frac{3}{2}}^+ a_1 + w)]\\
&\qquad = -(2j+1)(1 + 8\lambda\rho^{-2})a_2.
\end{aligned}$$

It turns out that the problem is more easily discussed in momentum space. Applying the transformation (11a) to (38), we obtain[21]

$$\begin{aligned}
&(1-\sigma^2)iw - 2\sigma(a_1 + a_2) = 0,\\
&d_2^+ d_{\frac{1}{2}}^-\{[(2j+1) + \sigma^2]a_1 - 2j\sigma(\sigma a_2 - iw)\}\\
&\qquad = 8(2j+1)\lambda a_1, \qquad (39)\\
&d_1^+ d_{-\frac{1}{2}}^-\{[(2j+1) - \sigma^2]a_2 - 2(j+1)\sigma(\sigma a_1 - iw)\}\\
&\qquad = 8(2j+1)\lambda a_2.
\end{aligned}$$

Between these equations, a_1 and a_2 may be eliminated, and there remains the following fourth-order differential equation for w:

$$\begin{aligned}
&[(1+\sigma^2)d_{-\frac{1}{2}}^- d_2^+ - 8\lambda][d_{\frac{1}{2}}^+ d_0^-(1+\sigma^2)^2 - 8\lambda(1-\sigma^2)]w\\
&\qquad = -64j\lambda(1+\sigma^2)w. \qquad (40)
\end{aligned}$$

Given any solution of (40), the corresponding solutions for a_1 and a_2 may be obtained from

$$\begin{aligned}
&-8i(2j+1)(1+\sigma^2)\sigma^{-1}a_1\\
&= [d_{\frac{1}{2}}^+ d_0^-(1+\sigma^2)^2 - 8\lambda(1-\sigma^2) - 8j(1+\sigma^2)]w,\\
&8i(2j+1)(1+\sigma^2)\sigma^{-1}a_2\\
&= [d_{\frac{3}{2}}^+ d_1^-(1+\sigma^2)^2 - 8\lambda(1-\sigma^2) + 8(j+1)(1+\sigma^2)]w.
\end{aligned} \tag{41}$$

It is not obvious that exact solutions of (40) exist in terms of rational functions or commonly occurring higher transcendental functions. We turn therefore to a consideration of the properties of any solution at the origin and at infinity. The behavior at the origin is revealed by considering the indicial equation that arises when one assumes a solution for w in ascending powers of σ. If $w \sim \sigma^r$ as $\sigma \to 0$, the indicial equation gives $r = \pm 2j$, $\pm 2(j+1)$; since $j \neq 0$, these are all distinct. Applying (15a) to each of the functions w, a_1 and a_2, we find that only $r = 2j$ and $2(j+1)$ are admissible.

At infinity, we assume an expansion for w in decreasing powers of σ; and if $w \sim \sigma^{-q}$ as $\sigma \to \infty$, the corresponding indicial equation is

$$\begin{aligned}
&32\lambda(2\lambda+1)\\
&\qquad = [(q-4)^2 - (2j)^2][(q-4)^2 - (2j+2)^2], \qquad (42)
\end{aligned}$$

say,

$$F(\lambda) = F_j(q). \tag{43}$$

At infinity, a_1 and a_2 are both $\sim \sigma^{-q+1}$, so applying (15a) to them and to w gives $q > 3$, or, if the assumed more restrictive criterion (15b) applies, then $q > 4$. In Fig. 1 we graph $F(\lambda)$ and $F_j(q)$ against λ and q, respectively, for a typical value of j ($j = 3$). Since $\lambda > 0$, only the portions of the graphs in the first quadrant are of interest. Including also the restriction $q > 4$, only those parts of the graphs drawn as solid curves are relevant.

w satisfies a fourth-order differential equation. If for a given λ there are n corresponding allowed values of q at infinity, then there are $4 - n$ independent linear relations which the solutions at the origin will have to obey in order to satisfy the conditions at infinity. Now there are two allowed forms for the solution at the origin. Therefore, for $n = 3$, we can in general always find a solution for this λ. For $n = 2$ there will, in general, be only a discrete set of values of λ for which a satisfactory solution exists, and for $n = 1$, in general, no solution will be possible.

[22] For the fermion–fermion system actually discussed by Kummer, the signs for V and A interactions are different but for the fermion–antifermion system they are the same.

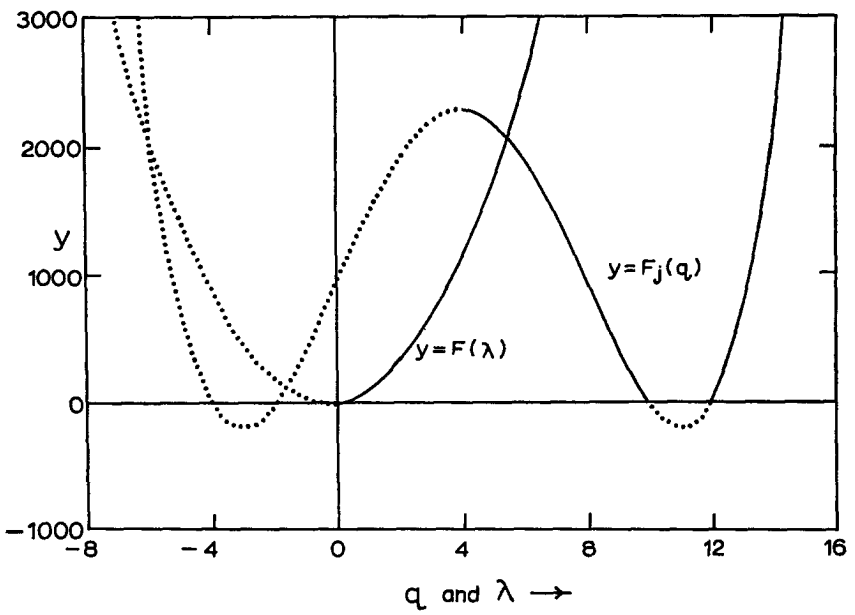


FIG. 1. The curves $y = F(\lambda)$ and (for $j = 3$) $y = F_j(q)$. The solid portions of the graphs are for ordinates where λ is positive and where $q > 4$.

From the graphs, we see that if $F(\lambda') = F_j(4)$, then for $0 < \lambda < \lambda'$ we can expect that a discrete set of values of λ may exist allowing a solution, and for $\lambda > \lambda'$ there will be no solution.

For large j, the upper bound λ' is given approximately by

$$\lambda' \simeq \tfrac{1}{2}j(j+1) - \tfrac{1}{4}. \tag{44}$$

For the vector interaction there is also a positive lower bound. In the inequality following Tiktopoulos' equation (7), replacing $e^2/(4\pi)^2$ with λ and setting $E = 0$, we find

$$\lambda > (8\pi)^{-1}. \tag{45}$$

If the restriction $b < -3$ can be relaxed to $b < -2$, then, if $F(\lambda'') = F_j(3)$, we see that for $\lambda'' < \lambda < \lambda'$ there are three acceptable values of q, and solutions could result for the whole continuum of this range of λ. We believe this to be an unlikely possibility.

It must be emphasized that it is not possible to state definitely with just the above considerations that acceptable solutions actually will occur in the way indicated. There is, for instance, the possibility of unique matching of a solution having a certain type of behavior at the origin with a solution having a certain type of behavior at infinity. This in fact happens with solutions of the P-sector differential equation in configuration space: All solutions are of the form $\rho^{-1}\mathfrak{L}_\zeta(\rho)$, where $\mathfrak{L}_\zeta$ is a modified Bessel function. If $\mathfrak{L}_\zeta$ is a K function, it has the correct behavior at infinity but not at the origin, and if it is an I function, the reverse is true for the range of ζ where ζ is real.

7. SOLUTIONS OF TYPE f^{SV}

When $j = 0$, the momentum-space radial equations obtained from Eq. (45) of I are

$$d_{\frac{1}{2}}^{+}d_0^{-}[(1-\sigma^2)s + 2i\sigma v] = \pm 16\lambda s,$$
$$d_2^{+}d_{\frac{1}{2}}^{-}[(1-\sigma^2)v + 2i\sigma s] = -8\lambda v. \tag{46}$$

Either s or v may be eliminated, leading (in general) to a fourth-order differential equation for the other variable. Applying relation (15a) to the behavior of s and v at the origin leads to the restriction $s \sim \sigma^2, \sigma^0$. Simultaneously with this, v can behave as σ^1 for either behavior of s, or another possibility is $v \sim \sigma^3$ when $s \sim \sigma^0$. There are thus two independent sets of solutions with suitable behavior at the origin.

At infinity we assume an expansion for s in decreasing powers of σ and, if $s \sim \sigma^{-q}$ as $\sigma \to \infty$, the corresponding indicial equation is

$$[(q-2)(q-4) \pm 16\lambda][(q-2)(q-6) - 8\lambda] = 0. \tag{47}$$

When the first factor vanishes, $v \sim \sigma^{-q-1}$ and, when the second factor vanishes, $v \sim \sigma^{-q+1}$. Applying (15b) then to both v and s at infinity, for the first factor to vanish we must have $q > 3$, and for the second factor to vanish we must have $q > 4$. Taking account of this restriction and also that $\lambda > 0$, we show with the solid portions of the graphs in Figs. 2 and 3 the simultaneous values of q and λ which satisfy (47) for a V and an A interaction, respectively.

Applying an argument similar to that used in Sec. 6, it follows that for a V interaction, a discrete spectrum

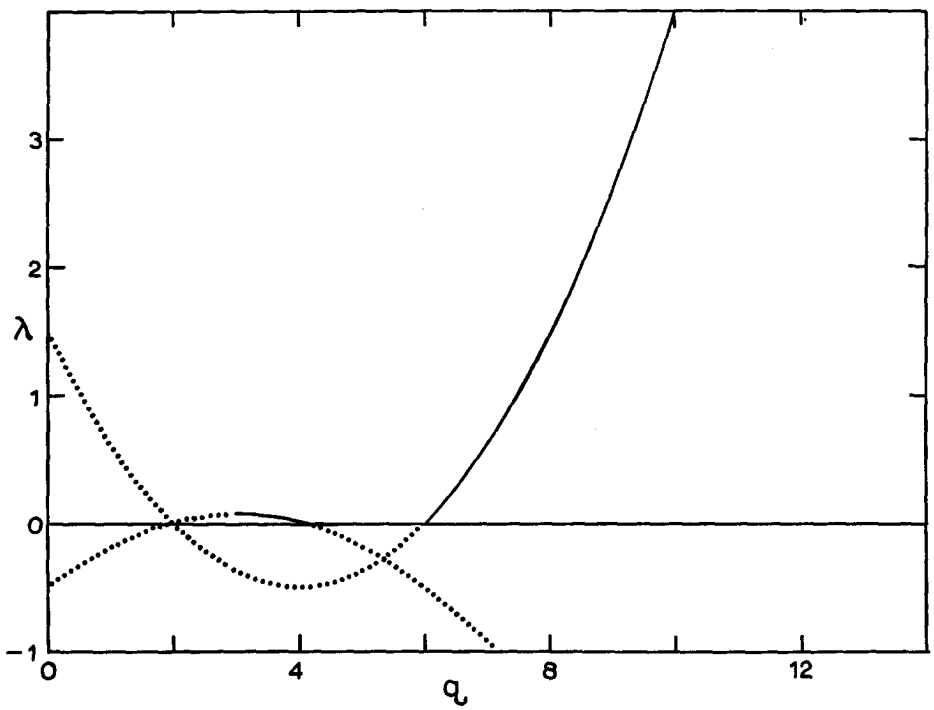


FIG. 2. A graph of Eq. (47) for a vector interaction. The solid portions show values of q for positive λ where S–V sector $j = 0$ formal solutions have an allowed asymptotic behavior.

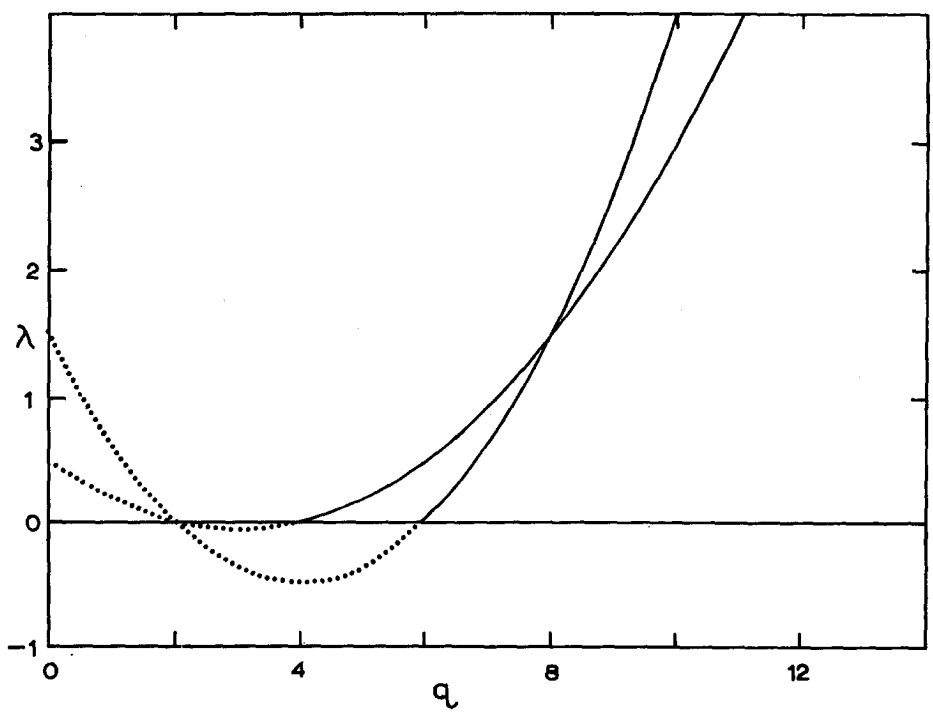


FIG. 3. A graph of Eq. (47) for an axial-vector interaction. The solid portions show values of q for positive λ where S–V sector $j = 0$ formal solutions have an allowed asymptotic behavior.

of λ could possibly be found for $0 < \lambda < \frac{1}{16}$, but for $\lambda \geq \frac{1}{16}$ there would be no acceptable solution. For an A interaction, a discrete spectrum is possible with no upper bound.

If negative values of λ are admitted, a special case arises when $\lambda = -\frac{1}{2}$. An exact solution of the differential equation for this value was found by Bertocchi *et al.*[23] for a V interaction. In our notation it is

$$s = d_{\frac{1}{2}}^{+} d_0^{-}\left(\frac{1}{1+\sigma^2}\right) = -\frac{8}{(1+\sigma^2)^3},$$
$$v = -i d_2^{+} d_{\frac{1}{2}}^{-}\left(\frac{\sigma}{1+\sigma^2}\right) = 4i\frac{\sigma(3+\sigma^2)}{(1+\sigma^2)^3}. \quad (48)$$

For this value of λ, the present author has found three other independent exact solutions of Eqs. (46), but neither they nor the forms (48) satisfy the criteria (15b).

When $j \neq 0$, the momentum-space radial equations equivalent to Eq. (43) of I are

$$d_{\frac{1}{2}}^{+} d_0^{-}[(1-\sigma^2)s + 2i\sigma(v_1+v_2)] = \pm 16\lambda s,$$
$$d_2^{+} d_{\frac{1}{2}}^{-}\{[(2j+1)-\sigma^2]iv_1 - 2(j+1)\sigma(i\sigma v_2 + s)\} = -8i(2j+1)\lambda v_1, \quad (49)$$
$$d_1^{+} d_{-\frac{1}{2}}^{-}\{[(2j+1)+\sigma^2]iv_2 - 2j\sigma(i\sigma v_1 + s)\} = -8i(2j+1)\lambda v_2.$$

Defining x by $s = d_{\frac{1}{2}}^{+} d_0^{-} x$, we have found a sixth-order differential equation for x. If $s \sim \sigma^{-q}$ as $\sigma \to \infty$, this equation leads to the indicial equation

$$[(q+2j-2)(q-2j-4) \pm 16\lambda][F(\lambda) - F_j(q)] = 0, \quad (50)$$

where $F(\lambda)$ and $F_j(q)$ are given by (42) and (43). The first factor in (50) is clearly the generalization of the first factor in (47). If q has a value which makes the first factor or the second factor in the left member of (50) vanish, then v_1 and v_2 together behave as σ^{-q-1} or as σ^{-q+1}, respectively. If we apply (15b), this means that $q > 3$ on the quadratic curve

$$(q+2j-2)(q-2j-4) \pm 16\lambda = 0 \quad (51)$$

and that $q > 4$ on the quartic curve (42). On both curves the limiting values ($q = 3$, $q = 4$) correspond to stationary values of λ. In the case of the quadratic curve, this value λ'' is given by

$$\lambda'' = \pm(2j+1)^2/16. \quad (52)$$

The properties of solutions of (49) near the origin can similarly be found by considering the sixth-order differential equation for x. Thus x has the independent types of behavior σ^r where $\frac{1}{2}r$ has the values $j+2$, $j+1$, j, $-j+1$, $-j$, and $-j-1$. The corresponding behavior of each member of the set $\{s, v_1, v_2\}$ obeys (15b) for each of the first three of these values, and the corresponding behavior of at least one member of the set $\{s, v_1, v_2\}$ does not obey (15b) for each of the other three of these values.

Using the arguments of Sec. 6, we can expect a range of λ to allow a continuous range of eigenvalues for which the set $\{s, v_1, v_2\}$ has acceptable behavior everywhere, a discrete range of eigenvalues, or no eigenvalue at all; this depends on finding in that range of λ a number n of acceptable values of q from (50) where, respectively, $n > 3$, $= 3$, < 3.

With a V interaction, a discrete spectrum is therefore possible for $0 < \lambda <$ (lesser of λ', λ''). When $j = \frac{1}{2}$, the upper bound is $\lambda'' = (13^{\frac{1}{2}} - 2)/8$; for all other values of j, $\lambda' < \lambda''$ and the upper bound is given by (52), i.e., it is $(2j+1)^2/16$. For λ greater than the upper bound, there will be no acceptable solution in general. The Tiktopoulos lower bound (45) must also apply in this sector.

With an A interaction, a discrete spectrum is possible for $0 < \lambda < \lambda'$, and for $\lambda > \lambda'$ there will be no acceptable solution in general.

8. SCALAR AND PSEUDOSCALAR INTERACTIONS

The same methods that we have applied in Secs. 6 and 7 can be used when the binding is due to the exchange of massless scalar or pseudoscalar particles. We quote the results.

A. f^{SV} Solutions

When $j = 0$, a discrete λ spectrum is possible in the range $0 < \lambda < \frac{1}{4}$ for a P interaction, and is possible for $0 < \lambda$ with no upper bound for an S interaction.

When $j \neq 0$, for both S and P interactions a discrete λ spectrum is possible in the range $0 < \lambda < \lambda'$, where λ' is the value corresponding to $q = 4$ on the relevant quartic curve. It is the positive root of

$$\lambda'(\lambda'+1) = j^2(j+1)^2 \quad (53)$$

and, for large j,

$$\lambda' \simeq j(j+1) - \tfrac{1}{2}. \quad (54)$$

B. f^{TA^-} Solutions

When $j = 0$, there is no acceptable solution for S or P interactions.

When $j \neq 0$, a discrete λ spectrum is possible for an S interaction in the range $0 < \lambda < \lambda'$, where λ' is the

[23] See Ref. 16. There appears to be a sign mistake in their equation (3.20), and the eigenvalue λ is negative. See also remarks on validity in Sec. 2.

positive root of

$$\lambda'(\lambda' - 1) = j^2(j+1)^2. \tag{55}$$

For large j,

$$\lambda' \simeq j(j+1) + \tfrac{1}{2}. \tag{56}$$

For a P interaction, there is again a discrete spectrum possible in the range $0 < \lambda < \lambda''$, where

$$\lambda'' = \tfrac{1}{4}(2j+1)^2. \tag{57}$$

The upper bound in this instance is determined by the quadratic curve. There is a distinctive feature of the f^{TA^-} solutions, common to both S and P interactions and resulting from the slightly different form the quartic curve takes as compared with other cases considered previously. Its equation is

$$\bar{F}(\lambda) \equiv 16\lambda(\lambda - 1) = [(q-4)^2 - (2j)^2] \times [(q-4)^2 - (2j+2)^2], \tag{58}$$

and the significant difference [e.g., from Eq. (42)] is the sign of the term linear in λ. Because $\bar{F}(\lambda)$ has a minimum value of -4 for $\lambda = \frac{1}{2}$, this means that for regions of the quartic graph having ordinates between -4 and 0, there are two positive values of λ which, if allowed eigenvalues, would produce the same types of behavior at infinity.

C. $g^{TA^\pm}$ Solutions

For all j, a discrete λ spectrum is possible for both S and P interactions in the range $0 < \lambda < \lambda'$, where

$$\lambda' = \tfrac{1}{4}(2j+1)(2j+3). \tag{59}$$

D. f^P, f^{TA^+}

There are no acceptable solutions of these forms for either S or P interactions.

9. CONCLUSIONS

The solutions f^P, f^{TA^+}, and $g^{TA^\pm}$ of the ladder-approximation differential Bethe–Salpeter equation for a spin-$\frac{1}{2}$ fermion–antifermion system bound to zero total mass by exchange of massless vector or axial vector particles do not satisfy a set of integrability criteria. But, barring uniformly unfavorable matching of the behaviors of either of the solutions f^{TA^-} or f^{SV} at the origin and infinity, we have shown that a discrete eigenvalue spectrum of the coupling constant λ will exist for solutions of these types satisfying the integrability criteria. A state of any angular momentum, parity, and charge parity can be selected by suitable choice of f^{TA^-} or f^{SV} and the quantum number j. ($j = 0$ is excluded for f^{TA^-}.) For both V and A interactions, upper bounds exist on the permissible values of λ except when $j = 0$ with an A interaction. These results appear to circumvent the "Goldstein problem." However, a word of caution is necessary. If we apply our method to the $g^{TA^\pm}$ equations for V and A interactions that have already been considered in Sec. 5, we find that a range $[0 < \lambda < \frac{1}{2}(j+1)^2]$ exists for which a discrete spectrum is possible. But this possibility is not realized since we know there is no spectrum in this case. This is, therefore, a case where we do have uniformly unfavorable matching of solutions which are acceptable at infinity with those unacceptable at origin, and vice versa. On the other hand, if the sign of λ is reversed, the range $0 < \lambda < \infty$ is possible on our argument, and this corresponds to Kummer's fermion–fermion A interaction system, where a spectrum with no upper bound does exist.

It is worth emphasizing that the criteria (15b) for acceptability of any formal solution lead in each case where there is a finite eigenvalue range possible to marginal exclusion of a continuum.

The method used in our investigation has been immediately applicable to a study of a fermion–antifermion system bound by exchange of massless scalar or massless pseudoscalar particles—the sixth-order equations that are involved being no more an obstacle than the one encountered in the S–V sector $j \neq 0$ case.

It seems reasonable to expect that the results found will carry over into the case where the exchange particles have nonzero mass, since solutions of the differential equations of type f^P, f^{TA^+}, etc., will still exist and since the configuration-space potential has the same singularity. Of course the numerical values of any discrete eigenvalues and of their upper bounds can be expected to depend on the ratio of the exchanged-particle mass to the mass of the binding fermion or antifermion.

When the total mass E is nonzero, there is in general no splitting into the S–V, T–A, and P Dirac-space sectors. If, however, we believe that the $E = 0$ solutions represent in some way the limit of the $E \neq 0$ solutions, then the probable existence of discrete eigenvalue spectra for λ for $E = 0$ solutions makes it hopeful that such spectra will exist for $E \neq 0$ solutions. Tiktopoulos's results for the vector interaction already foreshadow this.

It is encouraging that it appears unnecessary to introduce a high momentum cut-off.

ACKNOWLEDGMENT

I should like to thank Professor R. H. Dalitz for his interest in this work.

APPENDIX

In the (Wick-rotated) four-dimensional Euclidean configuration and momentum spaces the Bethe-Salpeter wavefunctions are related by the familiar Fourier transform integrals

$$f(p) = \int d^4x e^{-ip\cdot x} f(x),$$
$$f(x) = (2\pi)^{-4} \int d^4p e^{ip\cdot x} f(p). \tag{A1}$$

In an earlier paper (I) we introduced "bispherical" coordinates (R, ν, ω, ϕ) defined in terms of the usual four-dimensional Cartesian coordinates by

$$\begin{aligned} x_1 &= R \sin\nu \cos\phi, \\ x_2 &= R \sin\nu \sin\phi, \\ x_3 &= R \cos\nu \cos\omega, \\ x_4 &= R \cos\nu \sin\omega. \end{aligned} \tag{A2}$$

It is of some interest to examine the transformations (A1) in terms of bispherical coordinates in both configuration and momentum space. Denote the coordinates in momentum space by $(K, \nu', \omega', \phi')$.

A simultaneous eigenfunction of the operators $\mathbf{L}^2$ and $L_3^{\pm}$ in configuration space is [cf. Eq. (35) of I]

$$\begin{aligned} f(x) &= g_l(R) Z_{lm^+m^-}(\nu, \omega, \phi) \\ &= \bar{N} g_l(R) g_{lm^+m^-}(\nu) e^{im\phi} e^{im'\omega} \\ &= \bar{N} g_l(R) \sin^{|m|}\nu \cos^{|m'|}\nu \\ &\quad \times P_{l-\frac{1}{2}(|m|+|m'|)}^{(|m|,|m'|)}(\cos 2\nu) e^{im\phi} e^{im'\phi}, \end{aligned} \tag{A3}$$

where $m = m^+ + m^-$, $m' = m^+ - m^-$, and $\bar{N}$ is a normalization and phase factor.

In (A1),

$$d^4x = dR \cdot R^3 \, d\nu \sin\nu \cos\nu \, d\omega \, d\phi, \tag{A4}$$

$$\begin{aligned} p\cdot x = KR[&\sin\nu' \sin\nu \cos(\phi - \phi') \\ &+ \cos\nu' \cos\nu \cos(\omega - \omega')]. \end{aligned} \tag{A5}$$

Thus, from (A1),

$$\begin{aligned} f(p) = \bar{N} &\int_0^\infty dR R^3 g_l(R) \int_0^{\pi/2} d\nu \sin\nu \cos\nu g_{lm^+m^-}(\nu) \\ &\times \int_0^{2\pi} d\omega \exp(im'\omega) \\ &\times \exp[-iKR \cos\nu' \cos\nu \cos(\omega - \omega')] \\ &\times \int_0^{2\pi} d\phi \exp(im\phi) \\ &\times \exp[-iKR \sin\nu' \sin\nu \cos(\omega - \omega')]. \end{aligned} \tag{A6}$$

The integrations over ω and ϕ can be separately performed; they lead to

$$\begin{aligned} f(p) = (2\pi)^2 (-i)^{|m|+|m'|} e^{im\phi'} e^{im'\omega'} \bar{N} &\int_0^\infty dR R^3 g_l(R) \\ &\times \int_0^{\pi/2} d\nu \sin\nu \cos\nu g_{lm^+m^-}(\nu) \\ &\times J_{|m|}(KR \sin\nu' \sin\nu) J_{|m'|}(KR \cos\nu' \cos\nu). \end{aligned} \tag{A7}$$

From Bateman's expansion[24] and, alternatively, from Bailey's formulas,[25] we have been able to deduce the general result

$$\begin{aligned} &\int_0^{\pi/2} J_\alpha(S \sin\psi \sin\theta) \\ &\quad \times J_\beta(S \cos\psi \cos\theta) \sin^{\alpha+1}\theta \cos^{\beta+1}\theta P_n^{(\alpha,\beta)}(\cos 2\theta) \, d\theta \\ &= (-1)^n S^{-1} J_{\alpha+\beta+2n+1}(S) \sin^\alpha\psi \cos^\alpha\psi P_n^{(\alpha,\beta)}(\cos 2\psi). \end{aligned} \tag{A8}$$

This, with (A7), leads to

$$f(p) = g_l(K) Z_{lm^+m^-}(\nu', \omega', \phi'), \tag{A9}$$

where

$$K g_l(K) = (2\pi)^2 i^{-2l} \int_0^\infty dR R J_{2l+1}(KR) R g_l(R), \tag{A10}$$

and this is Eq. (11b).

[24] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge University Press, Cambridge, 1948), p. 370.
[25] W. N. Bailey, Quart. J. Math. Oxford Ser. 9, 141 (1938).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Approximations for the Frequency Spectrum of a Simple Lattice

E. M. BARODY
Battelle Memorial Institute, Columbus Laboratories, 505 King Avenue, Columbus, Ohio 43201

(Received 25 April 1969)

The simple cubic lattice with harmonic forces between nearest neighbors only is considered. Starting with the expression for the spectrum of squared frequency $G(x)$ as a Fourier transform of a product of Bessel functions, the asymptotic expansions about singular points are studied. For a range of values of the ratio of noncentral to central force constants σ, simple approximations are obtained which give quantitative information on the spectrum, not only very near the singular points, but over a substantial part of the frequency range. For $\sigma = \frac{1}{2}$, it has been found that a cusp at $x = \frac{1}{2}$ has been overlooked in previous work. In this case, essentially the entire spectrum is dominated by five singularities (at $x = 0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1$), and can be represented accurately by simple expressions. The approximations developed break down for very small σ, but appear to work well for $\frac{1}{4} \leq \sigma \leq 1$.

1. INTRODUCTION

An earlier paper[1] described calculations on the frequency spectrum and momentum autocorrelation function for a simple cubic lattice with harmonic forces between nearest neighbors only. For equal central- and noncentral-force constants, extensions of known analytical approximations were made. These led to simple expressions, which were surprisingly effective in yielding results obtained previously by predominantly numerical methods. Some study has now been made of how well similar approximations describe the frequency spectrum for other ratios of the force constants. Since these new calculations have succeeded better than might have been expected, a brief account of them is given here. As in the earlier work, the general background for the calculations may be found in the monograph of Maradudin, Montroll, and Weiss,[2] or in references cited there.

We take the ratio of noncentral- to central-force constants to be σ, and use Eq. (2.2) of I. In terms of $x = (\omega/\omega_L)^2$, the spectrum of squared frequency is then given by the Fourier transform

$$G(x) = (6/\pi)\int_0^\infty \cos 3\eta(1 - 2x)F(\eta)\, d\eta, \quad (1.1)$$

where

$$F(\eta) = J_0(\lambda\eta)J_0^2[(3 - \lambda)\eta/2]. \quad (1.2)$$

$J_0(z)$ is the zero-order Bessel function of the first kind, and $\lambda = 3/(1 + 2\sigma)$. In our considerations we emphasize the range $1 < \lambda < 3$, which corresponds to $0 < \sigma < 1$.

[1] E. M. Baroody, J. Math. Phys. **10**, 475 (1969). This paper is referred to as I.

[2] A. A. Maradudin, E. W. Montroll, and G. H. Weiss, *Theory of Lattice Dynamics in the Harmonic Approximation* (Academic Press Inc., New York, 1963).

2. BEHAVIOR OF THE SPECTRUM NEAR ITS SINGULAR POINTS

The location of the singular points of $G(x)$ and information on the function very near these points may be obtained from the asymptotic form of $F(\eta)$ for large η. In arriving at this information it is convenient to apply the theory of Fourier transforms as presented by Lighthill.[3] This may be done very directly, if we introduce the variable $u = (3/2\pi)(1 - 2x)$ and work in terms of the inverse transform

$$F(\eta) = \int_{-\infty}^{\infty} \exp(2\pi i u\eta) f(u)\, du, \quad (2.1)$$

where $f(u) = (\pi/3)G(x)$ is an even function of u. Suppose that in $f(u)$ there appears a pair of singular terms of the type

$$|u - b|^\alpha H(b - u) + |u + b|^\alpha H(u + b), \quad (2.2)$$

where α is real and nonintegral and $H(z)$ is the discontinuous function of Heaviside, which equals unity for $z > 0$ and vanishes for $z < 0$. According to Table 1 of Ref. 3 (p. 43), these terms imply a

[3] M. J. Lighthill, *Fourier Analysis and Generalised Functions* (Cambridge University Press, Cambridge, England, 1958). The procedure which we follow here differs from that of the earlier paper where the Fourier transform theory was used in a formal way in obtaining the asymptotic behavior of the momentum autocorrelation function from $G(x)$, but not in studying the singularities of $G(x)$ itself. The present procedure leads quickly to terms in $G(x)$ involving fractional powers of the distance from a singularity, but cannot yield terms involving integral powers, since these do not contribute to the asymptotic behavior of $F(\eta)$. In the earlier work an intuitive approach was used in the study of $G(x)$. This led to a term linear in $(1 - 3x)$ in Eq. (2.6) of I, a small term which proved to contribute consistently to the accuracy of the approximation. This may be seen very clearly by a comparison with numerical results published recently by Jelitto [R. J. Jelitto, J. Phys. Chem. Solids **30**, 609 (1969)]. For $(1 - 3x) = 0.005$, for example, Table 1 of Jelitto implies $G(\frac{1}{3}) - G(x) = 0.12834$, while Eq. (2.6) of I gives 0.12835. Without the linear term, the equation would give 0.12885. In the present work we also looked for linear terms (following plausible procedures), but found none which made helpful contributions.

corresponding term

$$\frac{2(\alpha!)}{|2\pi\eta|^{\alpha+1}}\cos[2\pi b\eta - \tfrac{1}{2}\pi(\alpha+1)] \qquad (2.3)$$

in the asymptotic expression for $F(\eta)$. We may, therefore, recognize singular terms like (2.2) which belong in $f(u)$ from an examination of

$$\begin{aligned}F(\eta) = {} & [2\lambda^{\frac{1}{2}}(3-\lambda)]^{-1}(2/\pi\eta)^{\frac{3}{2}} \\ & \times \{[\cos(3\eta - \tfrac{3}{4}\pi) - 2\cos(\lambda\eta + \tfrac{3}{4}\pi) \\ & - \cos[(3-2\lambda)\eta + \tfrac{3}{4}\pi]] + [8\lambda(3-\lambda)\eta]^{-1} \\ & \times [3(1+\lambda)\cos(3\eta - \tfrac{5}{4}\pi) \\ & + 2(3-\lambda)\cos(\lambda\eta + \tfrac{5}{4}\pi) + (5\lambda - 3) \\ & \times \cos[(3-2\lambda)\eta + \tfrac{5}{4}\pi]] + \cdots\}. \qquad (2.4)\end{aligned}$$

This expression follows from (1.2) and the asymptotic form of $J_0(z)$. We keep in mind, of course, that the expansion does not apply for $\lambda = 0$ or $\lambda = 3$, and that results based on it are likely to be poor for λ near these values.

Because the spectrum satisfies $G(x) = G(1-x)$, conclusions from the above relations may be stated for the domain $0 \le x \le \frac{1}{2}$. Comparison of the expression (2.3) with terms in (2.4) shows that, in specifying $G(x)$, three regions need to be distinguished:

A. $0 \le x \le x_1$,
B. $x_1 \le x \le x_2$,
C. $x_2 \le x \le \frac{1}{2}$,

where $x_1 = \frac{1}{6}(3-\lambda)$, and x_2 is the smaller of $\frac{1}{3}(3-\lambda)$ and $\frac{1}{3}\lambda$. The boundary points of *region A*, $x = 0$ and $x = x_1$, are always singular, the slope $G'(x)$ becoming infinite as these points are approached from within the region. The specification of a third singular point depends on λ. For $\lambda > \frac{3}{2}$, the slope becomes infinite as $x_2 = \frac{1}{3}(3-\lambda)$ is approached from *region C*. On the other hand, for $\lambda < \frac{3}{2}$, the slope becomes infinite as $x_2 = \frac{1}{3}\lambda$ is approached from *region B*.

To complete the general picture, comments on three special cases are needed. For $\lambda = \frac{3}{2}$, region C disappears and the singular point which bounds region B is $x = \frac{1}{2}$. There are a total of five singularities in the spectrum (rather than the usual six), and the central one is a double cusp, where the slope becomes infinite as $x = \frac{1}{2}$ is approached from either side. For $\lambda = 1$, one has the case treated in the earlier paper. Region B disappears and the singularities bounding region A are at $x = 0$ and $x = \frac{1}{3}$. For $\lambda = 3$ (vanishing noncentral forces), Eq. (1.1) involves a single Bessel function and $G(x) = [\pi x^{\frac{1}{2}}(1-x)^{\frac{1}{2}}]^{-1}$, the familiar result for a linear lattice.

With one exception, these results on the location of singular points confirm those published in 1953 by Rosenstock and Newell.[4] For $\lambda = \frac{3}{2}$ ($\sigma = \frac{1}{2}$) the sketch given in their Fig. 1 fails to show the singularity at the point corresponding to $x = \frac{1}{2}$.

As an illustration of the detailed use of Eqs. (2.2) and (2.3) to approximate $G(x)$ near singular points, consider the term in (2.4) which includes the factor $\cos[3\eta - (3/2\pi)]$ and corresponds to (2.3), with $b = 3/2\pi$ and $\alpha = \frac{1}{2}$. On using (2.2) and taking account of multiplying constants, we see that this term follows from

$$\begin{aligned}4[\pi^{\frac{1}{2}}\lambda^{\frac{1}{2}}(3-\lambda)]^{-1}[&|(3/2\pi) - u|^{\frac{1}{2}}\,H((3/2\pi) - u) \\ &+ |u + 3/2\pi|^{\frac{1}{2}}\,H(u + 3/2\pi)]\end{aligned}$$

in $f(u)$, or from

$$12\sqrt{3}[\pi^2\lambda^{\frac{1}{2}}(3-\lambda)]^{-1}[|x|^{\frac{1}{2}}\,H(x) + |1-x|^{\frac{1}{2}}\,H(1-x)]$$

in $G(x)$. In this expression, the first part gives the dominant term in the expansion of $G(x)$ near $x = 0$, while the second gives the corresponding term near $x = 1$. On working in this way with all the terms written explicitly in (2.4), we arrive at the following approximations for $G(x)$ in the regions A, B, and C:

Region A: near $x = 0$,

$$G(x) = \frac{12(3x)^{\frac{1}{2}}}{\pi^2\lambda^{\frac{1}{2}}(3-\lambda)}\left[1 + \frac{3(1+\lambda)x}{2\lambda(3-\lambda)} + \cdots\right], \qquad (2.5)$$

near $x = x_1 = \frac{1}{6}(3-\lambda)$,

$$\begin{aligned}G(x) = G(x_1) - {} & \frac{12\sqrt{2}(3-\lambda-6x)^{\frac{1}{2}}}{\pi^2\lambda^{\frac{1}{2}}(3-\lambda)} \\ & \times \left[1 - \frac{(3-\lambda-6x)}{12\lambda} + \cdots\right]; \qquad (2.6)\end{aligned}$$

Region B: $\lambda > \frac{3}{2}$, the slope $G'(x)$ bounded everywhere in the region, $\lambda \le \frac{3}{2}$, near $x_2 = \frac{1}{3}\lambda$,

$$\begin{aligned}G(x) = G(x_2) - {} & \frac{12(\lambda-3x)^{\frac{1}{2}}}{\pi^2\lambda^{\frac{1}{2}}(3-\lambda)} \\ & \times \left[1 - \frac{(5\lambda-3)(\lambda-3x)}{6\lambda(3-\lambda)} + \cdots\right]; \qquad (2.7)\end{aligned}$$

Region C: $\lambda > \frac{3}{2}$, just beyond $x_2 = \frac{1}{3}(3-\lambda)$,

$$\begin{aligned}G(x) = G(x_2) - {} & \frac{12(3x-3+\lambda)^{\frac{1}{2}}}{\pi^2\lambda^{\frac{1}{2}}(3-\lambda)} \\ & \times \left[1 - \frac{(5\lambda-3)(3x-3+\lambda)}{6\lambda(3-\lambda)} + \cdots\right], \qquad (2.8)\end{aligned}$$

$\lambda < \frac{3}{2}$, the slope $G'(x)$ bounded everywhere in the region.

[4] H. B. Rosenstock and G. F. Newell, J. Chem. Phys. **21**, 1607 (1953). We would like to thank Dr. Rosenstock for comments on the occurrence of the singularity at $x = \frac{1}{2}$. Once the question was raised, he could see from his methods that a cusp had been overlooked, but this omission had not come to his attention before. Dr. Rosenstock also kindly supplied numerical values of $G(x)$ for certain points in the spectrum for the case $\sigma = \frac{1}{4}$.

TABLE I. Coefficients β_j of Eq. (2.10) for two ratios of the force constants.

	σ			σ	
j	$\frac{1}{2}$	$\frac{1}{4}$	j	$\frac{1}{2}$	$\frac{1}{4}$
1	0.19444	0.20833	7	0.00063	0.00081
2	0.01597	0.02344	8	0.00049	0.00060
3	0.00360	0.00689	9	0.00039	0.00046
4	0.00167	0.00310	10	0.00031	0.00036
5	0.00112	0.00176	11	0.00025	0.00029
6	0.00083	0.00115	12	0.00021	0.00024

The leading terms in these expansions have been presented in a number of places in the literature, but less completely and explicitly than here. The second terms in the square brackets are probably new.

Equation (2.1) is equivalent to the expression involving the Laplace transform of $G(x)$ which was used in part of the calculations of the earlier paper. If η is replaced by $\frac{1}{6}i\eta$ and the variable x restored, (2.1) becomes

$$\exp(-\tfrac{1}{2}\eta)I_0(\tfrac{1}{6}\lambda\eta)I_0^2[\tfrac{1}{12}(3-\lambda)\eta] = \int_0^\infty \exp(-\eta x)G(x)\,dx, \quad (2.9)$$

which corresponds to Eq. (2.3) of I. This relation involving the zero-order modified Bessel function provides the most convenient basis for computing a substantial number of terms in the expansion of $G(x)$ near $x = 0$. When one wishes to determine $G(x)$ accurately over all of Region A, it is also very helpful (as was seen in I) to replace (2.5) with an expression which contains a correctly divergent derivative at $x = x_1 = \frac{1}{6}(3 - \lambda)$. That is, one uses

$$G(x) = \frac{12(3x)^{\frac{1}{2}}}{\pi^2\lambda^{\frac{1}{2}}(3-\lambda)}\left\{3\left[1 - \beta_1\left(\frac{x}{x_1}\right) - \beta_2\left(\frac{x}{x_1}\right)^2 - \cdots\right] - 2\left[1 - \frac{6x}{(3-\lambda)}\right]^{\frac{1}{2}}\right\}. \quad (2.10)$$

Numerical values of the coefficients β_j through $j = 12$ are listed in Table I for $\sigma = \frac{1}{2}$ and $\sigma = \frac{1}{4}$.

3. THE SPECTRA FOR $\sigma = \frac{1}{2}$ AND $\sigma = \frac{1}{4}$

For $\sigma = \frac{1}{2}$ ($\lambda = \frac{3}{2}$), the relationships developed above lead to a good description of $G(x)$ over the full range $0 \leq x \leq 1$. For Region A, Eqs. (2.6) and (2.10) work very well, as is shown by the comparison of several approximations given in Table II. With $G(\frac{1}{4})$ chosen to be 1.336 and only the first two terms in (2.6) applied, the two equations are in good agreement just below $x = x_1 = \frac{1}{4}$. In fact, $G(x)$ is easily obtained throughout the region to an accuracy of several parts per thousand.

TABLE II. Comparison of approximations for $G(x)$ in region A for $\sigma = \frac{1}{2}$.

	From Eq. (2.10)		From Eq. (2.6) with
x	Terms through $j = 6$	Terms through $j = 12$	$G(\frac{1}{4}) = 1.336$
0.20	0.820	0.819	0.823
0.22	0.937	0.936	0.939
0.23	1.012	1.011	1.012
0.24	1.110	1.108	1.107
0.25	1.345	1.341	1.336

It also turns out that the terms written explicitly in (2.7) lead to accurate values of $G(x)$ throughout Region B. With $\lambda = \frac{3}{2}$,

$$G(x) = G(\tfrac{1}{2}) - (4/\pi^2)(1-2x)^{\frac{1}{2}}(1+2x). \quad (3.1)$$

On the other hand, from (1.1),

$$G(\tfrac{1}{2}) = \frac{2(2\sigma+1)}{\pi\sigma}\int_0^\infty J_0(y/\sigma)J_0^2(y)\,dy. \quad (3.2)$$

For $\sigma \leq \frac{1}{2}$, this can be transformed to[5]

$$G(\tfrac{1}{2}) = [8(2\sigma+1)/\pi^3]K^2(k), \quad (3.3)$$

where $k^2 = \frac{1}{2}[1 - (1-4\sigma^2)^{\frac{1}{2}}]$ and $K(k)$ is the complete elliptic integral of the first kind. For $\sigma = \frac{1}{2}$, $G(\frac{1}{2}) = (16/\pi^3)K^2(2^{-\frac{1}{2}}) = 1.774$. Equation (3.1) should be an accurate approximation near $x = \frac{1}{2}$ and become poorer as $x = \frac{1}{4}$ is approached. That it is actually rather good throughout the region is shown by a test at $x = \frac{1}{4}$. One finds that $G = 1.774 - (3\sqrt{2}/\pi^2) = 1.344$, only six parts per thousand greater than the value obtained above. The conclusion that $G(x)$ is well described for all x is further supported by an evaluation of the normalization integral which gives 1.001. The complete curve for $G(x)$ as determined by our equations is given in Fig. 1.

For $\sigma = \frac{1}{4}$ ($\lambda = 2$), the approximations based on expansions around the singular points provide a surprising amount of quantitative information about $G(x)$, although they do not explicitly cover Region B. Equations (2.6) and (2.10) again work well for Region A and lead to $G = 1.37$ for $x = x_1 = \frac{1}{6}$. In Region C, Eq. (2.8) gives

$$G(x) = G(\tfrac{1}{3}) - \frac{2^{\frac{1}{2}}}{2\pi^2}(3x-1)^{\frac{1}{2}}(19-21x), \quad (3.4)$$

which should be accurate just beyond $x = x_2 = \frac{1}{3}$. It is not quite correct throughout the region, of

[5] See W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer-Verlag, New York, 1966), 3rd ed., pp. 103, 370. The statement on p. 103 would support (3.3) only for $\sigma < \frac{1}{2}$, but further consideration showed that extension to the closed interval is justified.

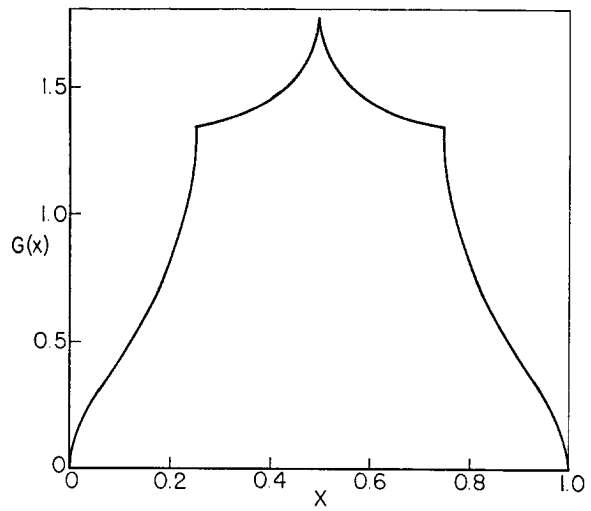


FIG. 1. The spectrum of squared frequency $G(x)$ for $\sigma = \frac{1}{2}$ as given by Eqs. (2.6), (2.10), and (3.1).

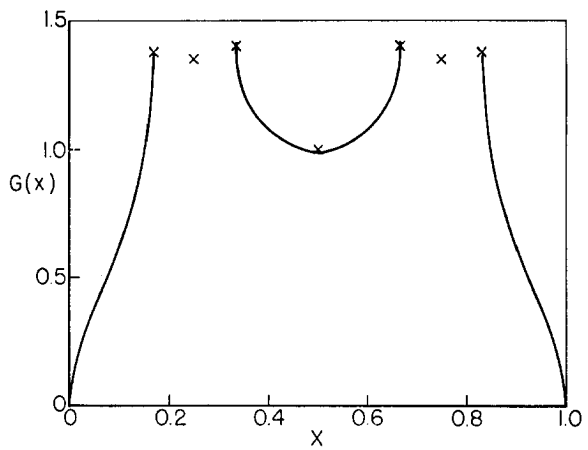


FIG. 2. The spectrum of squared frequency $G(x)$ for $\sigma = \frac{1}{4}$. The lines follow from Eqs. (2.6), (2.10), and (3.4). The crosses show the values of Rosenstock and Newell.

course, as is emphasized by a nonzero derivative at $x = \frac{1}{2}$. In view of our experience with the case $\sigma = \frac{1}{2}$, however, there is reason to believe that it is a useful approximation even near $x = \frac{1}{2}$. It yields $G(\frac{1}{3}) - G(\frac{1}{2}) = 17/4\pi^2 = 0.431$. Combining this with $G(\frac{1}{2}) = 0.988$, as calculated from (3.3), gives $G(\frac{1}{3}) = 1.419$.

This case is the only one which Rosenstock and Newell[4] treated numerically. Their most relevant results are $G(\frac{1}{6}) = 1.38$, $G(\frac{1}{3}) = 1.41$, and $G(\frac{1}{2}) = 1.00$, which are in remarkably good agreement with the values calculated above. In Fig. 2, the lines show $G(x)$ as given by our equations for Regions A and C, while the crosses are the values of Rosenstock and Newell.

The fact that (3.1) and (3.4) give good results throughout the appropriate regions of x is unexpected and, in part, fortuitous. Even so, when considered along with the earlier work for $\sigma = 1$, these results indicate that, for $\frac{1}{4} \leq \sigma \leq 1$, $G(x)$ can be rather accurately approximated by simple expressions arising from the singularities—not just very near the singular points themselves, but over a large part of the range $0 \leq x \leq 1$. Since our approximations use the asymptotic expansion of $J_0(z)$ in all three factors in $F(\eta)$, they should fail for small σ (λ near 3). For $\sigma = \frac{1}{8}$, a case which has been treated by Montroll,[6] they, in fact, give a very poor value of $G(x_2) - G(\frac{1}{2})$.

ACKNOWLEDGMENTS

The author has valued the interest and encouragement of associates at Battelle and is particularly indebted to Dr. A. C. Beer, Dr. E. J. Drauglis, and Dr. V. E. Wood for suggestions concerning the work and the preparation of the manuscript. He has also greatly appreciated the comments and information supplied by Dr. H. B. Rosenstock of the Naval Research Laboratory.

[6] E. W. Montroll, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* (University of California Press, Berkeley, Calif., 1956), Vol. III, p. 209.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Calculation of High-Field Distribution Functions in Semiconductors

M. O. VASSELL
The Bayside Laboratory, Research Center of General Telephone & Electronics Laboratories, Incorporated, Bayside, New York

(Received 24 August 1968; Revised Manuscript Received 15 May 1969)

The relation between various methods proposed recently for calculating the hot-carrier distribution function in semiconductors is discussed. In particular, it is shown that the method deduced heuristically by Rees from considerations of the stability of the steady state is an iterative prescription for solving a suitably chosen integral form of the Boltzmann equation. It is shown that this method is essentially an adaptation of Kellogg's method with additional sufficiency conditions imposed to guarantee the existence of a positive solution and the convergence of the iterative process. The essential ingredient of these conditions is that the kernel of the integral equation be positive. It is further pointed out that the self-scattering process introduced by Rees belongs to a larger class of operators that ensure the required positivity of the kernel.

I. INTRODUCTION

In a recent publication, Rees[1] formulated an iterative method for solving the hot-carrier problem which is, in principle, applicable to any combination of scattering mechanisms. The underlying basis for this approach, it was claimed, was an appeal to the stability of the steady state. It is the purpose of this paper to show how these results can be derived from the Boltzmann equation, and in what way this work is related to other formulations using the integral equation method. As we shall discover, the common feature of these formulations is that they lead to Fredholm equations of the second kind, for which there are well-known procedures for solution. Although we shall cite cases involving inhomogeneous as well as homogeneous equations, we shall discuss the details of a technique for homogeneous equations only, since this method appears to be the most effective for numerical calculations.

In Sec. II we discuss the general basis for the integral equation method, and cite some examples of this approach. Sufficiency conditions for a unique, positive solution of the homogeneous Fredholm equation are stated, and a discussion of the convergence of an iterative method for obtaining this solution is given in Sec. III. We translate these conditions into restrictions useful for the Boltzmann equation in Sec. IV by invoking the properties of this scattering operator. In the last section we show that the Rees method results naturally from this formalism.

II. THE INTEGRAL EQUATION METHOD

Of the many and varied methods for solving the hot-carrier problem for arbitrary scattering processes, the most general have proposed numerical solutions of some suitably chosen integral form of the Boltzmann equation:

$$\frac{e}{\hbar}\mathbf{E}\cdot\nabla_{\mathbf{k}}f(\mathbf{k}) = \hat{C}f = \int d\mathbf{k}'[f(\mathbf{k}')S(\mathbf{k}',\mathbf{k}) - f(\mathbf{k})S(\mathbf{k},\mathbf{k}')]. \quad (1)$$

As written, this equation governs the steady-state distribution of carriers of crystal momentum $\mathbf{k}$ for the case of uniform electric field $\mathbf{E}$, spatial homogeneity, and uniform temperature. The Fermi factors have been replaced by unity because we are studying semiconductors of low enough carrier concentrations so that Maxwell–Boltzmann statistics apply. As will become clear, however, the ideas presented here may be extended at will to more general cases. The collision integral is characterized by $S(\mathbf{k},\mathbf{k}')$, the rate of transition between states $\mathbf{k}$ and $\mathbf{k}'$. The procedure, most generally, is then to add to both sides of Eq. (1) terms that are linear in f and its gradient, and so obtain

$$\frac{e}{\hbar}[\mathbf{E}+\boldsymbol{\epsilon}(\mathbf{k})]\cdot\nabla_{\mathbf{k}}f + \alpha(\mathbf{k})f = \tilde{C}f, \quad (2)$$

where $\tilde{C}f$ represents the modified collision operator. The required integral form of the Boltzmann equation is obtained by integrating (2) along the characteristic curve of the partial differential operator, that is, by averaging over the collision-free trajectory of a particle drifting in the force field $e[\mathbf{E}+\boldsymbol{\epsilon}]$. Finally, the procedure for solution of the resulting integral equation is an iterative one, starting from some zeroth-order approximant that is dictated in general by the choice of $\boldsymbol{\epsilon}$ and the generalized relaxation time $\alpha^{-1}(\mathbf{k})$. For the procedure to be practical, of course, convergence must be rapid and the numerical work tractable.

[1] H. D. Rees, Phys. Letters **26A**, 416 (1968).

One proposal[2] has been to pick $\epsilon = 0$ and

$$\alpha(\mathbf{k}) = \int d\mathbf{k}' S(\mathbf{k}, \mathbf{k}') \equiv \lambda(\mathbf{k}),$$

for which choice $\tilde{C}f$ is the in-scattering contribution to the collision integral. In this case an inhomogeneous equation for the departure from equilibrium is obtained. The most convenient choice for the zeroth approximant is the relaxation-time solution of the Boltzmann equation that enters as the inhomogeneous term.

Another possibility[3] is to choose α and ϵ from the zeroth and first moments of the collision operator with respect to the momentum transfer $\mathbf{q} = \mathbf{k}' - \mathbf{k}$. In specific form this amounts to expanding $f(\mathbf{k}')/f_0(\mathbf{k}')$ to two terms in a Taylor series about $\mathbf{k}$, f_0 being the equilibrium solution. Then α is proportional to $(\mathbf{k} \cdot \mathbf{E})/E$, and ϵ lies in the $\mathbf{k}$ direction. In this case a local solution of Eq. (2) was proposed as the starting solution for iteration.

One situation where an appropriate choice for α and ϵ leads to a one-dimensional integral equation well suited to accurate computer calculations was pointed out by Budd.[4] It occurs when, as is the case for nonpolar optical scattering, the interaction matrix element is spherically symmetric, although the process itself may be inelastic. If one sets $\epsilon = 0$ and $\alpha^{-1} = \tau(\epsilon)$, the energy-dependent relaxation time with which the *entire* anisotropic part of the distribution relaxes, the modified collision term can be shown to depend only on the isotropic part of the distribution function. As a consequence of this simple dependence, the resulting homogeneous equation directly relates the entire distribution to its isotropic part, S_0. A one-dimensional homogeneous Fredholm equation for S_0, together with relations for the angular parts as quadratures involving S_0, can be immediately obtained by projection.

Although in principle straightforward, the above considerations could lead to a time-consuming numerical computation. In practice, therefore, it is useful to develop criteria for picking the optimum form of Eq. (2) from the limitless class available. All such criteria are ultimately based on certain desirable properties of the kernel of the integral equation, and in general are of two complimentary types. The first type deals with simplicity in form of the kernel. The collision integral itself may possess special properties, as in the last example, but even for general collision operators it is possible to find a form (2) that makes the kernel the simplest possible to handle numerically. Criteria of the second type are usually dictated by the method of solution used, and are especially important when an *iterative solution* of a homogeneous integral equation is contemplated. The reason is that such schemes are *inherently* designed to solve eigenvalue problems, and could converge to different eigensolutions, depending on the properties of the kernel, and the zeroth iterate chosen. However, we shall show in the next section that, by a judicious choice of separation (2), it is possible to guarantee convergence to the Boltzmann solution by iterating from any arbitrary positive function. The question might well be asked why it is necessary to consider eigenvalue problems, when in fact the transformed Boltzmann equation leads to an equation with known eigenvalue, whose eigenfunction one seeks. The answer is that alternative methods available (such as variational methods, approximation of the kernel by a degenerate kernel, and expansion of f in a complete set of functions) require in turn special intuition about the nature of the expected solution, detailed knowledge of the scattering rates in $\mathbf{k}$-space, or a reasonable means of truncating an infinite system of algebraic equations. The iterative method requires only the ability to accurately compute integrals in $\mathbf{k}$-space.

III. THE MODIFIED KELLOGG METHOD

In this section we discuss Kellogg's iterative method for solving the homogeneous integral equation,

$$f(\mathbf{k}) = \mu \int d\mathbf{k}' K(\mathbf{k}', \mathbf{k}) f(\mathbf{k}), \tag{3}$$

its convergence properties, and the conditions on K and the zeroth iterate that are required in order to extract the desired solution. Accordingly, let us choose an initial approximation θ_0 and calculate successive approximations by the prescription

$$\theta_n(\mathbf{k}) = \int d\mathbf{k}' K(\mathbf{k}', \mathbf{k}) \theta_{n-1}(\mathbf{k}') \tag{4}$$

for $n \geq 1$, the aim being to repeat the process until satisfactory convergence is achieved. Further, let $\varphi_1, \varphi_2, \cdots, \varphi_n, \cdots$ be the eigenfunctions of K and let $\mu_1, \mu_2, \cdots, \mu_n, \cdots$ (with $|\mu_i| < |\mu_j|$, for $i < j$) be the corresponding eigenvalues. It can be shown, following Kellogg,[5] that, if φ_r is the *first* of the eigenfunctions to

[2] H. F. Budd, J. Phys. Soc. Japan **18**, 142 (1963).
[3] M. O. Vassell and J. K. Percus, Bull. Am. Phys. Soc., Ser. II, **12**, 381 (1967).
[4] H. F. Budd, Phys. Rev. **158**, 798 (1967); J. Phys. Soc. Japan **21**, 420 (1966).
[5] See, for example, S. G. Miklin, *International Series of Monographs on Pure and Applied Mathematics. Integral Equations* (Pergamon Press, Inc., New York, 1964), Vol. 4, p. 94.

which θ_0 is *not orthogonal*, the sequence of approximations (4) converges to φ_r.[6] For this reason the essential difficulty with this method is that, if one does not specify *a priori* whether θ_0 is orthogonal to a certain eigenfunction, it remains unknown which of the eigenfunctions and eigenvalues will result. To circumvent this difficulty, we may argue as follows. It is clear that, if we know that θ_0 is not orthogonal to φ_1, the smallest eigenvalue and corresponding eigenfunction will be determined. To adapt procedure (4) to our purposes, what has to be done is twofold: First, we must guarantee that the solution of the Boltzmann equation is φ_1, i.e., that the lowest eigenvalue is $\mu = 1$; secondly, we must pick for θ_0 a function that is not orthogonal to φ_1, but at once orthogonal to all φ_n, $n > 1$.

To fulfill these requirements, the assumed positivity of the Boltzmann solution proves useful. According to a theorem of Jentzsch,[7] later refined and restated by Krein and Rutman,[8] if the kernel K is nonnegative and continuous except for jump discontinuities, then Eq. (3) has a *unique positive* eigenfunction, and this eigenfunction corresponds to the smallest of all the eigenvalues of Eq. (3).[9] This means that, if a choice for α in Eq. (2) is made so that the kernel K satisfies the conditions of Jentzsch's theorem, any method, iterative or otherwise, for calculating the eigenfunction possessing the lowest eigenvalue determines the Boltzmann solution. For the method of interest to us it is sufficient to pick θ_0 to be any arbitrary *positive* function.

Some of the formal arguments presented in the foregoing may be clarified by the following intuitive reasoning. If we suppose that the kernel K in Eq. (3) is obtained by inverting Eq. (2) with $\epsilon = 0$ but α chosen arbitrarily, then, by operating on both sides of Eq. (3) with $(e/\hbar)\mathbf{E}\cdot\nabla_{\mathbf{k}}$, we obtain

$$\frac{e}{\hbar}\mathbf{E}\cdot\nabla_{\mathbf{k}}f + \alpha(\mathbf{k})f = \mu\tilde{C}f = \mu[\hat{C}f + \alpha(\mathbf{k})f].$$

Since the collisions conserve particles, and $f \to 0$ sufficiently strongly as $\mathbf{k} \to \infty$ to be normalizable, further integration of this equation over all $\mathbf{k}$ yields the condition

$$(\mu - 1)\int d\mathbf{k}\alpha(\mathbf{k})f(\mathbf{k}) = 0.$$

This implies that, provided α and f are not mutually orthogonal, then $\mu = 1$ is an eigenvalue of K. Conversely, for any eigenvalue other than $\mu = 1$, the corresponding eigenfunction must be orthogonal to α. Now the solution of Eq. (1), when it exists, corresponds to $\mu = 1$ and is by hypothesis positive. Thus, when α is chosen to be of fixed sign, the Boltzmann solution is the only positive eigenfunction of the kernel K. It must be emphasized, however, that this condition is insufficient to guarantee that $\mu = 1$ is the smallest eigenvalue. In order to satisfy the latter condition, the class of α's that leads to an unambiguous determination of the Boltzmann solution via Eq. (4) must be further restricted by the conditions of Jentzsch's theorem

IV. PROPERTIES OF THE BOLTZMANN KERNEL

In order to take full advantage of the results of the last section, it is necessary to show under what conditions our kernels will satisfy the conditions of Sec. III. In general, the scattering rate is written for a combination of processes belonging to two classes, elastic and inelastic.[10] For processes that conserve electron energy, the *principle of detailed balancing* demands that the transition rate be symmetric, i.e., $S(\mathbf{k}', \mathbf{k}) = S(\mathbf{k}, \mathbf{k}')$. For inelastic processes, S is not symmetric, but by detailed balancing satisfies the relation

$$e^{-\epsilon_{\mathbf{k}'}/k_BT}S(\mathbf{k}', \mathbf{k}) = e^{-\epsilon_{\mathbf{k}}/k_BT}S(\mathbf{k}, \mathbf{k}'), \tag{5}$$

by virtue of which it is immediately symmetrizable. Furthermore, since S is essentially proportional to the square of a matrix element, it is positive.

With these properties, Jentzsch's theorem may be immediately applied to the equilibrium case. By introducing the redefinitions

$$f(\mathbf{k}) = e^{-\epsilon_{\mathbf{k}}/k_BT}g(\mathbf{k}),$$
$$\Lambda(\mathbf{k}', \mathbf{k}) = e^{-\epsilon_{\mathbf{k}'}/k_BT}S(\mathbf{k}', \mathbf{k}) = \Lambda(\mathbf{k}, \mathbf{k}'),$$

we may rewrite Eq. (1) to read

$$\int d\mathbf{k}'\Lambda(\mathbf{k}', \mathbf{k})[\mu g(\mathbf{k}') - g(\mathbf{k})] = 0. \tag{6}$$

In this case it is clear that the only positive solution is $g(\mathbf{k}) = \text{const}$, corresponding to the eigenvalue $\mu = 1$.

[6] As originally stated, the proof of convergence is given for symmetrical kernels and depends critically on the existence of a series representation for θ_1 in terms of the eigenfunctions of K. However, a rigorous proof for symmetric kernels, avoiding the use of eigenfunction expansions, can be found in F. Reisz and B. Sz.-Nagy, *Functional Analysis* (Frederick Ungar Publ. Co., New York, 1955), p. 240. A similar treatment for nonsymmetric, completely continuous operators is given by H. F. Buckner, "Numerical Methods for Integral Equations" in *Survey of Numerical Methods* (McGraw-Hill Book Co., New York, 1962), p. 454.

[7] R. Jentzsch, J. Reine Angew. Math. **141**, 235 (1912).

[8] M. G. Krein and M. A. Rutman, *Functional Analysis and Measure Theory* (American Mathematical Society, Providence, R.I., 1962), Transl. Series 1, Vol. 10, p. 272.

[9] A more detailed statement of these matters is given in the Appendix.

[10] See, for example, A. C. Smith, T. F. Janak, and R. B. Adler, *Electronic Conduction in Solids* (McGraw-Hill Book Co., New York, 1967), Chap. 7.

In order to discuss the general steady-state problem, let us first deduce the equivalent integral form for Eq. (2). We restrict ourselves without loss of generality to the choice $\epsilon = 0$ and α arbitrary. Consider then the equation

$$\frac{e}{\hbar}\mathbf{E}\cdot\nabla_{\mathbf{k}}f(\mathbf{k}) + \alpha(\mathbf{k})f = \tilde{C}f \equiv \int d\mathbf{k}'T(\mathbf{k}',\mathbf{k})f(\mathbf{k}'), \quad (7)$$

where the modified collision operator is

$$T(\mathbf{k}',\mathbf{k}) = S(\mathbf{k}',\mathbf{k}) + [\alpha(\mathbf{k}) - \lambda(\mathbf{k})]\delta(\mathbf{k}-\mathbf{k}'), \quad (8)$$

with

$$\lambda(\mathbf{k}) = \int d\mathbf{k}''S(\mathbf{k},\mathbf{k}''). \quad (9)$$

Now, in terms of the integrating factor e^F, where F is defined by

$$\frac{e}{\hbar}\mathbf{E}\cdot\nabla_{\mathbf{k}}F = \alpha(\mathbf{k}), \quad (10)$$

Eq. (7) reduces to

$$\frac{e}{\hbar}\mathbf{E}\cdot\nabla_{\mathbf{k}}[e^F f] = e^F\tilde{C}f. \quad (11)$$

Since

$$\int_0^\infty dt e^{-pt} = p^{-1}, \quad \text{for Re } p > 0,$$

the formal solution of Eq. (10) is given by

$$F(\mathbf{k}) = \left\{\int_0^\infty dt \exp\left[-\frac{et}{\hbar}\mathbf{E}\cdot\nabla_{\mathbf{k}}\right]\right\}\alpha(\mathbf{k})$$
$$= \int_0^\infty dt\alpha\left(\mathbf{k} - \frac{et}{\hbar}\mathbf{E}\right), \quad (12)$$

where use has been made of the translation operator. The arbitrary constant at our disposal has been chosen for convenience to be $F(\infty) = 0$, and is of no significance to the final result. Thus, by a similar manipulation on Eq. (11), by applying the result (12), and by inserting the initial condition $f(\mathbf{k}) \to 0$ for $\mathbf{k} \to \infty$, we obtain the required integral form

$$f(\mathbf{k}) = \int d\mathbf{k}'f(\mathbf{k}')\int_0^\infty ds \exp\left[-\int_0^s \alpha\left(\mathbf{k} - \frac{et}{\hbar}\mathbf{E}\right)dt\right]$$
$$\times T\left(\mathbf{k}';\mathbf{k} - \frac{es}{\hbar}\mathbf{E}\right)$$
$$\equiv \int d\mathbf{k}'f(\mathbf{k}')K(\mathbf{k}',\mathbf{k}). \quad (13)$$

It is apparent that K is positive if T is nonnegative. Since λ is by definition positive, being the integral of the positive function S, this is the case, for example, when $\alpha(\mathbf{k}) > \lambda(\mathbf{k})$, although less restrictive conditions can be stated from a knowledge of the details of the scattering rates. It is interesting to observe that the quantity $S(\mathbf{k}) \equiv \alpha(\mathbf{k}) - \lambda(\mathbf{k})$ is precisely the "self"-scattering rate introduced by Rees to simplify the numerical analysis, but that in our terms of reference S has the more essential role of ensuring positivity. Finally, one can easily show from Eqs. (8) and (13) that K is continuous except for a jump discontinuity when $k_\parallel = k'_\parallel$, where $k_\parallel$ is the component of $\mathbf{k}$ along $\mathbf{E}$.

We have shown therefore that the solution of the Boltzmann equation (7), with suitably chosen $\alpha(\mathbf{k})$, can be obtained by a convergent iterative process of the form (4), provided the zeroth iterate is positive.

V. REES'S METHOD

Let us now apply Eqs. (4) to the integral equation (13). We are at liberty to introduce for an appropriate θ_0 the sequence of iterated kernels $P_n(\mathbf{k}',\mathbf{k})$ defined by

$$\theta_n(\mathbf{k}) = \int d\mathbf{k}'\theta_0(\mathbf{k}')P_n(\mathbf{k}',\mathbf{k}). \quad (14)$$

The set of recurrence relations these satisfy from Eq. (4), i.e.,

$$P_n(\mathbf{k}',\mathbf{k}) = \int d\mathbf{k}''P_{n-1}(\mathbf{k}',\mathbf{k}'')K(\mathbf{k}'',\mathbf{k}), \quad (15)$$

is essentially Rees's result, except that K is further separated into two factors according to

$$K(\mathbf{k}'',\mathbf{k}) = \int d\mathbf{k}'''T(\mathbf{k}'',\mathbf{k}''')P_0(\mathbf{k}''',\mathbf{k}). \quad (16)$$

One may easily verify from Eq. (13) the following form for P_0:

$$P_0(\mathbf{k}',\mathbf{k}) = \frac{\hbar}{eE}\exp\left[-\int_0^\tau \alpha\left(\mathbf{k} - \frac{e}{\hbar}t\mathbf{E}\right)dt\right]$$
$$\times \delta(k_x - k'_x)\delta(k_y - k'_y)\eta(\tau). \quad (17)$$

Here $\tau = h(k_z - k'_z)/eE$, the time of flight of a particle drifting in the field $\mathbf{E}$ from $\mathbf{k}'$ to $\mathbf{k}$, and η is the unit step function:

$$\eta(x) = 1, \quad x > 0,$$
$$= 0, \quad x < 0.$$

When α is the set equal to λ, the probability of scattering out of $\mathbf{k}$ per unit time, P_0, measures the probability that no scattering occurs during this motion. Although no such physical meaning can be attributed to P_0 when $\alpha \neq \lambda$, it is, however, interesting to note that P_0 is formally a Green's function defined by

$$\left[\frac{e}{\hbar}\mathbf{E}\cdot\nabla_{\mathbf{k}} + \alpha(\mathbf{k})\right]P_0(\mathbf{k}',\mathbf{k}) = \delta(\mathbf{k}-\mathbf{k}'). \quad (18)$$

This formal property might have been anticipated from the fact that the transformation from Eqs. (7)–(13) can be achieved equivalently via Green's functions.

We have thus demonstrated that Rees's method is essentially an adaptation of Kellogg's method to the Boltzmann equation. We further remark that his self-scattering rate is one of a number of means for separating the Boltzmann equation in such a way that the equivalent integral equation has a positive kernel. It must also be emphasized that, whenever an iterative procedure of the type (4) is used to solve the Boltzmann equation, it is essential to verify that the kernel satisfies the conditions of Jentzsch and that the zeroth iterate is positive. In this connection it is interesting to note that one of the virtues of Budd's method[4] is that these conditions are built into the theory. Finally, other methods for calculating the smallest eigenvalue might prove useful for future study of the Boltzmann equation.

ACKNOWLEDGMENTS

The author is grateful to E. M. Conwell for valuable suggestions and a critical reading of the manuscript, and to J. F. Wagner and D. K. McGraw for many stimulating discussions. The author is further indebted to H. F. Budd for pointing out the relevance of the arguments in the last paragraph of Sec. III.

APPENDIX

The theorem quoted below is usually stated for a finite one-dimensional domain, but may be immediately generalized to an infinite three-dimensional domain. It is a restatement and generalization by Krein and Rutman[6] of a theorem of Jentzsch[5] on the uniqueness of a positive solution of the homogeneous Fredholm equation.

Consider the integral equation

$$\varphi(s) = \mu \int_a^b K(s, t)\varphi(t)\, dt. \tag{A1}$$

If the kernel $K(s, t)$ is nonnegative and satisfies the following conditions for $a \leq s \leq b$:

(i) $K(s, t)$ is measurable and summable with respect to t;

(ii) $\lim_{h \to 0} \int_a^b |K(s + h, t) - K(s, t)|\, dt = 0$;

(iii) there is an iterated kernel $K^{(N)}(s, t)$ such that

$$\int_a^b K^{(N)}(s, t)\varphi(t)\, dt > 0$$

for every nonnegative continuous function $\varphi(s)$, then Eq. (A1) has a *unique positive* eigenfunction, and the transposed equation

$$\varphi(s) = \lambda \int_a^b K(t, s)\varphi(t)\, dt \tag{A2}$$

has a unique almost-everywhere-positive solution corresponding to the same positive eigenvalue. The eigenvalue is the smallest of all the characteristic numbers of Eq. (A1).

The conditions (i) and (ii) are necessary and sufficient that the operator associated with the kernel K be completely continuous. The fact that continuity of the kernel itself is not required is of special importance to us.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Ising-Model Spin Correlations on the Triangular Lattice. III. Isotropic Antiferromagnetic Lattice*

JOHN STEPHENSON
Theoretical Physics Institute, University of Alberta, Edmonton, Alberta, Canada

(Received 20 February 1969; Revised Manuscript Received 5 June 1969)

The asymptotic behavior of the pair correlation $\omega_2(\mathbf{r}) = \langle\sigma_0\sigma_\mathbf{r}\rangle$ between two spins at sites $\mathbf{0}$ and $\mathbf{r}$ on an axis of an isotropic antiferromagnetic triangular lattice is investigated with the aid of the theory of Toeplitz determinants as developed by Wu. The leading terms in the asymptotic expansion are obtained for large spin separation at fixed nonzero temperature. Evidence is presented that the zero-point behavior of the correlation is of the form $\omega_2(\mathbf{r}) \sim \epsilon_0 r^{-\frac{1}{2}} \cos\frac{2}{3}\pi r$, where $r = |\mathbf{r}|$ is the spin separation and $\epsilon_0 = 2^{\frac{1}{2}}(E_0^T)^2 = 0.632226080\cdots$, E_0^T being the decay amplitude of the pair correlation at the Curie point (critical point) of an isotropic ferromagnetic triangular lattice. A special class of fourth-order correlations $\omega_4(\mathbf{r}) = \langle\sigma_0\sigma_\delta\sigma_\mathbf{r}\sigma_{\mathbf{r}+\delta}\rangle - \langle\sigma_0\sigma_\delta\rangle\langle\sigma_\mathbf{r}\sigma_{\mathbf{r}+\delta}\rangle$ between the four spins at sites $\mathbf{0}$, $\boldsymbol{\delta}$, $\mathbf{r}$, and $\mathbf{r}+\boldsymbol{\delta}$ on the same lattice axis, where $\boldsymbol{\delta}$ is a lattice vector, is reconsidered. The asymptotic form of the correlation for large separation of pairs of spins $r = |\mathbf{r}|$ is obtained for all fixed temperatures.

INTRODUCTION

Recently, Wu[1] has given a complete discussion of the asymptotic behavior of the Ising model pair correlation[2,3]

$$\omega_2(\mathbf{r}) = \langle\sigma_0\sigma_\mathbf{r}\rangle$$

between two spins at lattice sites $\mathbf{0}$ and $\mathbf{r}$ on the same row (or column) of a ferromagnetic rectangular lattice.[4] Wu obtained the asymptotic expansion of $\omega_2(\mathbf{r})$ for large spin separation $r = |\mathbf{r}|$ and fixed temperature T, in the three cases $0 < T < T_C$, $T = T_C$, the critical temperature (Curie point), and $T > T_C$. Wu's analysis of the asymptotic form of the Toeplitz determinant representing the pair correlation $\omega_2(\mathbf{r})$ is sufficiently general to permit immediate extension to many other cases of interest in the Ising model planar lattice,[5] with the exception of the antiferromagnetic triangular lattice.[6,7] In this paper, the application of Wu's method is made to the special case of the *isotropic* lattice, when the function which generates the elements of the Toeplitz determinant has a different analytic form from that considered by Wu. The isotropic antiferromagnetic triangular lattice has no critical point, in the sense that the thermodynamic functions have smooth temperature dependence, and there is no long-range order. The pair correlation decays to zero with increasing spin separation. The asymptotic expansion is obtained in Sec. 1 for $T > 0$, the main result being Eq. (1.31) with reference to Eqs. (1.1), (1.5), and (1.28) for definitions of the various symbols employed. The mathematically "critical" case $T = 0$ is considered in Sec. 2. Evidence is presented that the zero-point behavior of the pair correlation along a lattice axis is of the form

$$\omega_2(\mathbf{r}) \sim \epsilon_0 r^{-\frac{1}{2}} \cos\tfrac{2}{3}\pi r,$$

where $r = |\mathbf{r}|$ is the spin separation, and

$$\epsilon_0 = 2^{\frac{1}{2}}(E_0^T)^2 = 0.632226080\cdots,$$

where E_0^T is the decay amplitude of the pair correlation at the Curie point (critical point) of an isotropic ferromagnetic triangular lattice [see Eqs. (2.5)–(2.9)].

In Sec. 3, fourth-order correlations

$$\omega_4(\mathbf{r}) = \langle\sigma_0\sigma_\delta\sigma_\mathbf{r}\sigma_{\mathbf{r}+\delta}\rangle - \langle\sigma_0\sigma_\delta\rangle\langle\sigma_\mathbf{r}\sigma_{\mathbf{r}+\delta}\rangle$$

between four spins at sites $\mathbf{0}$, $\boldsymbol{\delta}$, $\mathbf{r}$, and $\mathbf{r}+\boldsymbol{\delta}$ on the same lattice axis, where $\boldsymbol{\delta}$ is a lattice vector, are reconsidered. The main result is Eq. (3.18) for the asymptotic form of $\omega_4(\mathbf{r})$ for large spin-pair separation $r = |\mathbf{r}|$ and fixed nonzero temperature. The exact form of $\omega_4(\mathbf{r})$ at zero temperature is quoted in Eq. (3.24).

1. PAIR CORRELATION FOR $T > 0$

Consider a planar isotropic triangular lattice with interaction energy $-J$ between neighboring pairs of spins, where J is negative. Antiparallel spin states are energetically preferred. It is supposed that the usual "thermodynamic limit" has been taken so that the lattice is "infinite" in two independent directions.

* Research supported in part by U.S. Air Force through grant No. AF-AFOSR-1310-67.

[1] T. T. Wu, Phys. Rev. **149**, 380 (1966).

[2] M. E. Fisher, Rept. Progr. Phys. **30**, 615 (1967).

[3] C. Domb, Advan. Phys. **9**, 149 (1960).

[4] Along a row $\mathbf{r} = (k, 0)$ and along a column $\mathbf{r} = (0, k)$ where $k = |\mathbf{r}|$ is an integer, and the lattice vectors are used as the basis.

[5] The form of the Toeplitz determinant used by Wu for pair correlations along a row of the quadratic lattice is that derived by E. W. Montroll, R. B. Potts, and J. C. Ward, J. Math. Phys. **4**, 308 (1963).

[6] A form of the Toeplitz determinant directly applicable to the triangular lattice is derived by J. Stephenson, J. Math. Phys. **5**, 1009 (1964). Alternatively, see Ref. 7.

[7] H. S. Green and C. A. Hurst, *Order Disorder Phenomena* (Interscience Publishers, Inc., New York, 1964).

The conventional notation

$$K = J/k_B T \quad \text{and} \quad v = \tanh K, \tag{1.1}$$

where k_B is Boltzmann's constant, will be used. Since J is negative, $-1 \leq v \leq 0$, $v = 0$ corresponds to $T = \infty$, and $v = -1$ corresponds to $T = 0$. The pair correlation $\omega_2(\mathbf{r})$ between two spins on the same lattice axis may be expressed as a Toeplitz determinant.[6,7] Since the lattice is isotropic, only one lattice axis need be considered. For spins on the same row, $\mathbf{r} = (k, 0)$, where k equals $r = |\mathbf{r}|$ and is the distance between the spins measured in units of the lattice spacing. Then $\omega_2(\mathbf{r})$ is equal to a Toeplitz determinant of order k with elements $a_{p,q} = a_{q-p}, p, q = 1, \cdots, k$, so

$$\omega_2(\mathbf{r}) = \det (a_{p,q})_{k\times k}, \tag{1.2}$$

where a_{q-p} is given by

$$a_{q-p} = (2\pi)^{-1}\int_{-\pi}^{\pi} d\omega e^{-i(q-p)\omega} \times \left(\frac{2v(1+v^2) - v^2(1-v)^2 e^{+i\omega} - (1-v)^2 e^{-i\omega}}{2v(1+v^2) - v^2(1-v)^2 e^{-i\omega} - (1-v)^2 e^{+i\omega}}\right)^{\frac{1}{2}}. \tag{1.3}$$

Equivalently, a_{q-p} is the coefficient of ξ^{q-p} in the expansion of the generating function

$$F(\xi) = -\xi^{-1}\left(\frac{1 - 2v\cos\theta\,.\,\xi + v^2\xi^2}{1 - 2v\cos\theta\,.\,\xi^{-1} + v^2\xi^{-2}}\right)^{\frac{1}{2}}, \tag{1.4}$$

where θ is a real angle defined by

$$\cos\theta = \tfrac{1}{2}(1 + e^{4K}), \quad 0 < \theta \leq \tfrac{1}{3}\pi. \tag{1.5}$$

$\theta = 0$ corresponds to $T = \infty$ and $\theta = \frac{1}{3}\pi$ corresponds to $T = 0$. The required expansion of the generating function $F(\xi)$ is made within the annulus $|v| < |\xi| < |v|^{-1}$. If, on the unit circle C_0, we set $\xi = e^{i\omega}$, we regain the integral form (1.3) for a_{q-p}:

$$a_{q-p} = (2\pi)^{-1}\int_0^{2\pi} d\omega e^{-i(q-p)\omega} F(e^{i\omega}). \tag{1.6}$$

(For the derivation of these formulas, see Ref. 6 or the monograph by Hurst and Green.[7]) By analogy with Sec. 2 of Wu's paper, we set

$$C(\xi) = -\xi F(\xi) = \left[\left(\frac{1 - ve^{i\theta}\xi}{1 - ve^{i\theta}\xi^{-1}}\right)\left(\frac{1 - ve^{-i\theta}\xi}{1 - ve^{-i\theta}\xi^{-1}}\right)\right]^{\frac{1}{2}}, \tag{1.7}$$

$$P(\xi) = [(1 - ve^{+i\theta}\xi)(1 - ve^{-i\theta}\xi)]^{-\frac{1}{2}},$$
$$Q(\xi) = [(1 - ve^{+i\theta}\xi)(1 - ve^{-i\theta}\xi)]^{\frac{1}{2}}, \tag{1.8}$$

so

$$C(\xi) = [P(\xi)Q(\xi^{-1})]^{-1},$$

where $P(\xi)$ and $Q(\xi)$ are both analytic for $|\xi| < 1 < |v|^{-1}$, and continuous and nonzero for $|\xi| \leq 1$. The correspondence between these formulas and those of Wu is immediately evident. Then, taking the result (2.6) from Wu's paper, we have

$$\omega_2 = D_k \sim (-1)^{k+1} R_{k+1}(-x_k), \tag{1.9}$$

where $(-1)^{k+1}R_{k+1}$ is exponentially (in k) close to its limit

$$R = \lim_{k\to\infty}(-1)^{k+1}R_{k+1} = (1 - v^2)^{\frac{1}{2}}(1 - 2v^2\cos 2\theta + v^4)^{\frac{1}{4}}. \tag{1.10}$$

R is obtained by using Szegö's theorem[1,5,6] to calculate the limiting value of the Toeplitz determinant generated by $C(\xi)$. And $(-x_k)$ is asymptotically given by

$$(-x_k) \sim (2\pi i)^{-1}\int_{C_0} d\xi \xi^{k-1} P(\xi^{-1})[Q(\xi)]^{-1} = (2\pi i)^{-1}\int_{C_0} d\xi \xi^{k-1}[(1 - ve^{i\theta}\xi^{-1})(1 - ve^{-i\theta}\xi^{-1}) \times (1 - ve^{i\theta}\xi)(1 - ve^{-i\theta}\xi)]^{-\frac{1}{2}}, \tag{1.11}$$

where the integration contour is the unit circle C_0. Alternatively, with $\xi = e^{i\omega}$, we have

$$(-x_k) \sim (2\pi i)^{-1}\int_0^{2\pi} d\omega \times e^{ik\omega}\,|(1 - ve^{i\theta}e^{i\omega})(1 - ve^{-i\theta}e^{i\omega})|^{-1}. \tag{1.12}$$

The integrand in (1.11) has branch-point singularities at $ve^{i\theta}$ and $ve^{-i\theta}$ inside the unit circle, and at $v^{-1}e^{i\theta}$ and $v^{-1}e^{-i\theta}$ outside the unit circle. v is negative, so we select $\arg(v) = \pi$ and write $v = |v|\,e^{i\pi}$. We take the ξ plane to have a cut inside the unit circle from a point A: $\xi = |v|\,e^{i(\pi-\theta)}$, along the radius AO to the origin O: $\xi = 0$, and thence along the radius OB to a point B: $\xi = |v|\,e^{i(\pi+\theta)}$. Here, $0 < \theta < \frac{1}{3}\pi$. Cuts which must be made outside the unit circle do not concern us here. Now deform the contour of integration by contracting it onto the cut. Along OA set $\xi = \xi_1\,|v|\,e^{i(\pi-\theta)}$, and along OB set $\xi = \xi_1\,|v|\,e^{i(\pi+\theta)}$, where ξ_1 is a new integration variable, $0 \leq \xi_1 \leq 1$. The result is

$$(-x_k) \sim \pi^{-1}v^k\int_0^1 d\xi_1\xi_1^{k-1} \times \{-ie^{-ik\theta}[(1 - v^2\xi_1)(1 - v^2e^{-i2\theta}\xi_1) \times (1 - e^{i2\theta}\xi_1^{-1})(1 - \xi_1^{-1})]^{-\frac{1}{2}} + \text{c.c.}\}. \tag{1.13}$$

After slight rearrangement,

$$(-x_k) \sim \text{Re}\; i2\pi^{-1}v^k e^{-ik\theta}e^{-i\theta}\int_0^1 d\xi_1\xi_1^k \times [(1 - \xi_1)(1 - e^{-i2\theta}\xi_1)(1 - v^2\xi_1) \times (1 - v^2e^{-i2\theta}\xi_1)]^{-\frac{1}{2}}. \tag{1.14}$$

Now we require the asymptotic expansion of an integral J of the form

$$J = \int_0^1 d\xi_1 \xi_1^k (1-\xi_1)^{-\frac{1}{2}} \prod_{i=1}^{3} (1 - b_i\xi_1)^{-\frac{1}{2}}, \quad (1.15)$$

where the b_i's are unequal, $b_i \neq 1$, and $|b_i| \leq 1$. In our case,

$$b_1 = e^{-i2\theta}, \quad b_2 = v^2, \quad b_3 = v^2 e^{-i2\theta}. \quad (1.16)$$

Rearrange the factors in the products in terms of the combination

$$x = \left(\frac{1-\xi_1}{1+a\xi_1}\right), \quad (1.17)$$

where a, if real, is greater than -1, and otherwise is at our disposal, so

$$(1-b_i\xi_1)^{-\frac{1}{2}} = (1+a\xi_1)^{-\frac{1}{2}} \left(\frac{1+a}{1-b_i}\right)^{\frac{1}{2}} \times \left[1 + \left(\frac{a+b_i}{1-b_i}\right)x\right]^{-\frac{1}{2}}.$$

Then

$$J = (1+a)^{\frac{3}{2}} \prod_{i=1}^{3} (1-b_i)^{-\frac{1}{2}} \times \int_0^1 d\xi_1 \xi_1^k (1-\xi_1)^{-\frac{1}{2}} (1+a\xi_1)^{-\frac{3}{2}} C\left(\frac{1-\xi_1}{1+a\xi_1}\right), \quad (1.18)$$

where

$$C(x) = \prod_{i=1}^{3} \left[1 + \left(\frac{a+b_i}{1-b_i}\right)x\right]^{-\frac{1}{2}} \quad (1.19)$$

and has the formal expansion

$$C(x) = \sum_{r=0}^{\infty} C_r x^r. \quad (1.20)$$

Inserting this expansion in the integrand (1.18) and formally interchanging the order of summation and integration, we get

$$J = (1+a)^{\frac{3}{2}} \prod_{i=1}^{3} (1-b_i)^{-\frac{1}{2}} \times \sum_{r=0}^{\infty} C_r \int_0^1 d\xi_1 \xi_1^k (1-\xi_1)^{r-\frac{1}{2}} (1+a\xi_1)^{-r-\frac{3}{2}}. \quad (1.21)$$

Now use Euler's integral representation of the hypergeometric function,[8]

$$J = \prod_{i=1}^{3} (1-b_i)^{-\frac{1}{2}} \Gamma(k+1) \times \sum_{r=0}^{\infty} C_r (1+a)^{-r} \Gamma(r+\tfrac{1}{2}) \Gamma(k+r+\tfrac{3}{2})^{-1} \times F(r+\tfrac{3}{2}, r+\tfrac{1}{2}, k+r+\tfrac{3}{2}; a(1+a)^{-1}), \quad (1.22)$$

[8] *Higher Transcendental Functions*, A. Erdélyi, Ed. (McGraw-Hill Book Co., New York, 1953), Vol. I, Chap. 2.

and formally expand the hypergeometric function:

$$J = \prod_{i=1}^{3} (1-b_i)^{-\frac{1}{2}} \Gamma(k+1) \times \sum_{r=0}^{\infty} C_r (1+a)^{-r} \Gamma(r+\tfrac{3}{2})^{-1} \times \sum_{s=0}^{\infty} [a(1+a)^{-1}]^s \Gamma(r+s+\tfrac{3}{2}) \Gamma(r+s+\tfrac{1}{2}) \times [\Gamma(k+r+s+\tfrac{3}{2}) \Gamma(s+1)]^{-1}, \quad (1.23)$$

which, after rearrangement with $r+s=m$, $r=n$, gives

$$J = \prod_{i=1}^{3} (1-b_i)^{-\frac{1}{2}} \Gamma(k+1) \times \sum_{m=0}^{\infty} [a(1+a)^{-1}]^m \Gamma(m+\tfrac{3}{2}) \Gamma(k+m+\tfrac{3}{2})^{-1} \times \sum_{n=0}^{m} C_n a^{-n} [\Gamma(n+\tfrac{3}{2}) \Gamma(m-n+1)]^{-1}, \quad (1.24)$$

whence an asymptotic expansion for J in fractional powers of k^{-1} may be obtained. The result is independent of a, though (1.24) conceals this, and remains valid even if $-a$ equals one of the b_i's.

Several complications may arise. The series representation of the hypergeometric function in (1.22) may be divergent, in which case the series must be interpreted as an asymptotic expansion of the integral. This difficulty can always be avoided by choosing $a = 0$, in which case everything simplifies to

$$J = \prod_{i=1}^{3} (1-b_i)^{-\frac{1}{2}} \Gamma(k+1) \times \sum_{r=0}^{\infty} C_r \Gamma(r+\tfrac{1}{2}) \Gamma(k+r+\tfrac{3}{2})^{-1}, \quad (1.25)$$

which is the form we shall use. Then, in our case, the series expansion for $C(x)$ diverges. The final expansion of J in fractional powers of k^{-1} is an asymptotic one.

We now use these results in (1.14). It is convenient to abbreviate

$$(1-b_3)^{-\frac{1}{2}} = (1 - v^2 e^{-i2\theta})^{-\frac{1}{2}} \text{ by } \rho e^{i\phi}, \quad (1.26)$$

where

$$\rho = (1 - 2v^2 \cos 2\theta + v^4)^{-\frac{1}{4}} \quad (1.27)$$

and

$$\phi = -\tfrac{1}{2} \arg (1 - v^2 \cos 2\theta + iv^2 \sin 2\theta). \quad (1.28)$$

Employing the asymptotic expansion (1.25) to the case of the integral in the expression (1.14) for $(-x_k)$ and substituting the result in (1.9), we get the asymptotic

expansion of the pair correlation:

$$\omega_2 \sim R(-x_k) \tag{1.29}$$
$$\sim (\tfrac{1}{2}\pi \sin\theta)^{-\frac{1}{2}} v^k \operatorname{Re} \exp\left[-i(k\theta + \tfrac{1}{2}\theta - \tfrac{1}{4}\pi - \phi)\right] \times \sum_{r=0}^{\infty} C_r \frac{\Gamma(k+1)\Gamma(r+\frac{1}{2})}{\Gamma(\frac{1}{2})\Gamma(k+r+\frac{3}{2})}. \tag{1.30}$$

Writing the first two terms in detail, we have

$$\omega_2 \sim (\tfrac{1}{2}\pi \sin\theta)^{-\frac{1}{2}} v^k \cdot k^{-\frac{1}{2}} \{\cos(k\theta + \tfrac{1}{2}\theta - \tfrac{1}{4}\pi - \phi) - (4k)^{-1}[\tfrac{3}{2} + (2\sin\theta)^{-1}\cos(k\theta + \tfrac{3}{2}\theta + \tfrac{1}{4}\pi - \phi) + v^2(1-v^2)^{-1}\cos(k\theta + \tfrac{1}{2}\theta - \tfrac{1}{4}\pi - \phi) + v^2\rho^2 \cos(k\theta + \tfrac{5}{2}\theta - \tfrac{1}{4}\pi - 3\phi)]\}. \tag{1.31}$$

The essential features of the behavior of the correlation are as follows: (i) an exponential decay $|v|^k$ with (ii) alternating sign $(-1)^k$ modified by (iii) an additional $k^{-\frac{1}{2}}$ decay and (iv) a slowly varying cosine envelope with period in k of $2\pi\theta^{-1} \simeq \pi |v|^{-\frac{1}{2}}$ at high temperatures. The decay amplitude is $(\frac{1}{2}\pi \sin\theta)^{-\frac{1}{2}}$.

2. PAIR CORRELATION AT $T = 0$

The considerations of the preceding section no longer apply at $T = 0$. The annulus in which the Laurent expansions of the generating functions $F(\xi)$ and $C(\xi)$ are required has now contracted to the unit circle. The branch-point singularities have coalesced in pairs on the unit circle at $\xi = e^{\pm i2\pi/3}$, and the procedure leading to the integral (1.11) for $-x_k$ is invalid. If we pursue the previous calculation further, it is apparent that the integral in (1.14) for $-x_k$ is divergent, and the limiting value R of $(-1)^{k+1}R_{k+1}$ is zero. However, $\omega_2 \sim R(-x_k)$, and setting $v = -1$, $\theta = \frac{1}{3}\pi$, $\phi = -\frac{1}{12}\pi$ in the leading term of the final asymptotic expansion (1.31) for ω_2 may yet provide some correct qualitative information about the behavior of ω_2 very close to $T = 0$, and perhaps even at $T = 0$. (Higher terms in the asymptotic expansion are divergent at $T = 0$.) "Near" $T = 0$,

$$\omega_2 \sim (2\pi^{-\frac{1}{2}}3^{-\frac{1}{4}}) \cdot k^{-\frac{1}{2}} \cos\tfrac{2}{3}\pi k, \tag{2.1}$$

which suggests that at $T = 0$ the essential features of the behavior of the correlation are (i) a $k^{-\frac{1}{2}}$ decay, with (ii) oscillating amplitude. Note that $\cos\frac{2}{3}\pi k$ is $+1$ when k is a multiple of 3, and $-\frac{1}{2}$ for other integer values of k.

From a numerical analysis of the correlation at $T = 0$, of which details may be found in Ref. 6, the author has previously suggested that the decay of ω_2 has the form

$$\omega_2 \sim \epsilon_0 k^{-\frac{1}{2}} \cos\tfrac{2}{3}\pi k, \tag{2.2}$$

where the numerical estimate of the constant ϵ_0 is

$$\epsilon_0 = 0.632226 \pm 2, \tag{2.3}$$

whereas $(2\pi^{-\frac{1}{2}}3^{-\frac{1}{4}}) = 0.8753\cdots$ in (2.1).

Now let us try to relate the decay constant ϵ_0 to the decay constant E_0^T of pair correlations at the ferromagnetic critical point.[1,9–11] E_0^T is the only other available critical-point constant for the triangular lattice! The pair correlation ω_2 on the ferromagnetic lattice is equal to the Toeplitz determinant whose elements are generated above the critical point by the function[5–7]

$$D(\xi) = -\xi^{-1}\left[\left(\frac{1 - A^{-1}\xi}{1 - A^{-1}\xi^{-1}}\right)\left(\frac{1 - B\xi}{1 - B\xi^{-1}}\right)\right]^{\frac{1}{2}}, \tag{2.4}$$

where $0 \leq B < 1$ and $A > 1$. [See Refs. (7) and (11) for the connection between the value of B and the lattice axis under consideration.] The ferromagnetic critical point is given by $A = 1$. The generating function $D(\xi)$ has branch-point singularities at $\xi = A^{\pm 1}$. There is a zero of order $\frac{1}{2}$ at $\xi = A$ outside the unit circle, and a pole of order $\frac{1}{2}$ at $\xi = A^{-1}$ inside the unit circle. As $A \to 1+$, the zero and pole tend to coincidence on the unit circle at $\xi = 1$. The decay of the correlation at the critical point is known to be[11]

$$\omega_2 \sim \frac{E}{k^{\frac{1}{4}}}\left(\frac{1+B}{1-B}\right)^{\frac{1}{4}}, \tag{2.5}$$

where

$$E = 0.645002448\cdots. \tag{2.6}$$

For the isotropic ferromagnetic triangular lattice,

$$E_0^T = 3^{-\frac{1}{8}}2^{\frac{1}{4}}, \; E = 0.668618986\cdots. \tag{2.7}$$

Now compare the forms of the generating functions in the ferromagnetic case above as $A \to 1+$, and in the antiferromagnetic case as $v \to -1$. The critical feature in the ferromagnetic case is that a zero of order $\frac{1}{2}$ and a pole of order $\frac{1}{2}$ coalesce on the unit circle C_0 at $\xi = 1$ as $T \to T_C$ from above, with the zero approaching $\xi = 1$ from inside C_0 and the pole approaching $\xi = 1$ from outside C_0. The antiferromagnetic generating function has similar features as $T \to 0$, $v \to -1$, with coincidences of a pole and a zero occurring at $\xi = e^{i2\pi/3}$ *and* at $\xi = e^{i4\pi/3}$. In view of this, one might suppose that the decay constant ϵ_0 on the antiferromagnetic lattice at $T = 0$ would be related to the square of the decay constant E_0^T at

[9] B. Kaufman and L. Onsager, Phys. Rev. **76**, 1244 (1949).
[10] M. E. Fisher, Physica **25**, 521 (1959).
[11] R. E. Hartwig and J. Stephenson, J. Math. Phys. **9**, 836 (1968).

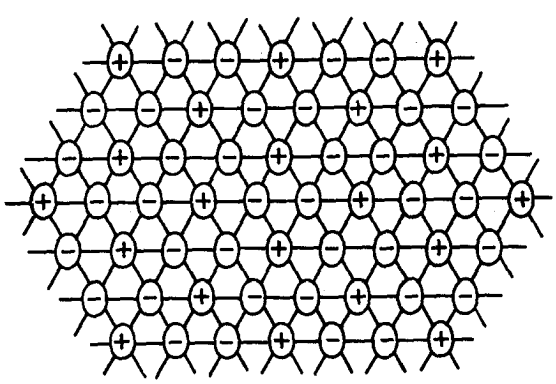


FIG. 1. A ground-state arrangement of spins on an isotropic antiferromagnetic triangular lattice which corresponds to the expectation values $\langle\sigma_0\sigma_r\rangle = \omega_3(\mathbf{r})$ at zero temperature. The spin at site 0 is an "up" spin (+). "Down" spins are denoted by (−).

the ferromagnetic critical point. Numerical inspection shows that the combination

$$2^{\frac{1}{2}}(E_0^T)^2 = 3^{-\frac{1}{4}}2E^2 = 0.632226080\cdots \quad (2.8)$$

is very close to the estimate for ϵ_0 in (2.3). It is natural to guess that

$$\epsilon_0 = 2^{\frac{1}{2}}(E_0^T)^2. \quad (2.9)$$

The zero-point behavior of the pair correlation may be related to the three-sublattice structure of the triangular lattice. The correlations fall in groups of three, corresponding to the alternation in sign, being positive on the sublattice for which k is a multiple of 3 and negative on the other two sublattices. There is a direct correspondence between the "average state" apparent from the expectation values $\langle\sigma_0\sigma_r\rangle = \omega_2$ and the ground states which contribute most of the zero-point entropy,[12] as discussed by Wannier (see Ref. 12, p. 361 and Fig. 8; Ref. 3, pp. 212–214. Also see Fig. 1 of this paper).

3. FOURTH-ORDER CORRELATIONS

In this section, reconsideration is given to fourth-order correlations between a nearest-neighbor pair of spins at (0, 0) and (1, 0) and a nearest-neighbor pair of spins at $(k, 0)$ and $(k + 1, 0)$.[13] All four spins lie on the same lattice axis. The required correlation function ω_4 is given by

$$\omega_4 = \langle\sigma_{0,0}\sigma_{1,0}\sigma_{k,0}\sigma_{k+1,0}\rangle - \langle\sigma_{0,0}\sigma_{1,0}\rangle\langle\sigma_{k,0}\sigma_{k+1,0}\rangle = -a_k a_{-k}, \quad (3.1)$$

where $k > 0$ is a measure of the separation of the pairs of spins. Here a_k, for all k, is given by

$$a_k = (2\pi)^{-1}\int_0^{2\pi}(-1)e^{-i(k+1)w}\left[\left(\frac{1 - ve^{i\theta}e^{i\omega}}{1 - ve^{i\theta}e^{-i\omega}}\right)\times\left(\frac{1 - ve^{-i\theta}e^{i\omega}}{1 - ve^{-i\theta}e^{-i\omega}}\right)\right]^{\frac{1}{2}}, \quad (3.2)$$

so the a_k are the same as the elements of the Toeplitz determinant representing the pair correlation ω_2. Employing the generating function $F(\xi)$ in (1.3), we have

$$a_k = (2\pi i)^{-1}\int_{C_0} d\xi\xi^{-(k+1)}F(\xi), \quad (3.3)$$

where C_0 is the unit circle. There is no pole at the origin if $k \leq -1$. This last expression is therefore most suited for calculation of a_{-k}, $k \geq 1$:

$$a_{-k} = -(2\pi i)^{-1}\int_{C_0} d\xi\xi^{k-2}\left[\left(\frac{1 - ve^{i\theta}\xi}{1 - ve^{i\theta}\xi^{-1}}\right)\times\left(\frac{1 - ve^{-i\theta}\xi}{1 - ve^{-i\theta}\xi^{-1}}\right)\right]^{\frac{1}{2}}. \quad (3.4)$$

The calculation of a_k, $k \geq 1$, will be similar to that for a_{-k}. In (3.2), set $\omega' = 2\pi - \omega$ and then $\xi = e^{i\omega'}$ on the unit circle in the ξ plane; for $k \geq 1$ we obtain

$$a_k = -(2\pi i)^{-1}\int_{C_0} d\xi\xi^{k}\left[\left(\frac{1 - ve^{i\theta}\xi^{-1}}{1 - ve^{i\theta}\xi}\right)\times\left(\frac{1 - ve^{-i\theta}\xi^{-1}}{1 - ve^{-i\theta}\xi}\right)\right]^{\frac{1}{2}}. \quad (3.5)$$

The asymptotic form of a_k and a_{-k}, and thence ω_4, follows from (3.4) and (3.5) by an analysis similar to that used in the discussion of $(-x_k)$ in Sec. 1. The ξ plane is cut inside the unit circle C_0 from A: $\xi = ve^{-i\theta}$ to O and from O to B: $\xi = ve^{i\theta}$, and the contour of integration contracted onto the cut AOB. Along OA, a new variable of integration ξ_1, $0 \leq \xi_1 \leq 1$, is introduced through $\xi = \xi_1 ve^{-i\theta}$ and along OB set $\xi = \xi_1 ve^{i\theta}$. Let us summarize the remainder of the calculation.

For a_{-k},

$$a_{-k} = (-1)2\pi^{-1}v^{k-1}\,\mathrm{Re}\,ie^{-ik\theta}\times\int_0^1 d\xi_1\xi_1^{k-1}\left[\frac{(1 - v^2\xi_1)(1 - v^2e^{-2i\theta}\xi_1)}{(1 - e^{-2i\theta}\xi_1)(1 - \xi_1)}\right]^{\frac{1}{2}}. \quad (3.6)$$

Introduce b_1, b_2, b_3, ρ, and ϕ as in Sec. 1, Eqs. (1.16) and (1.26)–(1.28). Then

$$a_{-k} = 2\pi^{-1}v^{k-1}\,\mathrm{Re}\,\{\exp[-i(k\theta + \tfrac{1}{2}\pi)]\}J', \quad (3.7)$$

[12] G. H. Wannier, Phys. Rev. 79, 357 (1950).
[13] J. Stephenson, J. Math. Phys. 7, 1123 (1966).

where

$$J' = \int_0^1 d\xi_1 \cdot \xi_1^{k-1} \left(\frac{1+a\xi_1}{1-\xi_1}\right)^{\frac{1}{2}} \left[\frac{(1-b_2)(1-b_3)}{(1-b_1)(1+a)}\right]^{\frac{1}{2}} \times D\left(\frac{1-\xi_1}{1+a\xi_1}\right), \quad (3.8)$$

in which the integrand has been rearranged in terms of the combination

$$x = (1-\xi_1)/(1+a\xi_1) \quad (3.9)$$

(if a is real, $a > -1$), and

$$D(x) = \left[1 + \left(\frac{a+b_2}{1-b_2}\right)x\right]^{\frac{1}{2}} \left[1 + \left(\frac{a+b_3}{1-b_3}\right)x\right]^{\frac{1}{2}} \times \left[1 + \left(\frac{a+b_1}{1-b_1}\right)x\right]^{-\frac{1}{2}}. \quad (3.10)$$

Introducing the formal series expansion of $D(x)$,

$$D(x) = \sum_{r=0}^{\infty} D_r x^r, \quad (3.11)$$

and using the Euler integral representation of the hypergeometric function, we get

$$J' \sim [(1-b_2)(1-b_3)]^{\frac{1}{2}}(1-b_1)^{-\frac{1}{2}}\Gamma(k) \times \sum_{r=0}^{\infty} D_r \Gamma(r+\tfrac{1}{2})\Gamma(k+r+\tfrac{1}{2})^{-1}(1+a)^{-r} \times F(r-\tfrac{1}{2}, r+\tfrac{1}{2}, k+r+\tfrac{1}{2}; a(1+a)^{-1}). \quad (3.12)$$

For simplicity choose $a = 0$, insert the values of the b_i, and observe that, from Eqs. (1.10) and (1.27),

$$(1-v^2)^{\frac{1}{2}}\rho^{-1} = R, \quad (3.13)$$

thus obtaining

$$a_{-k} \sim 2\pi^{-\frac{1}{2}}(2\sin\theta)^{-\frac{1}{2}} \times R \cdot v^{k-1} \operatorname{Re}\{\exp[-i(k\theta - \tfrac{1}{2}\theta + \tfrac{3}{4}\pi + \phi)]\} \times \sum_{r=0}^{\infty} D_r \Gamma(r+\tfrac{1}{2})\Gamma(k)[\Gamma(\tfrac{1}{2})\Gamma(k+r+\tfrac{1}{2})]^{-1}. \quad (3.14)$$

The leading term ($r = 0$) is

$$a_{-k} \sim (\tfrac{1}{2}\pi\sin\theta)^{-\frac{1}{2}} R v^k k^{-\frac{1}{2}} \cos(k\theta - \tfrac{1}{2}\theta + \tfrac{3}{4}\pi + \phi). \quad (3.15)$$

A similar calculation for a_k yields

$$a_k \sim \pi^{-\frac{1}{2}}(2\sin\theta)^{\frac{1}{2}} \times R^{-1}v^{k+1} \operatorname{Re}\{\exp[-i(k\theta + \tfrac{1}{2}\theta + \tfrac{1}{4}\pi - \phi)]\} \times \sum_{r=0}^{\infty} D'_r \Gamma(r+\tfrac{3}{2})\Gamma(k)[\Gamma(\tfrac{3}{2})\Gamma(k+r+\tfrac{3}{2})]^{-1}, \quad (3.16)$$

where D'_r are coefficients of x^r in the formal expansion of $[D(x)]^{-1}$. The leading term ($r = 0$) is

$$a_k \sim \pi^{-\frac{1}{2}}(\sin\theta)^{\frac{1}{2}} R^{-1} v^{k+1} k^{-\frac{3}{2}} \times \cos(k\theta + \tfrac{1}{2}\theta + \tfrac{1}{4}\pi - \phi). \quad (3.17)$$

Combining these results, for the leading term in the asymptotic expansion of the fourth-order correlation ω_4, we have

$$\begin{aligned}\omega_4 &\sim (-1)2\pi^{-1}v^{2k}k^{-2}\cos(k\theta - \tfrac{1}{2}\theta + \tfrac{3}{4}\pi + \phi) \times \cos(k\theta + \tfrac{1}{2}\theta + \tfrac{1}{4}\pi - \phi) \\ &= (-1)2\pi^{-1}v^{2k}k^{-2} \times \{[\sin(k\theta)]^2 - [\sin(\tfrac{1}{4}\pi - \tfrac{1}{2}\theta + \phi)]^2\}. \end{aligned} \quad (3.18)$$

We note the "cancellation" of the factors R and $(\sin\theta)^{\frac{1}{2}}$.

At high temperatures, the main features of the correlation are (i) an exponential v^{2k} decay, (ii) modified by an additional factor k^{-2}, together with (iii) a curious oscillation in sign, determined by the factor in curly brackets in (3.18). The correlation alternates regularly in sign as the separation k of the pairs of spins increases. The positive regions[14] of the lattice are of approximate length (in lattice spacings) Δk_+, given by

$$\Delta k_+ = (\tfrac{1}{2}\pi - \theta + 2\phi)/\theta, \quad (3.19)$$

and the negative regions are of approximate length Δk_-, given by

$$\Delta k_- = (\tfrac{1}{2}\pi + \theta - 2\phi)/\theta. \quad (3.20)$$

At high temperatures the correlation is very approximately

$$\omega_4 \sim \pi^{-1}v^{2k}k^{-2}\cos 2k\theta, \quad (3.21)$$

and the alternating positive and negative regions are of approximately equal length:

$$\Delta k_+ = \Delta k_- = \pi/2\theta. \quad (3.22)$$

As the temperature decreases, the regions of common sign shrink in length with $\Delta k_+ < \Delta k_-$, until at $T = 0$ the correlation is always negative or zero.

The above asymptotic expansions were obtained for $T > 0$, and we cannot expect to get valid results by setting $T = 0$. Nevertheless, the expression (3.18) for ω_4 has the limiting form

$$\lim_{T\to 0+} \omega_4 \sim -2\pi^{-1} \cdot k^{-2}(\sin\tfrac{1}{3}\pi k)^2, \quad (3.23)$$

and has the correct qualitative behavior, but the wrong decay amplitude, as we can see by comparison

[14] I.e., sets of adjacent spins along a lattice axis for which ω_4 is positive.

with the exact (correct) result for ω_4 at $T = 0$, which is[15]

$$\begin{aligned}\omega_4 &= -[2(\pi k)^{-1}\sin\tfrac{2}{3}\pi k]^2, && k \neq 0,\\ &= 0, && k \text{ a multiple of } 3,\ k \neq 0,\\ &= -3/(\pi k)^2, && \text{other integer values of } k. \end{aligned} \quad (3.24)$$

Numerically, for integer k, $k \neq 0$,

$$\omega_4 = -[2(\pi k)^{-1}\sin\tfrac{1}{3}\pi k]^2. \quad (3.25)$$

Thus the correct decay amplitude $(2/\pi)^2$ differs from that obtained from $\lim_{T\to 0}\omega_4$ by a factor $(2/\pi)$. A precisely similar difference occurs for the ferromagnetic fourth-order correlations[13] for which

$$\begin{aligned}\omega_4 &\sim A^{2k}/(2\pi k^2), && A < 1,\ T < T_C,\\ &\sim 1/(\pi k)^2, && A = 1,\ T = T_C,\\ &\sim A^{-2k}/(2\pi k^2), && A > 1,\ T > T_C. \end{aligned} \quad (3.26)$$

The behavior of the antiferromagnetic correlation ω_4 at $T = 0$ may be related to the three-sublattice structure of the triangular lattice, as in the discussion of the pair correlation in Sec. 2.

[15] The exact formula for ω_4 at $T = 0$ may be obtained by setting $v = -1$ in Eq. (1.3), calculating a_k explicitly, and substituting the result in Eq. (3.1). Alternatively, see Ref. 13, Eq. (7.4).

4. CONCLUDING REMARKS

In this paper, formulas for the asymptotic decay of pair correlations $\omega_2(\mathbf{r})$ have been obtained for fixed temperature when the two spins lie on the same axis of an isotropic triangular lattice. The oscillatory behavior of the correlations is peculiar to the triangular lattice, which exhibits one disordered phase over all temperatures. The qualitative form of the correlation is quite different from that associated with antiferromagnetic short-range order, which has oscillatory behavior confined to a simple alternation of the sign $(-1)^k$ when the spin-separation vector $\mathbf{r} = (k, 0)$. It is usually supposed that the isotropic triangular lattice merely corresponds to an anisotropic triangular lattice with Néel point T_N (antiferromagnetic critical point) at $T = 0$. Now an anisotropic triangular lattice exhibits antiferromagnetic long-range order below the Néel point T_N. But it is not immediately obvious what happens above the Néel point, since short-range antiferromagnetic order is incompatible with the oscillatory behavior of pair correlation derived above. The elucidation of this problem will be presented in the next paper of this series.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Ising-Model Spin Correlations on the Triangular Lattice. IV. Anisotropic Ferromagnetic and Antiferromagnetic Lattices*

JOHN STEPHENSON
Theoretical Physics Institute, University of Alberta, Edmonton, Alberta, Canada

(Received 20 February 1969; Revised Manuscript Received 5 June 1969)

A detailed discussion of pair correlations $\omega_2(\mathbf{r}) = \langle \sigma_0 \sigma_\mathbf{r} \rangle$ between spins at lattice sites **0** and **r** on the *axes* of anisotropic triangular lattices is given. The asymptotic behavior of $\omega_2(\mathbf{r})$ for large spin separation is obtained for ferromagnetic and antiferromagnetic lattices. The axial pair correlation for the ferromagnetic triangular lattice has the same qualitative behavior as that for the ferromagnetic rectangular lattice: There is long-range order below the Curie point T_C and short-range order above. It is shown that correlations on the anisotropic antiferromagnetic triangular lattice must be given separate treatment in *three* different temperature ranges. Below the Néel point T_N (antiferromagnetic critical point), the completely anisotropic lattice exhibits antiferromagnetic long-range order along the two lattice axes with the strongest interactions. Spins along the third axis with the weakest interaction are ordered ferromagnetically. Between T_N and a uniquely located tempetature T_D, there is antiferromagnetic short-range order along the two axes with the strongest interactions, and ferromagnetic short-range order along the other axis. T_D is named the *disorder temperature* because it divides the short-range-order region $T_N < T < T_D$ from the region $T_D < T < \infty$, in which the axial pair correlations have exponential decay with temperature-dependent oscillatory envelope. There is no singularity in the partition function at T_D, so there are only two thermodynamic phases: ordered below the Néel point, and disordered above. Correlations at T_D decay exponentially. Finally, special consideration is given to the anisotropic antiferromagnetic lattice when the two weakest interactions are equal, and $T_N = T_D = 0$. The single disordered phase exhibits exponential correlation decay with oscillatory envelope for $T > 0$. The exact values of the axial pair correlations at $T = 0$ are calculated. For large spin separation r along the strong interaction axis, $\omega_2 = (-1)^r$, and along the weak (equal) interaction axes

$$\omega_2 \sim 2^{\frac{1}{2}} E^2 \cdot r^{-\frac{1}{2}} \cos(\tfrac{1}{2}\pi r)[1 - (8r^2)^{-1} + \cdots],$$

where

$$2^{\frac{1}{2}} E^2 = 0.588352663\cdots,$$

and E is a decay constant relating to pair correlations at the Curie point of a square lattice.

INTRODUCTION

Twenty-five years ago, Onsager[1] published his solution of the Ising problem[2-4] for the thermodynamic properties of the two-dimensional anisotropic rectangular lattice in zero magnetic field. Soon afterwards, Kaufman and Onsager[5] presented an account of the short-range correlations[6] and announced the formula for the long-range order parameter which is equal to the square of the spontaneous magnetization.[7] The first published derivation of the spontaneous magnetization is that of Yang,[8] with a generalization to an anisotropic rectangular lattice by Chang.[9] More recently, Wu[10] has given an account of the way in which the short-range pair correlation along a row of the rectangular lattice behaves for large spin separation at various fixed temperatures. Kadanoff[11] has also calculated the general pair correlation in terms of a parameter linking the spin separation and the deviation of the temperature from the critical point. The results obtained for the quadratic lattice exhibit the following general features[3]:

(i) There is one singular point, or critical point, called the Curie point for a ferromagnet, and the Néel point for an antiferromagnet, below which there is long-range order, and above which there is short-range order.

(ii) The features of anisotropic and isotropic lattices are similar.

(iii) The transformation to an antiferromagnetic is trivial, and no new features in nonmagnetic properties are introduced (apart from an oscillation in sign of some of the pair correlations). In particular, the Néel point (antiferromagnetic critical point) is numerically equal to the Curie point (ferromagnetic critical point).

* Research supported in part by the U.S. Air Force through Grant No. AF-AFOSR-1310-67.

[1] L. Onsager, Phys. Rev. **65**, 117 (1944).
[2] E. Ising, Z. Physik **31**, 253 (1925).
[3] C. Domb, Advan. Phys. **9**, 149 (1960).
[4] M. E. Fisher. Rept. Progr. Phys. **30**, 615 (1967).
[5] B. Kaufman and L. Onsager, Phys. Rev. **76**, 1244 (1949).
[6] E. W. Montroll, R. B. Potts, and J. C. Ward, J. Math. Phys. **4**, 308 (1963).
[7] L. Onsager, Nuovo Cimento Suppl. **6**, 261 (1949); B. Kaufman and L. Onsager, "Long Range Order" (unpublished).
[8] C. N. Yang, Phys. Rev. **85**, 808 (1952).
[9] C. H. Chang, Phys. Rev. **88**, 1422 (1952).
[10] T. T. Wu, Phys. Rev. **149**, 380 (1966).
[11] L. P. Kadanoff, Nuovo Cimento, **44B**, 276 (1966).

We may contrast these qualitative features of a quadratic lattice with those of the triangular lattice, for which solutions were obtained in 1950 by a number of authors.[12–14] In particular, we mention the solutions of Houtappel[12] for the general *anisotropic* triangular lattice, and that of Wannier[13] for the isotropic lattice. The results obtained for the ferromagnetic triangular lattice exhibit the same general features as the quadratic lattice. However, for an *antiferromagnetic* lattice the situation is as follows:

(i) There is one singular point which is at zero temperature for the isotropic lattice.

(ii) The features of anisotropic and isotropic lattices are quite different.

(iii) The transformation to an antiferromagnet is not trivial in effect, and the Néel point is numerically lower than the Curie point of the corresponding ferromagnetic lattice.[15] The thermodynamic properties of the antiferromagnetic isotropic triangular lattice were commented on by Wannier.[13] In particular, he emphasized the unphysical appearance of a finite zero-point entropy, and obtained its exact value. Since then the antiferromagnetic triangular lattice has largely been put on one side as a "curiosity."

It is the purpose of this paper to examine carefully the properties of pair correlations $\omega_2(\mathbf{r}) = \langle \sigma_0 \sigma_r \rangle$ on the *axes* of general anisotropic triangular lattices,[16,17] with particular emphasis on the antiferromagnetic case. In Sec. 1, triangular lattices are classified into two types A or B, according to whether they may be transformed to the completely ferromagnetic lattice or to the completely antiferromagnetic lattice, respectively. (The significance of this apparently trivial classification is revealed by the end of Sec. 4.) In Sec. 2, a mathematically necessary classification of the generating functions for the elements of the Toeplitz determinant representing the pair correlation is made [see Eq. (2.9)]. Mathematical arguments in subsequent sections depend on the results of Secs. 1 and 2. In Sec. 3, correlations for the general *ferromagnetic* lattice are discussed in some detail. Much of the work is a generalization of Wu's results for the rectangular lattice,[10] and is needed later. The two temperature ranges $0 < T < T_C$ and $T > T_C$ and the special point $T = T_C$ are considered, where T_C is the Curie point. In the opening paragraphs of Sec. 4, we show that correlations on an *anisotropic antiferromagnetic* lattice must be given separate treatment in *three* different temperature ranges and at *two* special points. Below the Néel point T_N, the completely anisotropic lattice exhibits antiferromagnetic long-range order along the two lattice axes with the two strongest interaction energies. Spins along the third axis with the weakest interaction energy are ordered ferromagnetically in energetically unfavored orientations. A moment's consideration will show that this is also the way in which an antiferromagnetic triangular lattice attains its ground state. Between T_N and a uniquely located temperature T_D (Eq. 4.2), the lattice exhibits antiferromagnetic short-range order along the two axes with the strongest interaction energies, and ferromagnetic short-range order along the other axis. T_D is introduced here as the *disorder temperature* because it divides the short-range-order region $T_N < T < T_D$ just mentioned, from the region $T_D < T < \infty$, in which the pair correlation has exponential decay with a temperature-dependent oscillatory envelope (4.18). Notice that the antiferromagnetic lattice exhibits only two thermodynamic phases: an ordered phase below the Néel point T_N, and a disordered phase above. There is no singularity in the partition function at T_D. Special consideration is given to correlations at T_N and T_D. In fact, the correlations can be evaluated exactly at T_D [Eqs. (4.16) and (4.33)], a result which stems from the special form of the Eq. (4.2) determining T_D and its graphical interpretation. Some special anisotropic antiferromagnetic lattices are mentioned briefly in Sec. 5. Finally, in Sec. 6 the *exact* values of pair correlations at $T = 0$ are calculated for the special anisotropic antiferromagnetic lattice when the two weakest interactions are equal. In this case $T_N = T_D = 0$, so the Toeplitz determinants representing the correlations simplify, and can be evaluated exactly [see Eqs. (6.19) and (6.20)]. Some concluding remarks are made about possible generalizations of this work.

1. CLASSIFICATION OF ANISOTROPIC LATTICES

In this paper, we consider pair correlations between two spins on the same axis of a general anisotropic triangular lattice, with interaction energies $-J_1$, $-J_2$, $-J_3$ between parallel neighboring spins along row (**1**), column (**2**), and diagonal (**3**) axes, respectively [Fig. 1(a)]. The triangular lattice may alternatively be thought of as a quadratic lattice with rows (**1**) and columns (**2**) plus a single second-neighbor interaction along the diagonal direction (**3**) [Fig. 1(b)]. For a completely *ferromagnetic* lattice the J_l, $l = 1, 2, 3$, are

[12] R. M. F. Houtappel, Physica **16**, 425 (1950).
[13] G. H. Wannier, Phys. Rev. **79**, 357 (1950).
[14] For more detailed references, see Ref. 3.
[15] *Corresponding* ferromagnetic and antiferromagnetic lattices have interactions of equal magnitude and opposite sign.
[16] J. Stephenson, J. Math. Phys. **7**, 1123 (1966).
[17] H. S. Green and C. A. Hurst, *Order Disorder Phenomena* (Interscience Publishers, Inc., New York, 1964).

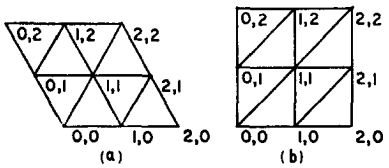


FIG. 1. (a) The triangular lattice and (b) the associated square lattice with a single second-nearest-neighbor bond.

all positive, and parallel spin states are energetically preferred. There is no loss of generality in supposing that $J_1 > J_2 > J_3 > 0$ in this case. For a completely *antiferromagnetic* lattice the J_l, $l = 1, 2, 3$, are all negative, and antiparallel spin states are energetically preferred. There is no loss of generality in supposing that $J_1 < J_2 < J_3 < 0$ in this case, so that the diagonal interaction J_3 is weakest. The pair correlation between two spins at sites **0** and **r** on the same lattice axis may be represented by a Toeplitz determinant[6,16] of order k, where k is a positive integer which is one greater than the number of lattice sites between **0** and **r** on the relevant lattice axis. Thus, for spins along row (**1**), $\mathbf{r} = (k, 0)$ and the separation of the sites **0** and **r** is $r \equiv |\mathbf{r}| = k$, measured in terms of lattice spacings. For spins along the diagonal (**3**), $\mathbf{r} = (k, k)$, and the separation r of the sites **0** and **r** depends on whether the lattice is triangular, as in Fig. 1(a), in which case $r = k$, or whether the lattice is square, as in Fig. 1(b), in which case $r = (2)^{\frac{1}{2}}k$. It is simplest to work in terms of k. The elements of the Toeplitz determinant are then

$$a_{p,q} = a_{q-p}, \quad p, q = 1, \cdots, k, \tag{1.1}$$

and depend only on the difference $(q - p)$ of row and column indices. Explicit formulas for the general element a_n, with $n = q - p$, have been derived elsewhere by the present author[16] and are alternatively available in the monograph by Hurst and Green.[17] If we use the notations

$$K_l = J_l/k_B T, \quad k_B \text{ Boltzmann's constant}, \tag{1.2}$$

$$C_l = \cosh 2K_l \quad \text{and} \quad S_l = \sinh 2K_l, \tag{1.3}$$

$l = 1, 2, 3$, then, for correlations along the (**3**) diagonal axis, we have

$$a_n = a_n(J_1, J_2, J_3)_3 = \frac{1}{2\pi}\int_{-\pi}^{\pi} d\omega e^{-in\omega} \times \frac{C_1C_2S_3 + S_1S_2C_3 - C_3\cos\omega + i\sin\omega}{|C_1C_2S_3 + S_1S_2C_3 - C_3\cos\omega + i\sin\omega|}. \tag{1.4}$$

The final subscript 3 in $a_n(J_1, J_2, J_3)_3$ refers to the (**3**) axis. Corresponding expressions for the other axes (**1**) and (**2**) are obtained by cyclic permutation of indices. Thus a_n is the coefficient of $e^{-in\omega}$ in the expansion of the generating function

$$A(\omega) = \frac{C_1C_2S_3 + S_1S_2C_3 - C_3\cos\omega + i\sin\omega}{|C_1C_2S_3 + S_1S_2C_3 - C_3\cos\omega + i\sin\omega|}. \tag{1.5}$$

The elements of the Toeplitz determinant have certain symmetry properties which will be useful in classifying the types of lattices and in reducing the number of cases which need be considered. These symmetry properties are, along the (**3**) axis,

$$\begin{aligned} a_n(J_1, J_2, J_3)_3 &= a_n(-J_1, -J_2, J_3)_3 \\ &= (-1)^{n+1}a_n(-J_1, J_2, -J_3)_3 \\ &= (-1)^{n+1}a_n(J_1, -J_2, -J_3)_3. \end{aligned} \tag{1.6}$$

These formulas show that the effect of reversing the signs of any two of the J_l, $l = 1, 2, 3$, may be taken into account quite easily. For example, properties of correlations on a lattice with *two* of the J_l negative may be obtained by transformation using (1.6) from correlations on a completely ferromagnetic lattice with all J_l positive. Lattices which may be transformed to a completely *ferromagnetic* lattice in this way will be called *Class A*. On the other hand, properties of correlations on a lattice with *one* of the J_l negative may be derived using (1.6) from correlations on a completely antiferromagnetic lattice with all J_l negative. Lattices which may be transformed to a completely *antiferromagnetic* lattice in this way will be called *Class B*. It may easily be seen that these cases cannot be transformed into one another, and that they exhaust the possibilities at this level of classification.

2. CLASSIFICATION OF GENERATING FUNCTIONS

Next we classify the types of generating functions by their distinguishing analytic structure. To facilitate the classification we employ the dual and inversion transformations.[3,18] The dual transformation relates a triangular lattice with parameters K_l to a honeycomb-lattice with parameters K_l^* given by

$$e^{-2K_l^*} = \tanh K_l \equiv v_l, \quad l = 1, 2, 3, \tag{2.1}$$

or alternatively,

$$\sinh 2K_l \sinh 2K_l^* = 1 \quad \text{and} \quad \cosh 2K_l^* = \coth 2K_l, \quad l = 1, 2, 3. \tag{2.2}$$

[18] I. Syozi and S. Naya, Progr. Theoret. Phys. (Kyoto) **24**, 829, (1960).

The inversion transformation relates triangular lattices with parameters K_l and K_l^+, $l = 1, 2, 3$, where

$$e^{-4K_3^+} = \frac{(v_1 + v_2v_3)(v_2 + v_1v_3)}{(1 + v_1v_2v_3)(v_3 + v_1v_2)}, \quad \text{and cyclic.} \tag{2.3}$$

Lattices for which K_l and K_l^+ are real exhibit a phase change at a critical temperature given by

$$|K_l| = |K_l^+|. \tag{2.4}$$

Successive application of the inversion and dual transformations (the order is irrelevant) is equivalent to relating a triangular lattice with parameters K_l by a star triangle transformation to a honeycomb lattice with parameters K_l^{+*}. It is easy to confirm that

$$\cosh 2K_3^{+*} = \coth 2K_3^+ = (C_1C_2S_3 + S_1S_2C_3)/S_3, \tag{2.5}$$

and thence rearrange expression (1.4) for the Toeplitz determinant element to

$$a_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} d\omega e^{-in\omega} \times \frac{(\cosh 2K_3^{+*} \sinh 2K_3 - \cosh 2K_3 \cos\omega + i \sin\omega)}{|\cosh 2K_3^{+*} \sinh 2K_3 - \cosh 2K_3 \cos\omega + i \sin\omega|}. \tag{2.6}$$

Now express $\sinh 2K_3$ in terms of $v_3 = \tanh K_3$, and multiply numerator and denominator in (2.6) by $(1 - v_3^2)$, which is a real positive quantity, to obtain

$$a_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} d\omega e^{-in\omega}(-e^{-i\omega}) \times \frac{(1 - v_3e^{2K_3^{+*}}e^{i\omega})(1 - v_3e^{-2K_3^{+*}}e^{i\omega})}{|(1 - v_3e^{2K_3^{+*}}e^{i\omega})(1 - v_3e^{-2K_3^{+*}}e^{i\omega})|}. \tag{2.7}$$

The form of the generating function in (2.7) is convenient for a study of correlations. Sometimes it will be useful to introduce an alternative notation

$$A^{-1} = v_3e^{2K_3^{+*}} \quad \text{and} \quad B = v_3e^{-2K_3^{+*}}, \tag{2.8}$$

so that $B = (v_3)^2A$, with $\operatorname{Re} K_3^{+*} > 0$. Let us observe that $\cosh 2K_3^{+*}$ is real, and distinguish the three cases

$$\begin{aligned} &\text{(a)}\ 1 < \cosh 2K_3^{+*}, \\ &\text{(b)}\ \cosh 2K_3^{+*} < -1, \\ &\text{(c)}\ -1 \leq \cosh 2K_3^{+*} \leq 1. \end{aligned} \tag{2.9}$$

These inequalities will now be re-expressed in terms of v_l variables, since their significance can then be appreciated more easily. In the following inequalities the upper sign refers to the case when $v_3 > 0$, and the lower sign to the case $v_3 < 0$. We first note that

$$(1 + v_1v_2v_3)(v_3 + v_1v_2) \gtrless (v_1 + v_2v_3)(v_2 + v_1v_3). \tag{2.10}$$

Case (a): We may take K_3^{+*} to be real and positive. Then $1 < \cosh 2K_3^{+*}$ if and only if

$$(v_1 + v_2v_3)(v_2 + v_1v_3) \gtrless 0, \tag{2.11}$$

which, in combination with (2.10), yields

$$(1 + v_1v_2v_3)(v_3 + v_1v_2) \gtrless (v_1 + v_2v_3)(v_2 + v_1v_3) \gtrless 0. \tag{2.12}$$

From this inequality and Eq. (2.3) we deduce that $0 < e^{-4K_3^+} < 1$, so K_3^+ is real and positive.

Case (b): We may take $K_3^{+*} = \operatorname{Re}(K_3^{+*}) + \frac{1}{2}i\pi$, with $\operatorname{Re}(K_3^{+*}) > 0$, so

$$\exp(2K_3^{+*}) = (-1)\exp(2 \operatorname{Re} K_3^{+*}).$$

Then $\cosh 2K_3^{+*} < -1$ if and only if

$$(1 + v_1v_2v_3)(v_3 + v_1v_2) \lesseqgtr 0, \tag{2.13}$$

which, in combination with (2.10), yields

$$(v_1 + v_2v_3)(v_2 + v_1v_3) \lesseqgtr (1 + v_1v_2v_3)(v_3 + v_1v_2) \lesseqgtr 0. \tag{2.14}$$

From this inequality and Eq. (2.3) we deduce that $1 < e^{-4K_3^+}$, so K_3^+ is real and negative.

Case (c): We may take $K_3^{+*} = i\frac{1}{2}\theta_3$, where θ_3 is real and positive. Then $-1 \leq \cosh 2K_3^{+*} = \cos\theta_3 \leq 1$ if and only if

$$(v_1 + v_2v_3)(v_2 + v_1v_3) \lesseqgtr 0 \lesseqgtr (1 + v_1v_2v_3)(v_3 + v_1v_2). \tag{2.15}$$

From this inequality and Eq. (2.3) we deduce that $e^{-4K_3^+}$ is negative, so K_3^+ is complex.

To use these inequalities, one determines which of them is satisfied for the lattice direction of interest and the temperature range under consideration, and one calculates K_3^+ from (2.3) and then $e^{2K_3^{+*}}$ from

$$e^{2K_3^{+*}} = \coth K_3^+. \tag{2.16}$$

It is important to note that the above inequalities are unaltered on changing the signs of both J_1 and J_2.

3 CLASS-*A* FERROMAGNETIC LATTICE WITH $J_1 > J_2 > J_3 > 0$

In this section we consider the completely ferromagnetic triangular lattice with $J_1 > J_2 > J_3 > 0$. To determine the analytic structure of the generating function, observe that $1 > v_1 > v_2 > v_3 > 0$, so that the only inequalities which may be satisfied are those

TABLE I.

T_C, Curie Point, J_1, J_2, and $J_3 > 0$	T_N, Néel Point, $J_1 < J_2 < J_3 < 0$	T_D, Disorder Point, $J_1 < J_2 < J_3 < 0$
$1 + v_1v_2v_3 = v_1 + v_2 + v_3 + v_1v_3 + v_2v_3 + v_1v_2$	$1 + v_1v_2v_3 = -v_1 - v_2 + v_3 - v_1v_3 - v_2v_3 + v_1v_2$	$v_3 + v_1v_2 = 0$
$C_1C_2S_3 + S_1S_2C_3 = C_3$ $C_1S_2C_3 + S_1C_2S_3 = C_2$ $S_1C_2C_3 + C_1S_2S_3 = C_1$	$C_1C_2S_3 + S_1S_2C_3 = C_3$ $C_1S_2C_3 + S_1C_2S_3 = -C_2$ $S_1C_2C_3 + C_1S_2S_3 = -C_1$	$C_1C_2S_3 + S_1S_2C_3 = -S_3$ $C_1S_2C_3 + S_1C_2S_3 = S_2$ $S_1C_2C_3 + C_1S_2S_3 = S_1$
$S_1S_2 + S_2S_3 + S_3S_1 = 1$ $z_1z_2 + z_2z_3 + z_3z_1 = 1$	$S_1S_2 - S_2S_3 - S_3S_1 = 1$ $z_1z_2 - z_2z_3 - z_3z_1 = 1$	$C_1C_2 - C_2C_3 - C_3C_1 = -1$ $z_1z_2 - z_1z_3 - z_3z_1 = -1$

of case (a) in (2.12) (upper sign). It follows that correlations along all three axes have the same qualitative features, so, without loss of generality, we consider only correlations in the diagonal (3) direction. Now K_3^+ and K_3^{+*} are real and positive. With notational abbreviations as in Eq. (2.8),

$$A^{-1} = v_3 e^{2K_3^{+*}} \quad \text{and} \quad B = v_3 e^{-2K_3^{+*}}, \tag{3.1}$$

we see that $0 \leq B < 1$, and that the *Curie point* T_C is determined by $A = 1$ or

$$v_3 e^{2K_3^{+*}} = 1. \tag{3.2}$$

In rearranged form

$$e^{-2K_3^+} \equiv \tanh K_3^{+*} = (1 - v_3)/(1 + v_3) \equiv e^{-2K_3}, \tag{3.3}$$

which is equivalent to $K_3^+ = K_3$. Now using the defining equation (2.3) for K_3^+, (3.3) can be written

$$(1 - v_3 - v_1v_2 + v_1v_2v_3)^2 = (v_1 + v_2 + v_2v_3 + v_3v_1)^2. \tag{3.4}$$

Taking the positive square root, we obtain the Curie-point equation[17]

$$1 + v_1v_2v_3 = v_1 + v_2 + v_3 + v_1v_2 + v_2v_3 + v_3v_1, \tag{3.5}$$

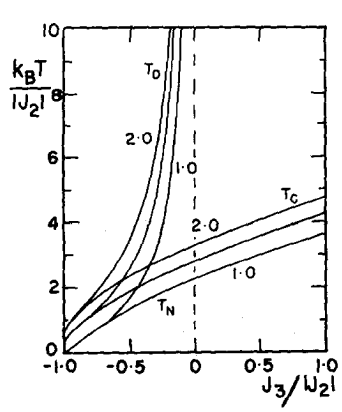


FIG. 2. Graphs of ferromagnetic Curie point T_C, and antiferromagnetic Néel point T_N, and disorder point T_D, in terms of $|J_2|/k_B$ vs $J_3/|J_2|$ for various fixed values of J_1/J_2. Graphs of T_C are in the right-hand half of the figure where $J_3 > 0$. Graphs of T_N are extensions of the T_C graphs into the left-hand half of the figure where $J_3 < 0$. The upper curves in the left-hand half of the figure are T_D curves. Ordering the graphs from the lowest, the values of J_1/J_2 for each graph are 1.0, 1.5, and 2.0.

which is symmetrical in v_1, v_2, and v_3. Various algebraic rearrangements of this equation are listed in the first column of Table I. For example, in terms of variables

$$z_l = e^{-2K_l}, \quad l = 1, 2, 3, \tag{3.6}$$

the Curie point T_C is determined from[3]

$$z_1z_2 + z_2z_3 + z_3z_1 = 1. \tag{3.7}$$

Graphs displaying the variation of k_BT_C/J_2 with J_3/J_2 for various fixed values of the ratio J_1/J_2 are sketched in the right-hand half of Fig. 2. It is straightforward to verify the following properties of A and B over the temperature intervals indicated:

$$\begin{aligned} &T < T_C\text{: } 0 < B < A < 1, \\ &T = T_C\text{: } A = 1,\ B = (\tanh K_3)^2, \\ &T > T_C\text{: } 0 < B < A^{-1} < 1. \end{aligned} \tag{3.8}$$

The generating function for the elements of the Toeplitz determinant now has the same form as that considered by Wu,[10] with the notational change $\alpha_2 \to A$, $\alpha_1 \to B$. Below the Curie point the generating function is

$$\frac{(1 - Ae^{-i\omega})(1 - Be^{i\omega})}{|(1 - Ae^{-i\omega})(1 - Be^{i\omega})|}, \tag{3.9}$$

with $0 < B < A < 1$. This form of the generating function gives rise to ferromagnetic long-range order. The limiting value of the correlation is, by application of Szegö's theorem,

$$\begin{aligned} \mathfrak{J}^2 &= \lim_{k \to \infty} \omega_2(k, k) \\ &= (1 - A^2)^{\frac{1}{4}}(1 - B^2)^{\frac{1}{4}}(1 - AB)^{-\frac{1}{2}} \\ &= [1 - (\sinh 2K_3^+/\sinh 2K_3)^2]^{\frac{1}{4}} \\ &= (1 - \mathfrak{K}^2)^{\frac{1}{4}}, \end{aligned} \tag{3.10}$$

where

$$\mathfrak{K}^2 = \frac{[(1 - v_1^2)(1 - v_2^2)(1 - v_3^2)]^2}{16(1 + v_1v_2v_3)(v_1 + v_2v_3)(v_2 + v_3v_1)(v_3 + v_1v_2)}, \tag{3.11}$$

and is symmetric in v_1, v_2, and v_3. $\mathfrak{J}$ is the spontaneous magnetization.[19,20] The asymptotic approach of the correlation to its limiting value has been studied by Wu,[10] and the results in his paper may be taken over with the appropriate notational changes:

$$\omega_2(k, k) - \mathfrak{J}^2 \sim \mathfrak{J}^2 \frac{A^{2k}}{2\pi k^2}\left[\frac{A}{1 - A^2}\right]^2\left[1 + O\left(\frac{1}{k}\right)\right]. \tag{3.12}$$

Above the critical point the generating function has the form

$$-e^{-i\omega}\frac{(1 - A^{-1}e^{i\omega})(1 - Be^{i\omega})}{|(1 - A^{-1}e^{i\omega})(1 - Be^{i\omega})|}, \tag{3.13}$$

with $0 < B < A^{-1} < 1$, which gives rise to ferromagnetic short-range order. From Wu's paper, the asymptotic form of the correlation is

$$\omega_2(k, k) \sim A^{-k}(\pi k)^{-\frac{1}{2}} \times (1 - A^{-2})^{-\frac{1}{4}}(1 - B^2)^{\frac{1}{4}}(1 - AB)^{-\frac{1}{2}}. \tag{3.14}$$

The decay of the correlation at the critical point[21] has been studied in detail by Wu[10] and by Hartwig and the present author.[22] For completeness, the corresponding result is

$$\omega_2(k, k) \sim \frac{E}{k^{\frac{1}{4}}}\left(\frac{1 + B}{1 - B}\right)^{\frac{1}{4}} \times \left[1 - \frac{1}{8k^2}\left(\frac{1}{8} - \frac{B}{(1 - B)^2}\right) + \cdots\right], \tag{3.15}$$

where

$$E = \exp\left(-\tfrac{1}{4}(1 + \gamma) - \sum_{s=2}^{\infty}\frac{\zeta(2s - 1)}{s4^s}\right) = 0.645002448\cdots, \tag{3.16}$$

and γ is Euler's constant. In particular, for the isotropic ferromagnetic triangular lattice at its Curie point, $A = 1$ and $B = 7 - 4(3)^{\frac{1}{2}}$, so

$$\omega_2(k, k) \sim E_0^T k^{-\frac{1}{4}}[1 + (192k^2)^{-1} + \cdots], \tag{3.17}$$

where

$$E_0^T = 3^{-\frac{1}{8}}2^{\frac{1}{4}}, E = 0.668618986\cdots. \tag{3.18}$$

Since the formulas above are valid for a general anisotropic ferromagnetic lattice, it is clear that, by specialization of the values of A and B, we may derive a variety of results. For example, if we set $J_3 = 0$ and

$$A = (\sinh 2K_1 \sinh 2K_2)^{-1}, \quad B = 0, \tag{3.19}$$

then the above formulas are valid for correlations along the diagonal direction of a quadratic lattice.

In summary, we note that Class A lattices exhibit a Curie point T_C which divides the temperature range $0 < T < \infty$ into two regions. Below T_C there is ferromagnetic long-range order and above T_C there is ferromagnetic short-range order.

4. CLASS-B ANTIFERROMAGNETIC LATTICE WITH $J_1 < J_2 < J_3 < 0$

In this section we consider the completely antiferromagnetic triangular lattice with $J_1 < J_2 < J_3 < 0$. To determine the structure of the generating function, observe that $-1 < v_1 < v_2 < v_3 < 0$, so that only lower inequalities in (2.12), (2.14), and (2.15) need be considered. It is especially important to observe the sign of $e^{-4K_3^+}$ in Eq. (2.3):

$$e^{-4K_3^+} = \frac{(v_1 + v_2v_3)(v_2 + v_1v_3)}{(1 + v_1v_2v_3)(v_3 + v_1v_2)}. \tag{4.1}$$

The choice of $J_1 < J_2 < J_3 < 0$ leads to the following conclusions. The factors $(v_1 + v_2v_3)$ and $(v_2 + v_1v_3)$ are negative at all temperatures. The sign of the factor $(v_3 + v_1v_2)$ depends on temperature, for the equation

$$v_3 + v_1v_2 = 0 \tag{4.2}$$

must be satisfied at some temperature T_D, above which $(v_3 + v_1v_2)$ is negative, and below which it is positive. We shall call T_D the *disorder point* (D ≡ disorder). When $T > T_D$, $e^{-4K_3^+}$ is negative and case (c) in Sec. 2 holds. When $T < T_D$, the lower inequality (2.14) of case (b) holds. Therefore, when $T < T_D$, K_3^+ is real and negative, and $K_3^{+*} = \text{Re}\,(K_3^{+*}) + \frac{1}{2}i\pi$ with $\text{Re}\,(K_3^{+*}) > 0$, so, with notational abbreviations analogous to those of Eq. (3.1),

$$A_3^{-1} = v_3 \exp(2K_3^{+*}) = (-1)v_3 \exp(2\,\text{Re}\,K_3^{+*}) \tag{4.3}$$

and $B_3 = v_3^2A_3$, we have $0 < A_3 < \infty$ and $0 < B_3 < 1$. The *Néel point* T_N (antiferromagnetic critical point) is determined by $A_3 = 1$, or

$$v_3e^{2K_3^{+*}} = 1. \tag{4.4}$$

In rearranged form,

$$e^{-2K_3^+} \equiv \tanh K_3^{+*} = (1 - v_3)/(1 + v_3) \equiv e^{-2K_3}, \tag{4.5}$$

which is equivalent to $K_3^+ = K_3$. Now using the defining equation (2.3) for K_3^+, (4.5) can be written

$$(1 - v_3 - v_1v_2 + v_1v_2v_3)^2 = (v_1 + v_2 + v_2v_3 + v_3v_1)^2, \tag{4.6}$$

[19] H. S. Green, Z. Physik. **171**, 129 (1963).
[20] R. B. Potts, Phys. Rev. **88**, 352 (1952).
[21] M. E. Fisher, Physica **25**, 521 (1959).
[22] R. E. Hartwig and J. Stephenson, J. Math. Phys. **9**, 836 (1968).

or

$$[(1 - v_3)(1 - v_1v_2)]^2 = [(1 + v_3)(v_1 + v_2)]^2. \quad (4.7)$$

Taking the (numerically) positive square root, we obtain the Néel-point equation[17]

$$1 + v_1v_2v_3 = -v_1 - v_2 + v_3 + v_1v_2 - v_2v_3 - v_3v_1, \quad (4.8)$$

which is symmetrical in v_1 and v_2, and may be rearranged in the three forms

$$\begin{aligned} -(v_1 + v_2)(1 + v_3) &= (1 - v_3)(1 - v_1v_2), \\ (v_3 - v_2)(1 - v_1) &= (1 + v_1)(1 + v_2v_3), \\ (v_3 - v_1)(1 - v_2) &= (1 + v_2)(1 + v_1v_3), \end{aligned} \quad (4.9)$$

which are satisfied if and only if J_1, $J_2 < J_3 < 0$. Various other algebraic rearrangements of this equation are listed in the second column of Table I. For example, in terms of z_l variables in Eq. (3.6), the Néel point T_N is determined from

$$z_1z_2 - z_2z_3 - z_3z_1 = 1. \quad (4.10)$$

Equation (4.2) which determines T_D may also be expressed in terms of z_l variables:

$$z_1z_2 - z_2z_3 - z_3z_1 = -1. \quad (4.11)$$

There is a remarkable "similarity" between the equation determining the Curie point T_C of a ferromagnetic lattice (3.7), and Eqs. (4.10) and (4.11) above determining the Néel point T_N and disorder point T_D of an antiferromagnetic lattice. Equation (4.11) for T_D may be rearranged in the three ways

$$\begin{aligned} 1 + z_1z_2 &= z_3(z_1 + z_2), \\ 1 - z_1z_3 &= z_2(z_3 - z_1), \\ 1 - z_2z_3 &= z_1(z_3 - z_2), \end{aligned} \quad (4.12)$$

which have a solution if and only if J_3 is the weakest interaction and is negative. Various other algebraic rearrangements of Eq. (4.11) for T_D are listed in the third column of Table I. Graphs displaying the variation of $k_BT_N/|J_2|$ and $k_BT_D/|J_2|$ with $J_3/|J_2|$ for various fixed values of the ratio J_1/J_2 are sketched in the left-hand half of Fig. 2.

Thus, for the anisotropic antiferromagnetic triangular lattice, we must consider correlations separately in the three temperature regions $0 < T < T_N$, $T_N < T < T_D$, and $T_D < T < \infty$, and at two special points T_N and T_D. However, as we shall see later, the lattice is ordered below T_N and disordered above T_N, so there are only two thermodynamic phases. The equations determining T_N and T_D were obtained using inequalities derived in Sec. 2 for correlations along the diagonal (3) direction of the triangular lattice, but the same equations and the division of the whole temperature range $0 < T < \infty$ into three regions arise also from analysis of correlations along the other lattice axes. In fact, alternative definitions of T_N and T_D can be given independent of the correlations, though this point is not considered further here.

A. Pair Correlation Along Diagonal (3) Axis

The interaction J_3 in the diagonal direction is weakest, and the behavior of the pair correlation in this direction is quite different from that along the other two axes. There are three temperature ranges and two special points T_N and T_D to consider.

Case 1: $0 < T < T_N$. Using the variables A_3 and B_3 defined in Eq. (4.3), we observe, when $0 < T < T_N$, that $0 < B_3 < A_3 < 1$. The generating function (2.7) now has the same form as that in Eq. (3.9), which was derived in Sec. 3 for a ferromagnet below its Curie point. Although J_3 is negative, there is ferromagnetic long-range order below T_N, and the pair correlations along the (3) axis are all positive. The long-range order parameter

$$\mathfrak{R} = \lim_{k \to \infty} \omega_2(k, k) \quad (4.13)$$

is given by Eq. (3.10).

Case 2: $T = T_N$. Now $A_3 = 1$ and $0 < B_3 < 1$, so that Eqs. (3.15) and (3.16) determine the behavior of the correlation. The Néel point may be "low" for an antiferromagnetic lattice (Fig. 2), so the decay amplitude of the critical point correlation in the (3) direction may be "large."

Case 3: $T_N < T < T_D$. In this temperature range, $1 < A_3 < \infty$ and $0 < B_3 < 1$. From Eqs. (3.13) and (3.14), we deduce that there is ferromagnetic short-range order. If the two weakest interactions J_3 and J_2 are comparable in strength, then T_N and T_D are "close" together (Fig. 2), and the short-range order region is "small." Note that, below T_D, the pair correlations along the (3) axis are all positive and decrease monotonically with increasing spin separation. J_3 is the weakest interaction, and below T_D its effect is swamped by the stronger interactions J_1 and J_2. As $T \to T_D$ from below, $(v_3 + v_1v_2) \to 0$ from above, and $K_3^+ \to -\infty$. The inversion lattice, which is antiferromagnetic along its (3) direction, has reached absolute zero! Also $A_3^{-1} = v_3e^{2K_3^{+*}} \to -v_3$ and $B_3 \to -v_3$, so the

asymptotic formula (3.14) for the correlation between T_N and T_D breaks down.

Case 4: $T = T_D$. T_D has been called the *disorder point* because it separates two disordered regions, in contrast to the Néel point which separates regions of long- and short-range order. At T_D the generating function (2.7) now simplifies to

$$-e^{-i\omega}\left(\frac{1 + v_3 e^{i\omega}}{1 + v_3 e^{-i\omega}}\right), \tag{4.14}$$

whence the Toeplitz determinant elements a_n are

$$\begin{aligned} a_n &= 0, \quad n \geq 1, \\ a_0 &= -v_3, \\ a_n &= -(-v_3)^{n-1}(1 - v_3^2), \quad n \leq -1. \end{aligned} \tag{4.15}$$

The Toeplitz determinant is trivial to evaluate at T_D, and the exact value of the correlation is

$$\omega_2(k, k) = (-v_3)^k, \tag{4.16}$$

which represents exponential decay. It is interesting to note that the zero-field pair correlation between two spins separated by a distance k on a linear chain with interaction energy $-J$ is just v^k, where

$$v = \tanh (J/k_B T).^{23}$$

Case 5: $T > T_D$. The factor $(v_3 + v_1 v_2)$ is negative, $e^{-4K_3^+}$ is negative, and case (c) in Sec. 2 holds. Set $K_3^{+*} = i\frac{1}{2}\theta_3$, where θ_3 is real and positive. The generating function (2.7) is now

$$-e^{-i\omega}\frac{(1 - v_3 e^{i\theta_3} e^{i\omega})(1 - v_3 e^{-i\theta_3} e^{i\omega})}{|(1 - v_3 e^{i\theta_3} e^{i\omega})(1 - v_3 e^{-i\theta_3} e^{i\omega})|}, \tag{4.17}$$

and has the same form as that considered by the present author in the preceding paper of this series[24] in a discussion of pair correlations for the isotropic antiferromagnetic triangular lattice. The calculation there is general enough so that it can be taken over to the present case with the notational changes $v \to v_3$ and $\theta \to \theta_3$. The asymptotic form of the correlation is [quoting the leading term only from Eq. (1.31) of Ref. 24]

$$\omega_2(k, k) \sim (\tfrac{1}{2}\pi \sin \theta_3)^{-\frac{1}{2}} v_3^k k^{-\frac{1}{2}} \times \cos (k\theta_3 + \tfrac{1}{2}\theta_3 - \tfrac{1}{4}\pi - \phi_3), \tag{4.18}$$

where

$$\phi_3 = -\tfrac{1}{2} \arg (1 - v_3^2 \cos 2\theta_3 + i v_3^2 \sin 2\theta_3) \tag{4.19}$$

[23] J. S. Marsh, Phys. Rev. **145**, 251 (1966).
[24] J. Stephenson, J. Math. Phys. **11**, 413 (1969) (preceding article).

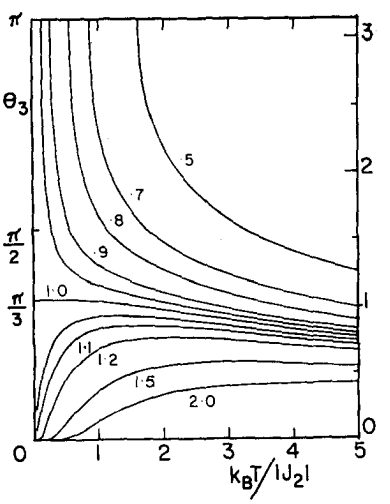


FIG. 3. Graphs of θ_3 vs $k_B T/|J_2|$ for various fixed values of J_3/J_2 in the special case $J_1 = J_2 < 0$. Ordering the graphs from the uppermost, the values of J_3/J_2 for each graph are 0.5, 0.7, 0.8, 0.9, 0.95, 1.0, 1.05, 1.1, 1.2, 1.5, 2.0.

and θ_3 is the real angle lying between 0 and π determined by

$$\cos \theta_3 = (C_1 C_2 S_3 + S_1 S_2 C_3)/S_3. \tag{4.20}$$

The exponential decay with oscillatory envelope of the pair correlation is characteristic of the triangular lattice above its T_D point. Figure 3 displays graphs of θ_3 vs a temperature variable $k_B T/|J_2|$ for various fixed values of J_3/J_2 in the special case when $J_1 = J_2$. Now $\cos \theta_3 = \cosh 2K_3^{+*} = \coth 2K_3^+$, so at T_D, $\theta_3 = \pi$. Then, as T increases from T_D to ∞, θ_3 decreases from π to 0, achieving the value $\frac{1}{2}\pi$ at a temperature $2T_D$. Notice that the nearest-neighbor pair correlation vanishes at $2T_D$, and is positive for lower temperatures.

B. Pair Correlation along Row (1) Axis

While discussing correlations in the (1) direction with the strongest interaction J_1, we shall refer to formulas in Sec. 2 and make the appropriate cyclic permutation of subscripts $3 \to 1$, $2 \to 3$, $1 \to 2$, in them. For the (1) direction

$$e^{-4K_1^+} = \frac{(v_2 + v_3 v_1)(v_3 + v_2 v_1)}{(1 + v_1 v_2 v_3)(v_1 + v_2 v_3)}. \tag{4.21}$$

When $T > T_D$, the factor $(v_3 + v_2 v_1)$ is negative, so $e^{-4K_1^+}$ is negative, and case (c) in Sec. 2 holds because the cyclic permutation of the lower inequality in (2.15) is satisfied:

$$(v_2 + v_3 v_1)(v_3 + v_2 v_1) > 0 > (1 + v_1 v_2 v_3)(v_1 + v_2 v_3). \tag{4.22}$$

When $T < T_D$, the cyclic permutation of the lower inequality in (2.12) is satisfied:

$$(1 + v_1 v_2 v_3)(v_1 + v_2 v_3) < (v_2 + v_3 v_1)(v_3 + v_2 v_1) < 0, \tag{4.23}$$

and we refer to case (a) in Sec. 2, where K_1^{+*} and K_1^+ are real and positive. With notational abbreviations analogous to those of Eq. (3.1),

$$A_1^{-1} = v_1 e^{2K_1{}^{+*}}$$

and

$$B_1 = v_1^2 A_1, \tag{4.24}$$

we have $-\infty < A_1 < 0$ and $-1 < B_1 < 0$. The Néel point T_N is determined by $A_1 = -1$, or

$$v_1 e^{2K_1{}^{+*}} = -1. \tag{4.25}$$

In rearranged form

$$e^{-2K_1{}^+} \equiv \tanh K_1^{+*} = (1 + v_1)/(1 - v_1) \equiv e^{+2K_1}, \tag{4.26}$$

which is equivalent to $K_1^+ = -K_1$. Now using the defining equation (4.21) for K_1^+, we can write (4.26) as

$$(1 + v_1 + v_2v_3 + v_1v_2v_3)^2 = (-v_2 + v_3 - v_3v_1 + v_1v_2)^2, \tag{4.27}$$

or

$$[(1 + v_1)(1 + v_2v_3)]^2 = [(1 - v_1)(v_3 - v_2)]^2. \tag{4.28}$$

Taking the (numerically) positive square root, we re-obtain the Néel-point equation (4.8). This checks the consistency of determining the Néel point from

$$|K_l^+| = |K_l|, \quad l = 1, 2, 3. \tag{4.29}$$

Now let us determine the asymptotic form of the pair correlation in the three temperature ranges $0 < T < T_N$, $T_N < T < T_D$, $T_D < T < \infty$, and at the two special points T_N and T_D. It turns out that correlations in the (1) direction, with the strongest interaction J_1, have the expected antiferromagnetic form below T_D.

Case 1: $0 < T < T_N$. Using the variables A_1 and B_1 defined in Eq. (4.24), we observe, when $0 < T < T_N$, that $-1 < A_1 < B_1 < 0$. The generating function in (2.7), with subscript 3 replaced by 1, now has the same form as that in Eq. (3.9), which was derived in Sec. 3 for a ferromagnet below its Curie point, except that A_1 and B_1 are now negative. There is antiferromagnetic long-range order below T_N, and the long-range order parameter

$$\mathfrak{R} = \lim_{k \to \infty} (-1)^k \omega_2(k, 0) \tag{4.30}$$

is given by Eq. (3.10) again, since

$$(-1)\frac{\sinh 2K_1^+}{\sinh 2K_1} = (-1)\frac{\sinh 2K_2^+}{\sinh 2K_2} = \frac{\sinh 2K_3^+}{\sinh 2K_3} = \mathcal{K}, \tag{4.31}$$

where $\mathcal{K}$ is the positive root of Eq. (3.11).

Case 2: $T = T_N$. Now $A_1 = -1$ and $-1 < B_1 < 0$ so that Eqs. (3.15) and (3.16) determine the behavior of the correlation. If the Néel point is "low," then the decay amplitude of the critical point correlation in the (1) direction will be "small."

Case 3: $T_N < T < T_D$. In this temperature range, $-\infty < A_1 < 1$ and $-1 < B_1 < 0$. From Eqs. (3.13) and (3.14), we deduce that there is antiferromagnetic short-range order. Note that, below T_D, the pair correlations along the (1) axis alternate in sign $(-)^k$ and decrease monotonically in magnitude with increasing spin separation. As $T \to T_D$ from below, $K_1^+ \to \infty$.

Case 4: $T = T_D$. At T_D the generating function for the (1) direction now simplifies to

$$-e^{-i\omega}\left(\frac{1 - v_1 e^{i\omega}}{1 - v_1 e^{-i\omega}}\right), \tag{4.32}$$

and the exact formula for the correlation is

$$\omega_2(k, 0) = v_1^k. \tag{4.33}$$

Case 5: $T > T_D$. Above T_D, the factor $(v_3 + v_1v_2)$ is negative, $e^{-4K_1{}^+}$ is negative, and the inequality (4.22) holds. Set $K_1^{+*} = i\frac{1}{2}\theta_1$, where θ_1 is real and positive. The generating function now has the same form as Eq. (4.17), with subscript 3 replaced by 1. Accordingly, the asymptotic form of the pair correlation is given by (4.18) and (4.19) with subscript 3 replaced by 1, and θ_1 is the real angle lying between 0 and π determined by

$$\cos\theta_1 = (C_2C_3S_1 + S_2S_3C_1)/S_1. \tag{4.34}$$

Figure 4 displays graphs of θ_1 vs the temperature variable $k_BT/|J_2|$ for various fixed values of J_3/J_2 in the special case when $J_1 = J_2$. Now $\cos\theta_1 = \cosh 2K_1^{+*} = \coth 2K_1^+$, so at T_D, $\theta_1 = 0$. Then, as T increases from T_D to ∞, θ_1 increases to some maximum value, which must be less than $\frac{1}{2}\pi$, since $e^{-4K_1{}^+}$ is finite, and subsequently decreases to zero again at $T = \infty$.

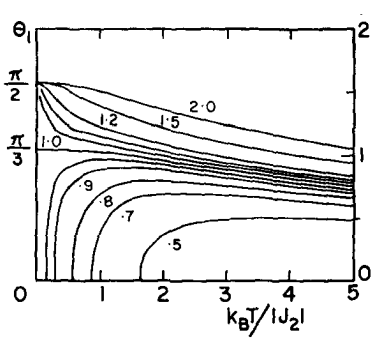


FIG. 4. Graphs of θ_1 vs $k_BT/|J_2|$ for various fixed values of J_3/J_2 in the special case $J_1 = J_2 < 0$. Ordering the graphs from the lowest, the values of J_3/J_2 for each graph are: 0.5, 0.7, 0.8, 0.9, 0.95, 1.0, 1.05, 1.1, 1.2, 1.5, 2.0.

Correlations in the (2) *direction* with intermediate strength interaction J_2 have the same qualitative features as those in the (1) direction. Formulas for this case may be obtained by an appropriate cyclic permutation of subscripts. At T_D, the correlations decay exponentially along the lattice axes, in a form suggesting that the spins along the lattice axes behave like "linear chains."[23]

5. SPECIAL CASES OF ANISOTROPIC ANTIFERROMAGNETIC LATTICES

In special limiting cases, the three-region lattice, with finite T_N and T_D points, may lose one or two of its regions. So far we have retained $J_1 < J_2 < J_3 < 0$. In the trivial special case $J_3 = 0$, we regain the anisotropic antiferromagnetic quadratic lattice ($T_D = \infty$) which has two phases. Now suppose we fix J_1 and J_2, and increase the strength of J_3. While $J_2 < J_3 < 0$, the lattice exhibits three regions. When $J_3 \to J_2 + <0$, $T_D \to 0$. The antiferromagnetic ordered and disordered regions below T_D are eliminated, and only the single triangular lattice type disordered region remains. Trivially $T_N = T_D = 0$, as is shown by Eqs. (4.9) and (4.12). Once $J_3 < J_2 < 0$, the J_2 interaction becomes weakest and the (2) and (3) directions interchange roles. The three region lattice system returns, and T_N and T_D rise again from $T = 0$. Increasing $|J_3|$ further, past $|J_1|$ for example, does not alter the situation in a qualitative manner. In the more special case when $J_1 = J_2 < 0$, a three-region system is obtained when $J_1 < J_3 < 0$. The lattice obtained when $J_3 < J_1 = J_2 < 0$ has one phase of triangular lattice type disorder with $T_N = T_D = 0$, and is considered in more detail in the next section. For details in the case of the completely isotropic lattice, see Ref. 24.

Another case of some interest is when $J_1 > J_2 > 0$ and $J_3 < 0$, which we may regard as that of a ferromagnetic quadratic lattice with a single second-neighbor antiferromagnetic interaction J_3. The formulas of Sec. 4 show that the (3) direction correlation is unaltered by changing the signs of J_1 and J_2. The modification to the discussion of correlations in the (1) and (2) directions consists in deleting the "anti" from antiferromagnetic, which is equivalent to using the transformation properties expressed in Eq. (1.6).

6. CLASS-*B* ANTIFERROMAGNETIC LATTICE WITH $J_3 < J_1 = J_2 < 0$

The distinguishing feature of this special case is that the two weakest interactions J_1 and J_2 are equal. From Eqs. (4.9) and (4.12), the Néel point T_N and disorder point T_D are at zero temperature, and the lattice exhibits one phase of triangular-lattice-type disorder. The asymptotic behavior of the pair correlations along a lattice axis l, $l = 1, 2, 3$, is determined over the whole temperature range $0 < T < \infty$ by

$$\omega_2(\mathbf{r}) \sim (\tfrac{1}{2}\pi \sin\theta_l)^{-\frac{1}{2}} v_l^k k^{-\frac{1}{2}} \cos(k\theta_l + \tfrac{1}{2}\theta_l - \tfrac{1}{4}\pi - \phi_l), \tag{6.1}$$

obtained by generalizing Eq. (4.18) to an arbitrary lattice axis l, where

$$\phi_l = -\tfrac{1}{2} \arg(1 - v_l^2 \cos 2\theta_l + i v_l^2 \sin 2\theta_l), \tag{6.2}$$

and θ_l are real angles for $T > T_D$ defined by

$$\cos\theta_3 = (C_1C_2S_3 + S_1S_2C_3)/S_3, \text{ and cyclic.} \tag{6.3}$$

In the special case under consideration here, $J_3 < J_1 = J_2 < 0$, so

$$\cos\theta_3 = 1 + S_1^2 e^{2K_3}/S_3 \tag{6.4}$$

and

$$\cos\theta_1 = \cos\theta_2 = C_1 e^{2K_3}. \tag{6.5}$$

The formula (6.1) was derived in Sec. 4 (see Ref. 24) under the assumption that $0 < |v_l| < 1$ for $T > 0$, and is invalid at $T = 0$. The remainder of this section is devoted to obtaining the exact values of the pair correlations along the lattice axes at $T = 0$. The (1) and (2) axes are equivalent since $J_1 = J_2$, but the (3) axis must be treated separately.

First we need the properties of θ_3 and $\theta_1 = \theta_2$, which are displayed in Figs. 3 and 4, where graphs of θ_3 and θ_1 as functions of a temperature variable $t = k_B T/|J_2|$ are sketched for various fixed values of J_3/J_2 with $J_1 = J_2$. For the isotropic lattice, set $\theta_1 = \theta_2 = \theta_3 = \theta$. θ has the zero-point value $\frac{1}{3}\pi$, and the curve of θ vs t for the isotropic lattice divides the curves of θ_3 and θ_1 vs t into two classes accordingly as $J_3 \gtrless J_2 = J_1$. The zero-point values of θ_3 and $\theta_1 = \theta_2$ when $J_3 < J_2 = J_1 < 0$ are

$$\lim_{T\to 0} \theta_3 = 0 \tag{6.6}$$

and

$$\lim_{T\to 0} \theta_1 = \tfrac{1}{2}\pi. \tag{6.7}$$

It is important to note that limiting processes $T \to 0$ and $J_3 \to J_2 = J_1$ are *not* interchangeable, and from the results of this section we cannot deduce any corresponding results for the isotropic lattice.[24]

A. (3) Direction at $T = 0$

Let $\theta_3 = 0$ and $v_3 = -1$ in Eq. (4.17) for the elements of the Toeplitz determinant representing the pair correlation. Then

$$a_0 = -1, \quad a_n = 0, \quad n \neq 0, \tag{6.8}$$

so the exact zero-point value of the correlation along the (3) axis is

$$\omega_2(k, k) = (-1)^k, \tag{6.9}$$

corresponding to a rigid arrangement of spins in regular antiferromagnetic order, alternately up and down [compare Eq. (4.16) for the correlation at T_D, and note that the present result (6.9) is *not* obtained by setting $T_D = 0$ in (4.16)]. This arrangement is to be expected at $T = 0$ because, when $J_3 < J_2 = J_1 < 0$, J_3 is the strongest interaction, and the ground state of a triangle of spins has energy $-|J_3|$. That is, all rows of spins in the (3) direction achieve antiferromagnetic order like a set of linear chains. For a lattice of N spins, there will be $O(N^{\frac{1}{2}})$ chains of spins in the (3) direction, which may be arranged in $2^{O(N^{\frac{1}{2}})}$ ways compatible with the ground state, so there is no zero-point entropy per site for an infinite lattice.

B. (1) Direction at $T = 0$

Spins along the (1) and (2) axes have similar properties since $J_1 = J_2$. Set $\theta_1 = \frac{1}{2}\pi$ and $v_1 = -1$ in the generating function for the (1) direction obtained by making the appropriate modification of Eq. (4.17); replace the subscript 3 by 1. Then

$$\begin{aligned} a_n &= (-1)^{\frac{1}{2}(n-1)}[-2/(\pi n)], \quad n \text{ odd}, \\ &= 0, \qquad\qquad\qquad\quad n \text{ even}. \end{aligned} \tag{6.10}$$

Since $a_{p,q} = a_{q-p} = 0$ when p and q are both odd or both even,

$$\omega_2(k, 0) = \det (a_{p,q})_{k\times k} = 0 \quad \text{for odd } k. \tag{6.11}$$

In particular, the nearest-neighbor correlation is zero in the (1) direction. For even spin separation, set $k = 2m$. The $k \times k$ Toeplitz determinant can be rearranged by m^2 row and column transpositions to the form

$$(-1)^m \det \begin{pmatrix} \mathsf{B}_m^T & 0 \\ 0 & \mathsf{B}_m \end{pmatrix} = (-1)^m (\det \mathsf{B}_m)^2, \tag{6.12}$$

where B_m is an $m \times m$ matrix whose (i, j) element b_{ij} is equal to $a_{(2i,2j-1)}$. Explicitly,

$$b_{i,j} = -\frac{2}{\pi}\frac{(-1)^{j-i-1}}{[2(j-i)-1]}, \quad i, j = 1, \cdots, m. \tag{6.13}$$

Now multiply the even numbered rows and columns of B_m by -1, and extract a factor $-2/\pi$ from each element to obtain a new matrix C_m whose determinant $\det \mathsf{C}_m = (-\frac{1}{2}\pi)^m \det \mathsf{B}_m$, and whose elements are

$$c_{i,j} = [2(i-j)+1]^{-1}, \quad i, j = 1, \cdots, m. \tag{6.14}$$

In detail,

$$\mathsf{C}_m = \begin{pmatrix} 1 & -1 & -\frac{1}{3} & -\frac{1}{5} & \cdots \\ \frac{1}{3} & 1 & -1 & -\frac{1}{3} & \cdots \\ \frac{1}{5} & \frac{1}{3} & 1 & -1 & \cdots \\ \frac{1}{7} & \frac{1}{5} & \frac{1}{3} & 1 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & & \end{pmatrix}_{m\times m}. \tag{6.15}$$

But C_m is precisely the Cauchy matrix which appears in the calculation of the pair correlation in the diagonal direction of a ferromagnetic square lattice at its Curie point.[5,10,21,22] Its determinant may be calculated exactly as

$$\det \mathsf{C}_m = \prod_{s=1}^{m} \frac{\Gamma(s)\Gamma(s)}{\Gamma(s-\frac{1}{2})\Gamma(s+\frac{1}{2})}, \tag{6.16}$$

and asymptotically for large m as

$$\det \mathsf{C}_m \sim \left(\frac{\pi}{2}\right)^m \frac{E}{m^{\frac{1}{4}}}\left(1 - \frac{1}{64m^2} + \cdots\right), \tag{6.17}$$

where $E = 0.645002448$ is given by Eq. (3.16). For even spin separation $k = 2m$, we therefore have

$$\begin{aligned} \omega_2(2m, 0) &= (-1)^m\left[\left(\frac{2}{\pi}\right)^m \det \mathsf{C}_m\right]^2 \\ &\sim (-1)^m \frac{E^2}{m^{\frac{1}{2}}}\left[1 - \frac{1}{32m^2} + \cdots\right]. \end{aligned} \tag{6.18}$$

Finally, combining the results for odd and even spin separation k, we have

$$\begin{aligned} \omega_2(k, 0) &= \cos(\tfrac{1}{2}\pi k)\left[\left(\frac{2}{\pi}\right)^{\frac{1}{2}k} \prod_{s=1}^{\frac{1}{2}k} \frac{\Gamma(s)\Gamma(s)}{\Gamma(s-\frac{1}{2})\Gamma(s+\frac{1}{2})}\right]^2 \\ &\sim 2^{\frac{1}{2}}E^2 k^{-\frac{1}{2}} \cos(\tfrac{1}{2}\pi k)[1 - (8k^2)^{-1} + \cdots], \end{aligned} \tag{6.19}$$

where

$$2^{\frac{1}{2}}E^2 = 0.588352663\cdots. \tag{6.20}$$

The above results are exact. Qualitatively, they are just what one gets by setting $\theta_1 = \frac{1}{2}\pi$ in (6.1) and letting $v_1 \to -1$, an invalid procedure! There is a curious similarity of form between Eq. (6.20), obtained here for pair correlations along the (1) direction when $J_3 < J_1 = J_2 < 0$, and that for pair correlations at $T = 0$ on the isotropic antiferromagnetic triangular lattice[24]:

$$\omega_2(k, 0) \sim \epsilon_0 k^{-\frac{1}{2}} \cos(\tfrac{2}{3}\pi k). \tag{6.21}$$

The exact value of ϵ_0 is plausibly conjectured to be

$$\epsilon_0 = 2^{\frac{1}{2}}(E_0^T)^2 = 0.632226080\cdots, \tag{6.22}$$

where E_0^T is the decay amplitude of the pair correlation

at the Curie point of an isotropic ferromagnetic triangular lattice, as in Sec. 3, Eqs. (3.17) and (3.18). In particular, note the appearance in (6.20) and (6.21) of a decay amplitude related to the *square* of a ferromagnetic critical point decay amplitude. This gives one additional confidence in the conjectured value of ϵ_0 in (6.22).

7. CONCLUDING REMARKS

The most important and interesting results obtained in this paper are those for the asymptotic behavior of pair correlations along the axes of a *two-dimensional*, *antiferromagnetic*, *anisotropic*, *triangular*, *Ising* lattice. The italicized words indicate five features which in conjunction determine the correlation properties we have derived here. It would be interesting to know whether the results we have obtained can be generalized to other antiferromagnetic lattices. The occurrence of a *disorder point* T_D, as we have defined it (4.2), is related to the existence of triangles in an anisotropic lattice. For this reason we may expect some of the present considerations to carry over to an antiferromagnetic Kagomé lattice, for example, and possibly also to some three-dimensional lattices, such as the hyper-Kagomé, hypertriangular, and face-centered cubic.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

On Hidden-Variable Theories

STANLEY P. GUDDER
Department of Mathematics, University of Denver, Denver, Colorado 80210

(Received 15 January 1969)

An abstract definition of a general hidden-variables theory is given, and it is shown that such a theory is always possible in the present framework of quantum mechanics and is, in fact, unique in a certain sense. It is noted that the Bohm–Bub hidden-variables example is contained in this theory and an attempt is made to clarify the position of this theory with respect to hidden-variable impossibility proofs. The general definition is used in the consideration of quantum-mechanical ordering and the measurement process.

1. INTRODUCTION

The idea that there is a possibility of a hidden-variables (HV) theory in quantum mechanics is almost as old as quantum mechanics itself. The first mathematical refutation of even the *possibility* of an HV theory in quantum mechanics was given by von Neumann in 1932.[1] However, the proponents of hidden variables dispensed with this refutation for various reasons and continued their insistence. Recently this has spurred a rash of new proofs concerning the impossibility of HV theories.[2–6] Strangely enough, there have also appeared explicit examples of HV theories.[7] Clearly there is something wrong here. One obviously cannot have an HV theory if it is impossible; yet the originators of these impossibility proofs have not explained the reason for this discrepency. The problem, in the author's opinion, is in the definition of an HV theory. The proponents of HV theories have an idea of what these theories should be and have given examples of such theories. The antagonists have a different idea of what an HV theory should be and have proved that such theories are impossible in the present general framework of quantum mechanics. These proofs are irrelevant since they do not refer to the HV theories as formulated by the advocates of these theories, and this point is stressed in a paper by Bub.[8] These proofs put much more stringent requirements on the HV theories than the HV researchers need or want for their theories. The opponents seem to think that the HV researchers advocate a return to classical mechanics or at least an embedding of quantum mechanics into a classical mechanical framework. However, this is not so. As the author sees it, the HV advocates feel that it is within the realm of possibility for determinism to be introduced in quantum mechanics. As a result,

[1] J. von Neumann, *The Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, N.J., 1953).
[2] J. Jauch and C. Piron, Helv. Phys. Acta **36**, 827 (1963); S. Gudder, Proc. Am. Math. Soc. **19**, 319 (1968).
[3] See, for example, J. Turner, J. Math. Phys. **9**, 1411 (1968).
[4] N. Zierler and M. Schlessinger, Duke J. **32**, 251 (1965).
[5] S. Kochen and E. Specker, J. Math. Mech. **17**, 59 (1967).
[6] B. Misra, Nuovo Cimento **47**, 843 (1967).
[7] D. Bohm, Phys. Rev. **85**, 166, 180 (1952); D. Bohm and J. Bub, Rev. Mod. Phys. **38**, 453 (1966).
[8] J. Bub, Intern. J. Theoret. Phys. (to be published).

at the Curie point of an isotropic ferromagnetic triangular lattice, as in Sec. 3, Eqs. (3.17) and (3.18). In particular, note the appearance in (6.20) and (6.21) of a decay amplitude related to the *square* of a ferromagnetic critical point decay amplitude. This gives one additional confidence in the conjectured value of ϵ_0 in (6.22).

7. CONCLUDING REMARKS

The most important and interesting results obtained in this paper are those for the asymptotic behavior of pair correlations along the axes of a *two-dimensional*, *antiferromagnetic*, *anisotropic*, *triangular*, *Ising* lattice. The italicized words indicate five features which in conjunction determine the correlation properties we have derived here. It would be interesting to know whether the results we have obtained can be generalized to other antiferromagnetic lattices. The occurrence of a *disorder point* T_D, as we have defined it (4.2), is related to the existence of triangles in an anisotropic lattice. For this reason we may expect some of the present considerations to carry over to an antiferromagnetic Kagomé lattice, for example, and possibly also to some three-dimensional lattices, such as the hyper-Kagomé, hypertriangular, and face-centered cubic.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

On Hidden-Variable Theories

STANLEY P. GUDDER
Department of Mathematics, University of Denver, Denver, Colorado 80210

(Received 15 January 1969)

An abstract definition of a general hidden-variables theory is given, and it is shown that such a theory is always possible in the present framework of quantum mechanics and is, in fact, unique in a certain sense. It is noted that the Bohm–Bub hidden-variables example is contained in this theory and an attempt is made to clarify the position of this theory with respect to hidden-variable impossibility proofs. The general definition is used in the consideration of quantum-mechanical ordering and the measurement process.

1. INTRODUCTION

The idea that there is a possibility of a hidden-variables (HV) theory in quantum mechanics is almost as old as quantum mechanics itself. The first mathematical refutation of even the *possibility* of an HV theory in quantum mechanics was given by von Neumann in 1932.[1] However, the proponents of hidden variables dispensed with this refutation for various reasons and continued their insistence. Recently this has spurred a rash of new proofs concerning the impossibility of HV theories.[2–6] Strangely enough, there have also appeared explicit examples of HV theories.[7] Clearly there is something wrong here. One obviously cannot have an HV theory if it is impossible; yet the originators of these impossibility proofs have not explained the reason for this discrepency. The problem, in the author's opinion, is in the definition of an HV theory. The proponents of HV theories have an idea of what these theories should be and have given examples of such theories. The antagonists have a different idea of what an HV theory should be and have proved that such theories are impossible in the present general framework of quantum mechanics. These proofs are irrelevant since they do not refer to the HV theories as formulated by the advocates of these theories, and this point is stressed in a paper by Bub.[8] These proofs put much more stringent requirements on the HV theories than the HV researchers need or want for their theories. The opponents seem to think that the HV researchers advocate a return to classical mechanics or at least an embedding of quantum mechanics into a classical mechanical framework. However, this is not so. As the author sees it, the HV advocates feel that it is within the realm of possibility for determinism to be introduced in quantum mechanics. As a result,

[1] J. von Neumann, *The Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, N.J., 1953).
[2] J. Jauch and C. Piron, Helv. Phys. Acta **36**, 827 (1963); S. Gudder, Proc. Am. Math. Soc. **19**, 319 (1968).
[3] See, for example, J. Turner, J. Math. Phys. **9**, 1411 (1968).
[4] N. Zierler and M. Schlessinger, Duke J. **32**, 251 (1965).
[5] S. Kochen and E. Specker, J. Math. Mech. **17**, 59 (1967).
[6] B. Misra, Nuovo Cimento **47**, 843 (1967).
[7] D. Bohm, Phys. Rev. **85**, 166, 180 (1952); D. Bohm and J. Bub, Rev. Mod. Phys. **38**, 453 (1966).
[8] J. Bub, Intern. J. Theoret. Phys. (to be published).

it may be possible to construct a measurement process describing the "collapse of the wavefunction," and as a by-product make von Neumann's projection postulate unnecessary. In this paper the author gives a general definition of what he feels an HV theory should be and proves that, in the present framework of quantum mechanics, such a theory is *always* possible and is, in fact, unique in a certain sense. This strengthens the results of previous authors who have given examples of such theories in specific circumstances. It also has the advantage of placing HV theories into an abstract framework so that the general properties of all such theories can be considered.

2. IMPOSSIBILITY PROOFS

Let us briefly consider some of the proofs of the impossibility of an HV theory. Some of these proofs have already been criticized,[9] but the author feels that he has a slightly different approach; and, furthermore, there are a few points that should be emphasized to motivate the present HV construction.

Von Neumann[10] claims that an HV theory would imply the existence of dispersion-free states (we shall give precise definitions in the next section) on the system of quantum-mechanical propositions. These propositions were assumed to consist of the set of all closed subspaces L_H of a Hilbert space H. (We shall not consider the technicality of superselection rules here.) Thus, he felt that an HV theory should imply that one can determine precisely the outcome of any experiment or measurement; hence, all dispersions must be zero and probabilistic considerations made to vanish. We contend that, although von Neumann's general idea is correct, his mathematical definition is too restrictive. We contend that the dispersion-free states need not be defined on the entire proposition system L_H since HV theories only propose to allow predictions concerning single measurements. Bohm[11] has repeatedly emphasized that the hidden parameters vary according with the different mutually exclusive experimental arrangements of matter that must be used in making different kinds of measurements. For example, to make a position or momentum measurement, two different sets of hidden variables would have to be used. Thus, the dispersion-free states must only be defined on outcomes corresponding to *single* measurements.

Similar criticisms can be made about the generalization of von Neumann's proof.[2] In these papers a more general structure is taken for the system of quantum-mechanical propositions. Although these papers fall short of the mark, what they are aiming for is the following theorem:

Theorem 1: If there is a full set of dispersion-free states on a quantum proposition system L, then L is a Boolean algebra.

However, even if this theorem were true, it would not make HV theories impossible. Again, this is because HV theories do not imply the existence of dispersion-free states defined on all of L but only those defined on Boolean sub-σ-algebras of L, since they only determine outcomes of individual experiments.

It is felt by some[3] that the paper of Zierler and Schlessinger,[4] which considers imbeddings of the quantum proposition system into Boolean algebras, gives a proof of the impossibility of HV theories. However, HV theories do not suggest that quantum mechanics be imbedded in a classical structure. Similarly, Kochen and Specker[5] define an HV theory as an imbedding of the set of quantum observables into the set of dynamical variables on a phase space. Again, this is a claim that one can imbed quantum mechanics into a classical system, which is not what the HV proponents mean an HV theory to be. Finally, Misra's proof[6] takes a different axiomatic structure for quantum mechanics, but also claims that hidden variables imply the existence of dispersion-free states, which are now taken as monotonic positive functionals on the system of quantum observables.

3. GENERAL HIDDEN-VARIABLE THEORIES

We first give an English-language version of what we feel the HV researchers mean by an HV theory.

"The state m of a quantum-mechanical system is not complete in the sense that another variable ξ can be adjoined to m so that the pair (m, ξ) completely determines the system. That is, a knowledge of (m, ξ) enables one to predict precisely the outcome of any single measurement. Furthermore, an average over the values of ξ gives the usual quantum state m."

We shall now attempt to translate this version of an HV theory into a mathematical-language version. We use the general proposition system formulation of quantum mechanics.[12] A *proposition system* is

[9] D. Bohm and J. Bub, Rev. Mod. Phys. **38**, 470 (1966); J. Bell, *ibid.* **38**, 447 (1966); **40**, 228 (1968); see also Ref. 4.

[10] See Ref. 1, pp. 295–328.

[11] See Bohm, Ref. 7.

[12] J. Jauch, *Foundations of Quantum Mechanics* (Addison–Wesley Publ. Co., Reading, Mass., 1968); V. Varadarajan, *Geometry of Quantum Theory* (D. Van Nostrand, Princeton, N.J., 1968); G. Mackey, *Mathematical Foundations of Quantum Mechanics* (W. A. Benjamin, New York, 1963).

a partially ordered set $L = \{a, b, c, \cdots\}$ with first and last elements zero and one, respectively, an orthocomplementation $a \to a'$ satisfying $a'' = a$, $a \leq b$ implies $b' \leq a'$, and $a \vee a' = 1$; furthermore, if a_i is a sequence of disjoint ($a_i \leq a_j'$, $i \neq j$) elements, then $\bigvee a_i$ exists in L. A state m is a map $m: L \to [0, 1]$ such that $m(1) = 1$ and $m(\bigvee a_i) = \sum m(a_i)$ if a_i is a disjoint sequence. A set of states M on L is *full* if for any $a \neq 0$ there is an $m \in M$ such that $m(a) = 1$ and if $a \neq b$ there is an $m \in M$ such that $m(a) \neq m(b)$. A *quantum proposition system* is a pair (L, M) where L is a proposition system and M is a full set of states on L. Notice that (L, M) generalizes the standard quantum-structure (L_H, M). A *dispersion-free state* s on L is a map $s: L \to \{0, 1\}$ such that $s(1) = 1$ and

$$s\left(\bigvee_{i=1}^{n} a_i\right) = \sum_{i=1}^{n} m(a_i)$$

if a_i is a finite disjoint set of propositions. Now single measurements correspond to Boolean sub-σ-algebras of L. To say that the results of a measurement (corresponding to a Boolean sub-σ-algebra $B \subseteq L$) are completely determined means that one has a dispersion-free state s on B. We denote the set of dispersion-free states on B by M_B. We are now ready to formulate our definition of an HV theory.

Definition: A quantum proposition system (L, M) *admits an* HV *theory* if there is a probability space (Ω, F, μ) with the property that, for any maximal Boolean sub-σ-algebra $B \subseteq L$, there is a map H_B from $M \times \Omega$ onto M_B such that:

(i) $\omega \to H_B(m, \omega)(a)$ is measurable for every $m \in M$, $a \in B$;

(ii) $\int_\Omega H_B(m, \omega)(a)\, d\mu(\omega) = m(a)$ for every $m \in M$, $a \in B$.

Denote the set of maximal Boolean sub-σ-algebras of L by $\mathfrak{B}$. We call $((\Omega, F, \mu), \{H_B: B \in \mathfrak{B}\})$ an HV theory for (L, M). An HV theory $((\Omega, F, \mu), \{H_B: B \in \mathfrak{B}\})$ is *minimal* if $H_B(m, \omega_1) = H_B(m, \omega_2)$, for all $m \in M$ and $B \in \mathfrak{B}$, implies $\omega_1 = \omega_2$.

We consider only maximal Boolean sub-σ-algebras so that the theory does not get too cumbersome. This is really only a technicality since any Boolean sub-σ-algebra is contained in a maximal one. The probability space (Ω, F, μ) may be thought of as the space of hidden variables. If an HV theory is minimal, there is a minimal number of hidden variables—just enough to give all the dispersion-free states. Notice that the Bohm–Bub theory[13] gives part of an HV theory as defined above. Bohm and Bub only consider one measurement such as a spin measurement and hence their theory gives the map $H_B: M \times \Omega \to M_B$ for one single Boolean σ-algebra B. The Bohm–Bub example is not minimal since many hidden variables correspond to each dispersion-free state. The nonminimal hidden-variable theories are probably the more interesting ones. In the literature,[14] two-valued homomorphisms are often considered instead of dispersion-free states. However, we now show that these two concepts are equivalent. But first we need some definitions.

Two propositions $a, b \in L$ are *compatible* (written $a \leftrightarrow b$) if there exist mutually disjoint propositions a_1, b_1, c such that $a = a_1 \vee c$, $b = b_1 \vee c$. Let B_2 denote the Boolean algebra with two elements zero and one. A *two-valued homomorphism* f on L is a map $f: L \to B_2$ such that $f(a') = f(a)'$; and if $a \leftrightarrow b$, then $f(a \vee b) = f(a) \vee f(b)$ and $f(a \wedge b) = f(a) \wedge f(b)$. We identify B_2 with the pair of real numbers $\{0, 1\}$ in the natural way.

Lemma: m is a dispersion-free state on L if and only if m is a two-valued homomorphism on L.

Proof: Suppose m is a two-valued homomorphism on L. Notice first that there is an $a \in L$ such that $m(a) = 1$ since if $m(b) = 0$, then $m(b') = 1$. Now, for $a \in L$,

$$m(1) = m(1 \vee a) = (m(1) \vee m(a) \geq m(a).$$

Hence $m(1) = 1$. Next we notice that m preserves order since if $a \leq b$, then $m(b) = m(a \vee b) = m(a) \vee m(b) \geq m(a)$. It follows that if a and b are disjoint, so are $m(a)$ and $m(b)$. Let a_i, $i = 1, \cdots, n$, be mutually disjoint. If $m(\bigvee a_i) = 1$, then $\bigvee m(a_i) = 1$. It follows that there is an a_j such that $m(a_j) = 1$ and $m(a_i) = 0$, $i \neq j$. Hence,

$$\sum_{i=1}^{n} m(a_i) = 1 = m(\bigvee a_i).$$

If $m(\bigvee a_i) = 0$, then $\bigvee m(a_i) = 0$ and $m(a_i) = 0$, $i = 1, \cdots, n$. Hence,

$$\sum_{i=1}^{n} m(a_i) = 0 = m(\bigvee a_i)$$

and m is a dispersion-free state. Conversely, suppose m is a dispersion-free state on L. Then

$$1 = m(a \vee a') = m(a) + m(a')$$

and, hence,

$$m(a') = 1 - m(a) = m(a)'.$$

[13] See Bohm and Bub, Ref. 7.

[14] See, for example, Refs. 4 and 5.

If $a \leftrightarrow b$, we can write

$$a \vee b = [a \wedge b] \vee [b \wedge (a \wedge b)'] \vee [a \wedge (a \wedge b)'],$$

where the terms in square brackets are mutually disjoint. It then follows that

$$m(a \vee b) = m(a) + m(b) - m(a \wedge b).$$

If $m(a \vee b) = 1$, then either $m(a) = 1$ or $m(b) = 1$, so

$$m(a) \vee m(b) = 1 = m(a \vee b).$$

If $m(a \vee b) = 0$, then both $m(a) = 0$ and $m(b) = 0$, so

$$m(a) \vee m(b) = 0 = m(a \vee b).$$

It now follows that $m(a \wedge b) = m(a) \wedge m(b)$ and, thus, m is a two-valued homomorphism.

We are now ready to prove our main theorem.

Theorem 2: Any quantum proposition system (L,M) admits a minimal HV theory $((\Omega, F, \mu), \{H_B: B \in \mathfrak{B}\})$. Furthermore, $((\Omega, F, \mu), \{H_B: B \in \mathfrak{B}\})$ is the unique minimal HV theory in the sense that if $((\Omega', F', \mu'), \{H'_B: B \in \mathfrak{B}\})$ is another HV theory, there is a measurable map τ from Ω' onto Ω such that

$$H_B(m, \tau\omega')(a) = H'_B(m, \omega')(a)$$

for every $B \in \mathfrak{B}$, $\omega' \in \Omega'$, $m \in M$, and $a \in B$,

$$\mu'(\tau^{-1}(\Lambda)) = \mu(\Lambda)$$

for every $\Lambda \in F$, and if $((\Omega', F', \mu'), \{H'_B: B \in \mathfrak{B}\})$ is minimal, then τ is one-to-one.

Proof: If B is a maximal Boolean sub-σ-algebra, it is in particular a Boolean algebra and, hence, by Stone's representation theorem,[15] there is an isomorphism h_B from B onto a Boolean algebra G_B of open and closed subsets of a compact space Ω'_B. If $m \in M$, define $\tilde{m}_0$ on G_B by $\tilde{m}_0(\Lambda) = m(h_B^{-1}(\Lambda))$. We now show $\tilde{m}_0$ is countably additive on G_B. Let $\Lambda_i \in G_B$, $i = 1, 2, \cdots$, be disjoint and suppose $\bigcup \Lambda_i \in G_B$. Because of the compactness of Ω'_B, all but finitely many Λ_i's are empty. We thus have

$$\tilde{m}_0(\bigcup \Lambda_i) = m(h_B^{-1}(\bigcup \Lambda_i)) = m(\bigvee h_B^{-1}(\Lambda_i)) = \sum m(h_B^{-1}(\Lambda_i)) = \sum m_0(\Lambda_i).$$

Hence, $\tilde{m}_0$ is countably additive on G_B and so, by the Hahn extension theorem,[16] $\tilde{m}_0$ has a unique extension to a measure $\tilde{m}$ on the σ-field H_B generated by G_B. Notice that K_B is contained in the Borel field of Ω'_B. Thus, $(\Omega'_B, K_B, \tilde{m})$ is a probability space. We now form the product probability space (Ω_B, F_B, μ_B), where

$$\Omega_B = \prod_{m \in M} \Omega_m, \quad F_B = \prod_{m \in M} K_m, \quad \mu_B = \prod_{m \in M} \tilde{m},$$

where $\Omega_m = \Omega'_B$ and $K_m = K_B$ for all $m \in M$. That is, Ω_B is the Cartesian product of Ω'_B with itself to the cardinality of M, F_B is the σ-field generated by the cylinder sets according to Kolmogorov's construction,[17] and μ_B is the product measure defined by

$$\mu_B(S_{m_1} \times S_{m_2} \times \cdots \times S_{m_n} \times \prod{}' \Omega_m) = \tilde{m}_1(S_{m_1})\tilde{m}_2(S_{m_2}) \cdots \tilde{m}_n(S_{m_n}),$$

where $\prod' \Omega_m = \prod \{\Omega_m: m \neq m_1, \cdots, m_n\}$. The space (Ω_B, F_B, μ_B) exists by Kolmogorov's theorem.[17] By Tychonoff's theorem, note that Ω_B is compact and F_B is contained in the Borel field. We can then use Kolmogorov's theorem again to form the product probability space (Ω, F, μ), where

$$\Omega = \prod_{B \in \mathfrak{B}} \Omega_B, \quad F = \prod_{B \in \mathfrak{B}} F_B, \quad \text{and} \quad \mu = \prod_{B \in \mathfrak{B}} \mu_B.$$

Now $B \in \mathfrak{B}$ defines the map $H_B: M \times \Omega \to M_B$ by $H_B(m, \omega)(a) = 1$, if $\omega(B)(m) \in h_B(a)$, and

$$H_B(m, \omega)(a) = 0,$$

otherwise. We now show that $H_B(m, \omega)(\cdot)$ is a dispersion-free state. Certainly $H_B(m, \omega)(1) = 1$. Suppose a_i are mutually disjoint, $i = 1, \cdots, n$. If

$$H_B(m, \omega)(\bigvee a_i) = 0,$$

then

$$\omega(B)(m) \notin h_B(\bigvee a_i) = \bigcup h_B(a_i).$$

Thus,

$$\omega(B)(m) \notin h_B(a_i), \quad H_B(m, \omega)(a_i) = 0,$$

$i = 1, \cdots, n$, and

$$0 = H_B(m, \omega)(\bigvee a_i) = \sum_{i=1}^{n} H_B(m, \omega)(a_i).$$

If $H_B(m, \omega)(\bigvee a_i) = 1$, then

$$\omega(B)(m) \in h_B(\bigvee a_i) = \bigcup h_B(a_i).$$

Since $h_B(a_i)$, $i = 1, \cdots, n$, are mutually disjoint, there is a $1 \leq j \leq n$ such that $\omega(B)(m) \in a_j$, $\omega(B)(m) \notin a_i$, and $i \neq j$. Hence,

$$H_B(m, \omega)(a_i) = \delta_{ij}$$

and

$$1 = H_B(m, \omega)(\bigvee a_i) = \sum_{i=1}^{n} H_B(m, \omega)(a_i),$$

and $H_B(m, \omega)(\cdot)$ is a dispersion-free state. Now $\omega \to H_B(m, \omega)(a)$ is measurable since it has the value

[15] M. Stone, Trans. Am. Math. Soc. **40**, 37 (1936).

[16] P. Halmos, *Measure Theory* (D. Van Nostrand, Princeton, N.J., 1950), p. 54.

[17] A. Kolmogorov, *Foundations of Probability* (Chelsea Publ. Co., New York, 1950), p. 29.

one on the cylinder set

$$\Lambda = \{\omega \in \Omega: \omega(B) \in h(a) \times \prod \{\Omega_{m'}: \Omega_{m'} = \Omega'_B, m' \neq m\}\}$$
$$= \{h(a) \times \prod \{\Omega_{m'}: \Omega_{m'} = \Omega'_B, m \neq m'\} \times \prod \{\Omega_{B'}: B' \neq B\}\},$$

and the value zero otherwise. Also,

$$\int_\Omega H_B(m, \omega)(a)\, d\mu(\omega)$$
$$= \int_\Lambda d\mu(\omega) = \mu(\Lambda)$$
$$= \mu(h_B(a) \times \prod \{\Omega_{m'}: \Omega_{m'} = \Omega'_B, m' \neq m\})$$
$$= \tilde{m}(h_B(a)) = m(a),$$

for all $m \in M$, $a \in B$. Now the map $\omega \to \omega(B)(m)$ is onto Ω'_B by definition. It can be shown[18] that the Stone space Ω'_B is isomorphic to the set of two-valued homomorphisms on B. By the lemma this latter set is the same as M_B and, hence, H_B is surjective. Now suppose $((\Omega', F', \mu'), \{H'_B: B \in \mathfrak{B}\})$ is another hidden variable theory. Define $\tau: \Omega' \to \Omega$ by $\tau(\omega')(B)(m) = H'_B(m, \omega')$. τ is surjective since H'_B is. The measurability of τ follows from the measurability of $\omega' \to H'_B(m, \omega')$, and $H_B(m, \tau\omega')(a) = H'_B(m, \omega')(a)$ by definition. The rest of the proof is left to the reader.

4. REMARKS

As was mentioned in Sec. 2, Kochen and Specker define an HV theory as an imbedding of the partial algebra of observables into an algebra of dynamical variables on a phase space. Their main claim is that it is not enough to preserve the statistics in this imbedding but also functional relationships between observables must be preserved. They give a simple example in which the statistics are preserved but in which the random variables corresponding to different observables are independent. Although their example has a primitive similarity to our construction, this type of behavior does not occur here. In this theory, observables are σ-homomorphisms from the Borel subsets of the real line $B(R)$ into L. There is a one-to-one correspondence between observables and Boolean sub-σ-algebras of L in the sense that the range $R(x)$ of an observable x is a Boolean sub-σ-algebra and every (separable) Boolean sub-σ-algebra is the range of an observable.[19]

Let x be an observable with range $R(x)$ and let B be a maximal Boolean sub-σ-algebra containing $R(x)$. Now, of course, $R(x)$ may be contained in many different maximal Boolean sub-σ-algebras and these correspond to different experimental arrangements used to measure x. In our construction we have represented B (via an isomorphism h_B) by a Boolean algebra of subsets of a set Ω'_B. One would say that a function

$$f: \Omega'_B \to R$$

represents x if

$$x(E) = h_B^{-1}(f^{-1}(E)) \quad \text{for all} \quad E \in B(R).$$

Now there is, in general, no function f representing x, although there would be if h_B^{-1} were a σ-homomorphism.[19] However, such functions would exist in special cases (e.g., if x had only a finite number of values), so let us suppose there is one, say f. Now if an observable y is a function of x, say $y = g(x)$ [this means $y(E) = x(g^{-1}(E))$ for all $E \in B(R)$], then $R(y) \subseteq R(x) \subseteq B$. If y is represented by a function $p: \Omega'_B \to R$, then one can show that $p = g \circ f$, and thus functions of observables are preserved in this theory. More generally, suppose that x and y are compatible observables [i.e., $x(E) \leftrightarrow y(F)$ for all E, $F \in B(R)$]. If we are to think of these observables as being measured at the same time, then we must find a Boolean sub-σ-algebra B containing $R(x) \cup R(y)$. Such a B always exists.[19] It can be shown that there exists an observable z and Borel functions u and v such that $R(z) \subseteq B$, $x = u(z)$, and $y = v(z)$.[19] Now if we can represent x, y, and z by functions f, g, and h, respectively, then it can be shown that $f = u \circ h$ and $g = v \circ h$. So, again, functions of compatible observables are preserved. One would get independence of the representing functions f and g if one considered different maximal Boolean sub-σ-algebras B_1 and B_2 containing $R(x)$ and $R(y)$, respectively; but this would physically involve one measurement for x and a separate measurement for y, and one would then expect to obtain independent values.

Turner[20] defines an HV theory in terms of an imbedding σ of the propositions L (and states) into a Boolean algebra L' of subsets of a phase space, although he only requires a statistics-preserving imbedding. He then discusses the quantum ordering of the propositions. He makes the claim that if σ preserves the quantum order, then, according to the work of Zierler and Schlessinger,[21] the imbedding would be trivial and hence impossible. This claim seems incorrect. In fact, Zierler and Schlessinger have shown that an order-preserving imbedding is *always* possible. What they do show is that there may

[18] P. Halmos, *Lectures on Boolean Algebras* (D. Van Nostrand, Princeton, N.J., 1963), p. 78.
[19] V. Varadarajan, Commun. Pure Appl. Math. **15**, 129 (1962).
[20] See Ref. 3.
[21] See Ref. 4.

be no order-preserving imbedding σ that also satisfies $\sigma(a \vee b) = \sigma(a) \vee \sigma(b)$ when $a \leftrightarrow b$. Next, Turner claims that the HV theory of Bohm–Bub induces an imbedding of the propositions of the 3-dimensional Hilbert space $|\psi\rangle = \psi_1 |s_1\rangle + \psi_2 |s_2\rangle + \psi_3 |s_3\rangle$ that is not order preserving. This is incorrect for two reasons. First, the Bohm–Bub theory implies no such imbedding. The theory concerns only one measurement, that of spin, which corresponds only to a Boolean algebra B with eight elements. Of course, this Boolean algebra has an imbedding h_B of the strongest kind, but none is implied for the set of all propositions. Secondly, this imbedding h_B preserves quantum order, as does any HV theory.[22] We make this precise in the next paragraph.

We say that an HV theory $((\Omega, F, \mu), \{H_B: B \in \mathcal{B}\})$ for (L, M) *preserves quantum order* if $a \leq b$ if and only if, for every $B \in \mathcal{B}$ satisfying $a, b \in B$, we have

$$H_B(m, \omega)(a) \leq H_B(m, \omega)(b) \quad \text{for all} \quad m \in M, \omega \in \Omega.$$

Using Theorem 2, one can now prove the following:

Theorem 3: Any HV theory preserves quantum order.

There is an unusual kind of phenomenon that occurs in HV theories that does not appear in the quantum theory, however. In quantum mechanics the distribution for the values of an observable is independent of the type of measurement used to find these values. This is not, in general, true for HV theories. For example, suppose we are interested in a particular outcome represented by the proposition $a \in L$. Suppose the "completed" state of the system is (m, ω). If we make a measurement corresponding to a Boolean sub-σ-algebra $B_1 \in \mathcal{B}$ with $a \in B_1$, then the outcome occurs if $H_{B_1}(m, \omega)(a) = 1$. Now if we make a measurement corresponding to a Boolean sub-σ-algebra $B_2 \in \mathcal{B}$ with $a \in B_2$, it is entirely possible that $H_{B_2}(m, \omega) \neq H_{B_1}(m, \omega)(a)$. This is easily seen to occur in specific examples.[22] Thus, although a is seen to occur in one measurement, it does not occur in a different type of measurement even if the completed state (m, ω) is the same. It is possible that this kind of phenomena can be shown experimentally, giving empirical evidence that an HV theory is necessary for a finer understanding of nature than that given by quantum mechanics.

This general HV theory can also be used to describe a model for the measurement process. When the system is not being measured, then the states will evolve according to some dynamical law $V_t: M \to M$. However, when a measuring instrument (corresponding to some Boolean sub-σ-algebra B) is introduced into the system's environment, the hidden-variable space (Ω_B, F_B, μ_B) enters and the completed states $M \times \Omega_B$ evolve according to a new law $U_t: M \times \Omega_B \to M \times \Omega_B$. After a certain time (which, in practice, would be short) the completed states will approach a certain point $(m_0, \omega_0) \in M \times \Omega_B$, where m_0 is the "wavefunction collapsed to" in the usual quantum theory, and m_0 restricted to B gives the resulting dispersion-free state s. In this case, we would have $H_B(m, \omega) = s$, if (m, ω) is the initial completed state.

ACKNOWLEDGMENTS

The author would like to thank J. Tutsch and J. Bub for many helpful discussions concerning HV theories.

[22] For more details and a slightly different viewpoint see the forthcoming paper by J. Tutsch, "Measurement in Quantum Mechanics," Phys. Rev. (to be published).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Domain of Dependence

ROBERT GEROCH*
Department of Mathematics, Birkbeck College. London, England

(Received 13 March 1969)

The various properties of the domain of dependence (Cauchy development) which have been found particularly useful in the study of gravitational fields are reviewed. The basic techniques for constructing proofs and counterexamples are described. A new tool—the past and future volume functions—for treating certain global properties of space-times is introduced. These functions are used to establish two new theorems: (1) a necessary and sufficient condition that a space-time have a Cauchy surface is that it be globally hyperbolic; and (2) the existence of a Cauchy surface is a stable property of space-times.

1. INTRODUCTION

The concept of the domain of dependence, or Cauchy development, of a spacelike 3-surface has recently found a number of important applications to global problems in general relativity, e.g., in the proofs (and often in the statements) of theorems about singularities[1-6] in the initial-value problem for gravitational and other fields[7] and in various related questions. Roughly speaking, a point p is in the domain of dependence of a surface S if initial data, given on S, completely and uniquely determine the state of any system at p. The domain of dependence and causal structure[4,8,9] must be ranked as the two most important tools presently available for the study of global properties of space-times.

Global hyperbolicity was invented for dealing with hyperbolic differential equations on a manifold.[10,11] A space-time is said to be globally hyperbolic if, given any two of its points, the collection of all timelike or null (causal) curves joining these points is compact (in a suitable topology). This condition requires, intuitively, that there be no "asymptotic regions," "holes," or "singularities" in the region between the points. Global hyperbolicity is the natural condition to impose, for example, to ensure the existence and uniqueness of solutions—in particular, Green's-function solutions—of hyperbolic equations.[10,11] This condition has also been used in the study of gravitational fields, e.g., in the initial-value problem.[12]

The goals of the present paper are, firstly, to summarize those properties of the domain of dependence which have been found useful in the study of gravitational fields and, secondly, to obtain some new results concerning the domain of dependence and its relation to global hyperbolicity.

In Sec. 2 we introduce the three basic definitions: the domain of dependence, the Cauchy horizon, and a Cauchy surface. (The Cauchy horizon consists of certain boundary points of the domain of dependence. A surface is a Cauchy surface if its domain of dependence is the entire space-time.) We illustrate these definitions with seven examples, intended to suggest a number of basic features of the domain of dependence, the Cauchy horizon, and a Cauchy surface and—what appears to be of at least equal importance in this subject—a number of properties that these objects do not possess. In Sec. 3 we prove a selection of results. Using the constructions and counterexamples of Secs. 2 and 3, the reader should have little trouble in deciding, from the almost infinite number of plausible-sounding conjectures, which are true and which are false. Since there are such a large number of useful but comparatively minor properties of the domain of dependence, it is almost impossible to try to remember them all; for this reason the examples and constructions are as important as the statements of theorems. Sections 2 and 3 are a review in the sense that all of the material, if not published before, is known.

In Sec. 4 we define global hyperbolicity and introduce a topology on certain sets of curves.

Global hyperbolicity and the existence of a Cauchy surface—two properties of space-times each widely

* Present address: Department of Physics, Syracuse University, Syracuse, New York.

[1] R. Penrose, Phys. Rev. Letters **14**, 57 (1965).
[2] S. W. Hawking, Proc. Roy. Soc. (London) **294A**, 511 (1966).
[3] S. W. Hawking, Proc. Roy. Soc. (London) **295A**, 490 (1966).
[4] S. W. Hawking, Proc. Roy. Soc. (London) **300A**, 187 (1967).
[5] R. Geroch, "Singularities," in *Proceedings of the Midwestern Relativity Conference*, L. Witten, Ed. (to be published).
[6] R. Penrose, *Battelle Recontres*, C. M. DeWitt and J. A. Wheeler, Eds. (W. A. Benjamin, Inc., New York, 1968), Chap. VII.
[7] Y. Choquet-Bruhat and R. Geroch, "Global Aspects of the Cauchy Problem in General Relativity" Comm. Math. Phys. (to be published).
[8] E. H. Kronheimer and R. Penrose, Proc. Cambridge Phil. Soc. **63**, 481 (1967).
[9] R. Geroch, "Space-Time Structure from a Global Viewpoint," in *Proceedings of The International School of Physics Enrico Fermi, 1969* (Academic Press Inc., to be published).
[10] J. Leray, "Hyperbolic Partial Differential Equations" (mimeographed notes, Princeton, 1952).
[11] Y. Choquet-Bruhat, *Battelle Recontres*, C. M. DeWitt and J. A. Wheeler, Eds. (W. A. Benjamin, Inc., New York, 1968), Chap. IV.
[12] Y. Choquet-Bruhat, *Gravitation: An Introduction to Current Research*, L. Witten, Ed. (John Wiley & Sons, Inc., New York, 1962), Chap. 4.

used in its own area of relativity theory—are equivalent: a necessary and sufficient condition that a space–time have a Cauchy surface is that it be globally hyperbolic. The proof, given in Sec. 5, involves the past and future volumes. Because these volume functions appear to be a useful technique for dealing with global properties of space–times, their properties are summarized.

Finally, in Sec. 6 we show that the existence of a Cauchy surface is a stable property of space–times. That is, arbitrary sufficiently small variations in the metric (in a suitable topology) do not destroy the existence of a Cauchy surface.

2. DEFINITIONS AND EXAMPLES

By a *space–time* M we understand a 4-dimensional manifold with a smooth metric g_{ab} of signature $(+, -, -, -)$. We shall also assume that M is time-oriented, that is, that the light cones of M are divided into two systems, past and future.[13]

The fundamental notion is that of the domain of dependence, or Cauchy development, of an achronal set S in M. (A set is said to be *achronal*[9] if no two of its points may be joined by a timelike curve.) It will be convenient to think of S as some spacelike or null 3-surface; in fact, as we shall see later, the domain of dependence is uninteresting except in this case. Roughly speaking, a point p is in the domain of dependence of S if the state of any system at p can be completely specified by initial conditions given on S. Because signals must travel along timelike or null curves, we should expect that the initial data on S would completely determine the situation at p if and only if every such curve from p strikes S. More precisely, *the future domain of dependence of* S, $D^+(S)$, *consists of those points* p *such that every past-directed timelike curve from* p *without a past endpoint*[14] *intersects* S.[1,2,6,9,15] The *past domain of dependence*, $D^-(S)$, is defined in a similar way by interchanging the roles of past and future. Finally, the *(total) domain of dependence* is the set $D(S) = D^+(S) \cup D^-(S)$.

The following remarks will serve to clarify these definitions. It was necessary, first of all, to include the condition "without a past endpoint" in the definition of the domain of dependence, since otherwise $D^+(S)$ would always coincide with S itself: If q is any point not in S, then one can always find a past-directed timelike curve from q (*with* a past endpoint) which does not intersect S. Secondly, although it might appear to be more appropriate to define the domain of dependence using timelike and null curves rather than just timelike curves, the only effect of such a change, as it turns out, is to eliminate certain boundary points from $D^+(S)$. It will later be seen to be more convenient to have such boundary points included in $D^+(S)$. Finally, the definition of the domain of dependence could be applied equally well when S is not achronal, but, once again, the consequence is to introduce additional complications without adding anything really new.

Note that the domain of dependence involves only the conformal structure of the space–time.[16] That is to say, if we replace the metric g_{ab} by $\Omega^2 g_{ab}$, where Ω is any strictly positive scalar field, then each $D^+(S)$ is unchanged.

Fortunately, it turns out that a handful of counter-examples is sufficient to dispose of many plausible-sounding conjectures concerning the domain of dependence. Thus one of the easiest ways to develop an intuitive feeling for the properties of $D^+(S)$, as well as to check and discover new properties, is through examples. We shall give seven examples here. (In all examples involving Minkowski space, we adopt the usual coordinates t, x, y, and z.)

Example 1: Let M be Minkowski space, and let S be the x axis, i.e., the curve given by $t = y = z = 0$. From any point not on S, we may always find a past-directed timelike curve, without a past endpoint, which fails to intersect S. Thus $D^+(S) = D^-(S) = D(S) = S$. This property would, of course, continue to hold were S any spacelike curve or 2-surface. We shall see in Sec. 3 that, at least locally, S must be a spacelike or null 3-surface if $D^+(S)$ is to contain any points other than those of S.

Example 2: Let M be Minkowski space, and let S be the hyperboloid of revolution given by

$$S = \{(t, x, y, z) \mid t^2 - 1 = x^2 + y^2 + z^2, t < 0\}.$$

It will be seen from Fig. 1 that

$$D^+(S) = \{(t, x, y, z) \mid t^2 \geq x^2 + y^2 + z^2 \geq t^2 - 1, t \leq 0\},$$

$$D^-(S) = \{(t, x, y, z) \mid t^2 - 1 \geq x^2 + y^2 + z^2, t < 0\}.$$

[13] If M were not time-orientable, then we could always find a covering space of M—representing exactly the same physical universe—which is. Many of the results can be applied directly to the non-time-orientable case, but the statements and proofs become somewhat more complicated.

[14] A point q is said to be an *endpoint* of a curve γ if γ enters and remains in every neighborhood of q. Note that q need not necessarily be a point of γ.

[15] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience Publishers, Inc., New York, 1965), Vol. II.

[16] In fact, the notion of the domain of dependence can be generalized to certain casual spaces. See R. Geroch, E. H. Kronheimer, and R. Penrose, "Ideal Points in Space–Time" (to be published).

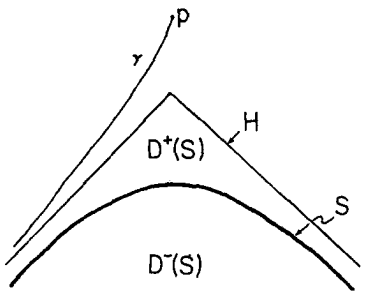


FIG. 1. The domains of dependence of a hyperboloid of revolution in Minkowski space (two spatial dimensions suppressed). Null geodesics in this figure are straight lines inclined at 45° to the vertical. The point p is not in $D^+(S)$ because the past-directed timelike curve γ from p has no past endpoint and does not intersect S.

[From the point $p = (1, 0, 0, 0)$, for example, there is a past-directed timelike curve, γ in the figure, which fails to intersect S, and so $p \notin D^+(S)$.] The only achronal set in M which contains S is S itself, and so we cannot further extend S while keeping it achronal. That is to say, S is a spacelike surface in Minkowski space such that $D(S) \neq M$, and such that S is not included in any achronal set whose domain of dependence *is* M. However, there are other surfaces in Minkowski space—for example, the plane S' given by $t = 0$—such that $D(S') = M$.

Consider, in the above example, the "future boundary" of $D^+(S)$, that is, the past light cone

$$H = \{(t, x, y, z) \mid t^2 = x^2 + y^2 + z^2, t \leq 0\}$$

of the origin. It is easily verified that H is achronal, and that $D^+(H) = H$, $D^-(H) = D(S)$. [Note that if, in our original Minkowski space, the point at the origin had been removed from the manifold, then we should have had $D^-(H) = H$.] More generally, let H be any achronal subset of $D^+(S)$. If p is a point of $D^+(H)$, then every past-directed timelike curve γ, without a past endpoint, from p intersects H. But $H \subset D^+(S)$, and so γ intersects S. Thus $D^+(H) \subset D^+(S)$, and, by a similar argument, $D^-(H) \subset D(S)$.

Example 3: Let M be Minkowski space with the single point at the origin, $O = (0, 0, 0, 0)$, removed, and let S be the closed, spacelike 3-disk

$$S = \{(t, x, y, z) \mid x^2 + y^2 + z^2 \leq 4, t = -1\}$$

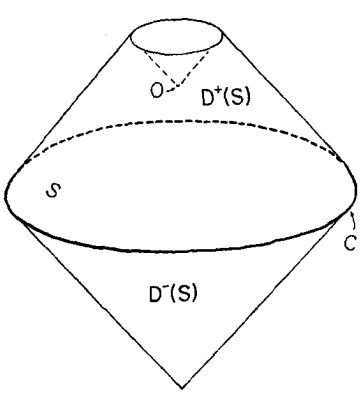


FIG. 2. The domains of dependence of a disk in Minkowski space with the single point O removed (one spatial dimension suppressed). Points to the future of O are not in $D^+(S)$. The set $D^+(S)$ [respectively, $D^-(S)$] includes its future (respectively, past) boundary whether or not the boundary circle C is included in S.

(see Fig. 2). Then, evidently,

$$D^-(S) = \{(t, x, y, z) \mid -3 \leq t \leq -1, x^2 + y^2 + z^2 \leq (t + 3)^2\}.$$

Were it not for the fact that the point O has been removed from Minkowski space to form M, $D^+(S)$ would be just the time-reverse of $D^-(S)$. However, in the example as given, no point p inside the "future light cone" of O can be in $D^+(S)$, for a past-directed timelike curve from p to O has no past endpoint and does not intersect S. (Of course, O cannot be an endpoint of a curve, for O is not a point of the manifold M.) Thus

$$D^+(S) = \{(t, x, y, z) \mid (t - 1)^2 \geq x^2 + y^2 + z^2 \geq t^2, t > 0, \text{ or } (t - 1)^2 \geq x^2 + y^2 + z^2, 0 \geq t \geq -1\}.$$

Suppose now we replace the set S by S', consisting of S with its bounding 2-sphere,

$$C = \{(t, x, y, z) \mid t = -1, x^2 + y^2 + z^2 = 4\},$$

removed. Then

$$D^{\pm}(S') = D^{\pm}(S) - C.$$

That is to say, the domain of dependence of S' differs from that of S only in that the boundary region C is absent. Note that the rest of the boundaries of D^+ and D^- are the same for S as for S'. This example illustrates a characteristic feature of the domain of dependence: D^+ (D^-) always contains its entire future (past) boundary. (These boundaries will later be defined more precisely.)

Example 4: Let M be the (extended) Schwarzschild solution, and let S be the 3-surface given, in the usual coordinates, by $t = 0$ (see Fig. 3). Then $D^+(S)$ consists of all points on S or to the future of S, and $D^-(S)$ consists of all points on S or to the past of S. Thus

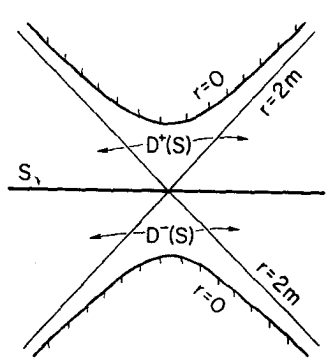


FIG. 3. The domains of dependence of a surface S in the Schwarzschild solution. Each point in the figure (except those on the "singularity," $r = 0$) represents a single "r = const" 2-sphere in the full 4-dimensional space–time. (Thus, for example, the topology of S is $S^2 \times R$.) A curve in the full space–time is represented in the figure by the set of points corresponding to the 2-spheres that curve intersects. Radial null geodesics are represented in the figure by straight lines at an angle of 45° to the vertical, while nonradial null geodesics are curved lines whose angles with the vertical are always less than 45°.

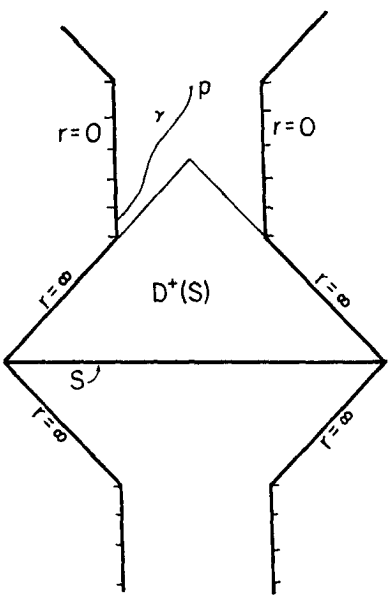


FIG. 4. The future domain of dependence of a surface S in the Reissner–Nördstrom solution. Each point in the figure (except at the "singularity" $r = 0$ and at "infinity," $r = \infty$) represents one of the $r = \text{const}$ 2-spheres in the four-dimensional space-time. A conformal transformation has been applied to bring the asymptotic regions (near $r = \infty$) into the figure. (The same transformation, applied to the Schwarzschild solution, produces a hexagonal-shaped figure.) Radial null geodesics are again straight lines at 45° to the vertical.

$D(S) = M$ and $D^+(S) \cap D^-(S) = S$. [It is always true that $D^+(S) \cap D^-(S) = S$.]

Example 5: Let M be the (extended) Reissner–Nördstrom solution, and let S be the 3-surface given, in the usual coordinates, by $t = 0$ (see Fig. 4). The set $D^+(S)$ is shown in the figure. Points which are sufficiently far up the "throat" between the two asymptotically flat regions, such as p in the figure, are not in $D^+(S)$, for there exists a past-directed timelike curve γ from p which "reaches the singularity at $r = 0$," and so fails to intersect S. In this example, $D(S) \neq M$. In fact, in Examples 3 and 5 there exists *no* achronal set whose domain of dependence is the whole space-time.

Examples 4 and 5 illustrate two ways in which the presence of singularities can influence the domain of dependence.

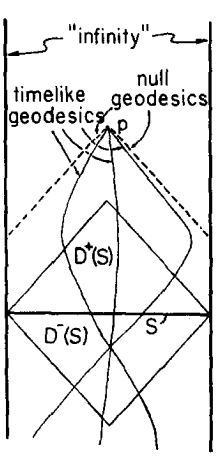


FIG. 5. The domains of dependence of a surface in anti-de Sitter space (two spatial dimensions suppressed). Here "infinity" is represented by a timelike surface, and so certain timelike curves (from p, for example) can reach infinity without first intersecting S. However, every past-directed timelike *geodesic* from p intersects S.

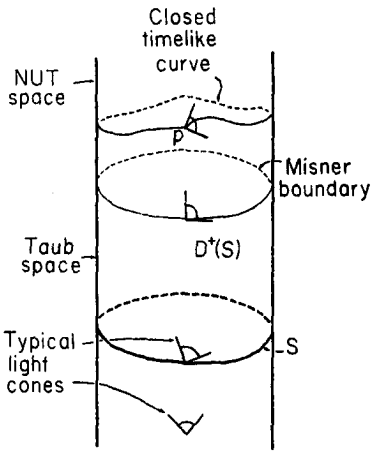


FIG. 6. The future domain of dependence of a surface in Taub–NUT space (two spatial dimensions suppressed: the circles in the figure would be 3-spheres in the full space-time). The metric is invarient under rotations of the cylinder about its axis. Through each point p to the future of the Misner boundary there passes a closed timelike curve, and so no such point may be in $D^+(S)$.

Example 6: Let M be anti-de Sitter space (a space-time, topologically Euclidean, of constant negative curvature), and let S be one of the homogeneous spacelike sections. The sets $D^+(S)$ and $D^-(S)$, as well as a number of typical timelike and null geodesics, are shown in Fig. 5. Note that in this example there exist points, such as p in the figure, such that every past-directed timelike geodesic from p reaches S, but such that $p \notin D^+(S)$. Intuitively, information can come in from "timelike infinity" and affect p without its being registered on S. This example illustrates the point that "timelike curve" cannot be replaced by "timelike geodesic" in the definition of the domain of dependence. (Note also that the null geodesics from p fail to intersect S.)

Example 7: Our final example illustrates the behavior of the domain of dependence in the presence of closed timelike curves. Let M be Taub–NUT space,[17] and let S be one of the compact, homogeneous, nonisotropic spacelike sections in the "Taub part" of the space (Fig. 6). Those points which lie on or below the Misner boundary between Taub and NUT space and on or above S are in $D^+(S)$. Through any point p which lies in the "NUT part" of the space (i.e., above the Misner boundary), there passes a closed timelike curve. The curve which begins at p and which continually describes this closed timelike curve does not strike S and does not have a past endpoint. Thus, no point of NUT space can lie in $D^+(S)$. The above property, evidently, holds in general for any achronal S in any space-time: If p lies on a closed timelike curve, then $p \notin D(S)$. (In the Gödel

[17] E. T. Newman, L. Tamburino, and T. Unti, J. Math. Phys. **4**, 915 (1963); C. W. Misner, J. Math. Phys. **4**, 924 (1963).

universe, to take a more extreme example, any two points may be joined by a timelike curve. Thus, there does not even exist a nonempty achronal set in the Gödel universe.)

In several of the examples above—e.g., in the Schwarzschild solution and in Minkowski space—we have found an achronal set S such that $D(S) = M$. When this is the case, S is said to be a *Cauchy surface.*[1,2] To say that S is a Cauchy surface for M is a statement about both S and M. Thus, in Example 2, we exhibited an achronal set S in Minkowski space such that $D(S) \neq M$, and such that S is not contained in any Cauchy surface. However, Minkowski space certainly has a Cauchy surface. On the other hand, neither the Reissner–Nördstrom solution, anti-de Sitter space, nor Taub–NUT space have any Cauchy surface.

Intuitively, that S be a Cauchy surface for M means that initial data on S determines the entire evolution of M, past and future.[7] In general, it is not possible to tell, by examining only a neighborhood of S, whether or not S will be a Cauchy surface. Thus, a space-time M (for example, the Reissner–Nördstrom solution) in which it appears, during the early stages of evolution (i.e., near S), that S will be a Cauchy surface may, at some much later time, develop so as to have no Cauchy surface[9]. It is for this reason that the assumption of the existence of a Cauchy surface is a rather unpleasant condition to impose, e.g., in theorems about singularities.

Another useful concept is that of the Cauchy horizon, or "future boundary" of $D^+(S)$. More precisely, the *future Cauchy horizon* $H^+(S)$ of S is defined[4,6] as the collection of all points p of $D^+(S)$ such that no future-directed timelike curve from p intersects $D^+(S)$ other than, of course, at p itself. The *past Cauchy horizon* $H^-(S)$ and *(total) Cauchy horizon* $H(S)$ are defined in analogy with $D^-(S)$ and $D(S)$. We see from Examples 1 and 3 that S may intersect, or even coincide with, its Cauchy horizons.

It follows immediately from this definition that $H^+(S)$ is achronal, and so sets such as $D^+(H^+(S))$, $H^+(H^+(S))$, and $D^-(H^+(S))$ are defined. From the discussion of Example 2, we see that the first two are always identical to $H^+(S)$, while the third must be at least as large as $H^+(S)$ and no larger than $D(S)$.

3. PROPERTIES OF THE DOMAIN OF DEPENDENCE

In this section we shall establish a number of standard results concerning the domain of dependence, Cauchy surfaces, and Cauchy horizons. As most of these results are known and many are fairly easy to prove, we shall not attempt to give the proofs in great detail. It is possible to construct quite a long list of properties of the domain of dependence, and for this reason it is perhaps more feasible to keep in mind a few constructions and counterexamples rather than to try to remember the various properties themselves. Therefore we present here a small selection of results, designed to illustrate certain of the observations of Sec. 2 and to exhibit some of the more important techniques for constructing proofs.

It is convenient to define the *past*[8,9] of a point p, $I^-(p)$, as the collection of points (p excluded) which may be reached from p by a past-directed timelike curve. The *future* of p, $I^+(p)$, is defined in a similar way by interchanging past and future. Note that $I^\pm(p)$ are always open sets. The past (future) of a set is defined as the union of the pasts (futures) of the points of that set. For example, in this notation the definition of the Cauchy horizon takes the form

$$H^+(S) = D^+(S) - I^-[D^+(S)].$$

We saw in Example 1 that, when S is a region of dimensionality less than three in Minkowski space, $D(S) = S$. This property holds in a general space-time:

Property 1: Let p be a point of $D^+(S)$. Then $I^-(p) \cap I^+[S]$ is contained in $D^+(S)$. Furthermore, $I^-(p) \cap S$ is a 3-surface.

Proof: Let $q \in I^-(p) \cap I^+[S]$, and let γ be a past-directed timelike curve from q without a past endpoint. We must show that γ intersects S. Since $q \in I^-(p)$, γ can be extended into the future to p. It follows from the fact that $p \in D^+(S)$ that the resulting curve γ' must intersect S. But $q \in I^+(S)$, and so, since S is achronal, γ' must intersect S at a point to the past of q. Thus, γ intersects S.

To prove the second part of the theorem, choose a nowhere-vanishing vector field t^a on M.[9,18] A trajectory of t^a is defined as a curve, maximally extended, which is everywhere tangent to t^a. There is precisely one trajectory through each point of M, and no trajectory has an endpoint. Since $I^-(p) \cap I^+[S] \subset D^+(S)$, any trajectory which enters the region $I^-(p) \cap I^+[S]$ must continue into the past to intersect S. Furthermore, any point of $I^-(p) \cap S$ must lie on a trajectory which enters the region $I^-(p) \cap I^+[S]$. Let s be a point of $S \cap I^-(p)$, and choose a small neighborhood O of s such that every trajectory which enters O intersects S in O. Since S is achronal, every such trajectory intersects S in the region $O - I^+(s) - I^-(s)$. Therefore,

[18] L. Markus, Ann. Math. 62, 411 (1955).

the neighborhood O of s intersects S in an achronal 3-surface.

In the examples of Sec. 2 (especially in Example 3), $D^+(S)$ appears to contain all points which could reasonably be considered as in its "future boundary." This property is necessary to ensure that $H^+(S)$, which is defined as a subset of $D^+(S)$, can be interpreted as this future boundary. Our precise statement asserts that, while $D^+(S)$ itself need not be closed, $D^+(S)$ differs from its closure only in points of the closure of S.

Property 2: $\overline{D^+(S)} = D^+(S) \cup \bar{S}$. In particular, if S is closed, so is $D^+(S)$.

Proof: Since $D^+(S) \supset S$, $\overline{D^+(S)} \supset D^+(S) \cup \bar{S}$. Let p_i be a sequence of points of $D^+(S)$, and let p be an accumulation point of this sequence. We must show that $p \in D^+(S)$ or $p \in \bar{S}$. Suppose that $p \notin \bar{S}$, so that there is some neighborhood O of p which does not intersect S. Let γ be any past-directed timelike curve from p, without a past endpoint. Since the p_i accumulate at p, there is some p_j and some timelike curve γ' from p_j into the past such that γ' joins γ in O and thereafter coincides with γ. But $p_j \in D^+(S)$, and so γ' must intersect S. The intersection cannot occur in O, and so must take place after γ' and γ coincide. Consequently, γ intersects S, and so $p \in D^+(S)$.

Note, however, that it is not true in general that $\overline{D^+(S)} = D^+(\bar{S})$. For example, let S be the surface $t = 0$, $x^2 + y^2 + z^2 > 0$ in Minkowski space.

There are, in fact, a number of relations between the operations of closure, D^+, and H^+. For example, either all or none of the following five sets are closed: S, $D^+(S)$, $D(S)$, $H^+(S)$, and $H(S)$. The proof is along the lines of Property 2, but incorporates the fact that, whenever S is achronal, so is $\bar{S}$, and points of $\bar{S} - S$ can never appear in $D(S)$. One normally considers the domain of dependence only in the case when S is closed. If we think of S as a surface on which to give initial data, then this restriction is not a very severe one, for initial data on S automatically determine, by continuity, the data on $\bar{S}$.

In order that a point p be in $D^+(S)$, we require that every past-directed timelike curve from p, without a past endpoint, intersect S. It might be asked, therefore, what is the effect on $D^+(S)$ if "timelike curve" is replaced by "timelike or null curve." We certainly expect that the domain of dependence, defined in this alternative way, would be no larger than $D^+(S)$. It turns out that neither is $D^+(S)$ made significantly smaller. The only effect of including null curves in the definition of the domain of dependence is to eliminate certain points of the Cauchy horizon from $D^+(S)$.

A continuous curve γ is said to be a *causal curve* if, given any two points p and q of γ and any neighborhood $O \subset M$ of the closed interval $[p, q]$ of γ, there is a smooth curve from p to q in O whose tangent vector is everywhere timelike or null.[19]

Property 3: Let $p \in D^+(S) - H^+(S)$. Then every past-directed causal curve from p, without past endpoint, intersects S.

Proof: Let $\gamma(t)$ be a past-directed causal curve from p, without past endpoint, where t has range $[0, \infty)$ and $\gamma(0) = p$. Let d denote the distance function between pairs of points of M with respect to some positive-definite metric on M. Since $p \in D^+(S) - H^+(S)$, we may choose a point $q \in D^+(S) \cap I^+(p)$ such that $d(p, q) < 1$. Since $\gamma(1) \in I^-(q)$, we may find a past-directed timelike curve $\gamma'(t)$, defined for $t \in [0, 1]$, from q such that

$$\gamma(t) \in I^-(\gamma'(t)), \quad d(\gamma(t), \gamma'(t)) < (1 + t)^{-1} \quad (1)$$

for all $t \in [0, 1]$. Since $\gamma(2) \in I^-(\gamma'(1))$, we may extend the curve γ' so that $\gamma'(t)$ is timelike and subject to (1) for $t \in [0, 2]$. Continuing in this way, we obtain a past-directed timelike curve $\gamma'(t)$, defined and subject to (1) for all $t \in [0, \infty)$. The curve γ' cannot have a past endpoint, for if there were such an endpoint it would [by (1)] also be an endpoint of γ. Therefore, since $q \in D^+(S)$, γ' intersects S. It now follows from (1) that some points of γ are to the past of S, and hence not in $D^+(S)$. Let s be the first point at which γ leaves $D^+(S)$. Since $\overline{D^+(S)} = D^+(S) \cup \bar{S}$, s is either in $\bar{S}$ or in $D^+(S)$. But s cannot lie to the future of S, for in this case $s \in I^-(q) \cap I^+[S] \subset$ interior $D^+(S)$. Therefore, s is in $\bar{S}$. Finally, s must be in S itself, for otherwise a timelike curve, drawn from q and passing through s, would violate the fact that $q \in D^+(S)$.

Example 3 shows that, when $p \in H^+(S)$, the question of whether or not causal curves from p reach S becomes more complicated. For some $p \in H^+(S)$ it may happen that every past-directed causal curve from p intersects S, while this property need not necessarily hold for all points of $H^+(S)$.

We have seen in Sec. 2 that, in many examples in which there is a Cauchy horizon, $H^+(S)$ is a null surface. This property does not hold in general: If, for example, S consists of those points of the plane

[19] In particular, a smooth curve is a causal curve if and only if its tangent vector is everywhere timelike or null. The above definition is rather awkward because it is necessary, for the purposes of Sec. 4, to admit causal curves which are not smooth.

$t = 0$ in Minkowski space for which x, y, and z are rational, then $S = D^+(S) = H^+(S)$, and $H^+(S)$ is neither a surface nor null. However, $H^+(S)$ always contains null geodesics through its points not included in S. This important property of the Cauchy horizon has been used in theorems about singularities.[4-6]

Property 4[4,7,9]*:* Let $p \in H^+(S) - S$. Then there exists a segment Γ of a past-directed null geodesic from p which remains entirely in $H^+(S)$, and which either has no past endpoint or else has a past endpoint contained in $\bar{S}$.

Proof: Choose a sequence of points p_i which lie to the future of p and approach p as a limit. Since $p \in H^+(S)$, none of the p_i lie in $D^+(S)$. Therefore, through each p_i we may draw a past-directed timelike curve γ_i, without past endpoint, such that the γ_i do not intersect S and, in particular, do not enter $D^+(S)$.

Define a partial limit of the γ_i to be a segment $\Gamma' \subset I^+[S]$ of a past-directed null geodesic from p such that, given any point $q \in \Gamma'$ and any neighborhoods U and V of Γ' and q, respectively, an infinite number of γ_i remain in U, at least until they reach V. We first show that each partial limit is contained in the Cauchy horizon. Let $q \in \Gamma'$. Then each past-directed timelike curve from q enters $I^-(p) \cap I^+[S] \subset D^+(S)$, and so intersects S. Furthermore, no point of $I^+(q)$ is in $D^+(S)$, for otherwise at least one γ_i would enter $D^+(S)$. Thus $q \in H^+(S)$.

Let Γ' be a partial limit, and suppose that Γ' has a past endpoint $q \in I^+[S]$. Choose a small compact neighborhood V of q. The sequence of points at which the γ_i first leave V must, by compactness, have an accumulation point r. Since the γ_i cannot enter $D^+(S)$, $r \notin I^-(q)$. Since the γ_i are timelike and pass arbitrarily close to q, $I^+(r) \supset I^+(q)$. Therefore, there is a segment $\Gamma'' \subset V$ of a null geodesic joining q to r. The tangent vectors to Γ' and Γ'' must agree at q, for otherwise Γ'', and hence the γ_i, would enter $I^-(p) \cap I^+[S] \subset D^+(S)$.[9] Therefore, $\Gamma' \cup \Gamma''$ is a partial limit of the γ_i. We have shown that any partial limit which has a past endpoint in $I^+[S]$ may be extended, as a partial limit, beyond that endpoint.

It follows that there exists a partial limit Γ which either has no past endpoint or else has a past endpoint $s \notin I^+[S]$. In the latter case, since $\Gamma \subset D^+(S)$, we must have $s \in \overline{D^+(S)} = D^+(S) \cup \bar{S}$, and so $s \in \bar{S}$.

Examples 2 and 3 show that there may be more than one null geodesic Γ from p. These same two examples also illustrate the fact that each of the three possibilities admitted by Property 4 (Γ has no past endpoint, Γ has a past endpoint in S, and Γ has a past endpoint in $\bar{S} - S$) can occur.

We next consider some results involving Cauchy surfaces. It is an immediate consequence of the definitions that S is always a Cauchy surface for the space-time given by the interior of $D(S)$, provided that set contains S. Thus all results about Cauchy surfaces apply equally well to the interior of the domain of dependence. We first show that, as expected, the condition that S be a Cauchy surface is precisely the condition that $H(S)$ be empty.

Property 5: Let S be nonempty and M connected. Then a necessary and sufficient condition that S be a Cauchy surface is that $H(S)$ be empty.

Proof: If S is a Cauchy surface, then $D(S) = M$, and so $H(S)$ is empty. Suppose, conversely, that $H(S)$ is empty. In particular, $H(S)$ is closed, and so, therefore, is $D(S)$. We show that $D(S)$ is also open. Let $p \in D(S)$. Since $p \notin H^+(S)$, there is a point r to the future of p in $D(S)$; since $p \notin H^-(S)$, there is a point s to the past of p in $D(S)$. We conclude from Property 1 that the open neighborhood $I^-(r) \cap I^+(s)$ of p is in $D(S)$. Since $D(S)$ is both open and closed, $D(S) = M$.

One often uses Properties 4 and 5 together to establish the existence of a Cauchy surface. By assuming some condition (such as an energy condition) which implies that null geodesics of the type required for Property 4 cannot exist, it follows that $H(S)$ is empty and, therefore, that S is a Cauchy surface.

The following result, using this type of argument, gives a characterization of a Cauchy surface in terms of the behavior of null geodesics.

Property 6[6]*:* A necessary and sufficient condition that an achronal set S be a Cauchy surface is that S be closed and every null geodesic in M intersect and then re-emerge from S.

Proof: If S is a Cauchy surface, then $D(S)$ is closed and so, therefore, is S. By Property 3, every null geodesic intersects S. In the proof of Property 3, it was shown in addition that every causal curve—and, in particular, every null geodesic—re-emerges from S.

Suppose, conversely, that S is closed and that every null geodesic intersects and then re-emerges from S. Since S is achronal, it follows that, for each point r of M, either $r \in S$, or every past-directed null geodesic

from r intersects S (i.e., $r \in I^+[S]$), or every future-directed null geodesic from r intersects S (i.e., $r \in I^-[S]$). Therefore, since S is closed, any timelike curve which passes from $I^+[S]$ into $I^-[S]$ must intersect S.

The theorem follows from Properties 4 and 5 if we can show that $H^+(S) \cap S$ is empty. Suppose, on the contrary, that there were some point $p \in H^+(S) \cap S$. Every past-directed null geodesic from p emerges from S, and therefore from $H^+(S)$. Consequently, since $H^+(S)$ is achronal, we may, without loss of generality, assume that no point on a past-directed null geodesic from p intersects $H^+(S)$ except, of course, at p itself. [If there were such a geodesic, replace p by the last point in $H^+(S)$ on that geodesic.] Choose a spacelike 2-surface C, topologically a 2-sphere, in a small neighborhood of p such that every past-directed null geodesic from p intersects C, and such that each point of C lies on such a null geodesic. (C is a cross section of the null cone of p.) Each point $q \in C$ is not in $H^+(S)$, and so there exists a q' which is to the future of q and which is in either $I^-[S]$ or $D^+(S)$. Letting C' denote the union of the q', it follows that $I^-[C']$ is open and contains C. Consequently, there exists a point $r \in I^+(p)$ such that every past-directed timelike curve from r enters $I^-[C']$. But $I^-[C'] \subset I^-[S] \cup D^+(S)$, and so every such curve must intersect S. Therefore $r \in D^+(S)$, which contradicts the fact that $p \in H^+(S)$.

The condition "S be closed" is actually necessary in Property 6, as the following example will show. Let S be the surface in Minkowski space given by

$$S = \{(t, x, y, z) \mid t = 1, x^2 + y^2 + z^2 \geq 1, \text{ or } 0 < t < 1, x^2 + y^2 + z^2 = t^2\}$$

(see Fig. 7). Then every null geodesic intersects and re-emerges from S; but, since S is not closed, it cannot be a Cauchy surface. If, in the above example, we remove from the space–time the set of points with $t \leq 0$, then S is closed. (The point at the origin, which was previously in the closure of S, is now no longer a point of the manifold.) However, S is still not a Cauchy surface: Now it does not satisfy the other condition of Property 6—namely, that every null geodesic re-emerge from S.

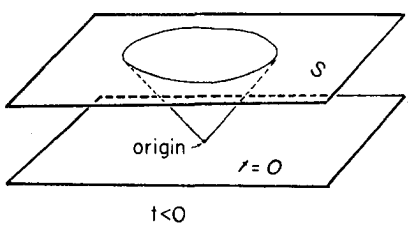


FIG. 7. An achronal set S in Minkowski space (one spatial dimension suppressed). The point marked "origin" is not in S, and so $D^+(S)$ cannot include any point to the future of this point. However, it will be observed that every null geodesic intersects and then re-emerges from S. If, in this space–time, we remove the points on and to the past of the plane $t = 0$, then S is closed. However, certain null geodesics (namely, those which would have passed through the origin) do not now re-emerge from S.

Geodesics are not a very useful way to characterize the domain of dependence. Despite Property 6, it is possible that every past-directed null geodesic from a point p intersects and then re-emerges from S, even though $p \notin D^+(S)$. (Let S be the surface $t = 0$, $x^2 + y^2 + z^2 \geq 1$ in Minkowski space.) In Example 6, every timelike geodesic intersects and re-emerges from S, but S is not a Cauchy surface. If, however, every timelike and every null geodesic intersects S, then S is necessarily a Cauchy surface. In fact, if the point p is such that every past-directed timelike and null geodesic from p intersects S, then $p \in D^+(S)$. The converse is false, e.g., Examples 3 and 6.

Our next property of the domain of dependence asserts that nothing very interesting from a topological viewpoint can take place in $D^+(S)$. Suppose, for example, that we have two universes which, at some time, join together to form a single universe. We might wish to give initial data for this system by specifying a Cauchy surface S consisting of two disjoint pieces, one in each universe. The idea would then be that the initial data on S would predict the separate evolution of the two universes, and also their ultimate confluence. Property 7 asserts, among other things, that such a specification is impossible.

Property 7[6,9]*:* Let S be a Cauchy surface for the space-time M. Then M is topologically $S \times R$. In particular, if M is connected, so is S.

Proof: Choose a nowhere-vanishing timelike vector field t^a on M. Exactly one trajectory of t^a passes through each point of M. Since S is a Cauchy surface, each trajectory intersects S at just one point. We may therefore define a scalar field φ, a "time parameter," by imposing the two conditions $t^a \nabla_a \varphi = 1$ and $\varphi = 0$ on S. It is always possible to rescale t^a so that φ assumes arbitrarily large and arbitrarily small values along each trajectory.[20] Now consider the one-parameter family of surfaces S_φ given by $\varphi = \text{const}$. Exactly one of these 3-surfaces passes through each point of M. Furthermore, any two of these surfaces, S_φ and $S_{\varphi'}$, are homeomorphic: The mapping from S_φ onto $S_{\varphi'}$ is obtained by following along the trajectories of t^a. Hence, $M = S \times R$.

[20] This rescaling may be accomplished as follows: The total length of a trajectory is a lower semicontinuous function on S. Therefore, this function is the limit of some monotonically increasing sequence φ_i ($i = 1, 2, \cdots$) of functions on S. Rescale t^a so that $\varphi = i$ on the surface whose proper distance from S along the trajectories is φ_i.

Examples 5, 6, and 7 show that the converse of Property 7 is false: M may be a topological product without its having a Cauchy surface. In Sec. 5 we shall obtain a stronger version of Property 7: The one-parameter family of 3-surfaces can be so chosen that each is a Cauchy surface.

Our last two properties of the domain of dependence are proved most easily using a topology on sets of curves to be introduced in Sec. 4. The proofs will be given at the end of that section.

Property 8[2]*:* Let $p \in D^+(S) - H^+(S)$. Then the closure of the set $I^-(p) \cap I^+[S]$ is compact.

Note in addition that when $p \in D^+(S) - H^+(S)$, then $\overline{I^-(p) \cap S}$, a closed subset of $\overline{I^-(p) \cap I^+[S]}$, is also compact. The converse of Property 8 is false. In fact, it is possible for $\overline{I^-(p) \cap I^+[S]}$ to be compact for every $p \in M$ even though S is not a Cauchy surface. (If, however, we assume in addition that M has no closed causal curves, then the converse, as stated above, is true.)

Property 9, which has been used in theorems about singularities, will be required for Sec. 6.

Property 9[2,9]*:* Let $p \in D^+(S) - H^+(S)$. Then there is a timelike geodesic from p to S whose length is at least as large as that of every other smooth causal curve from p to S.

4. GLOBAL HYPERBOLICITY

The notion of global hyperbolicity was introduced by Leray[10] in order to deal with questions of existence and uniqueness of solutions of hyperbolic differential equations on a manifold. It has been used, in particular, in the study of the Cauchy problem in general relativity.[12]

In order to define global hyperbolicity, it is necessary to introduce a topology on certain collections of curves in the space–time M. Let p and q be two points of M, and let $C(p, q)$ denote the collection of all causal curves from p to q. The set of all curves in $C(p, q)$ which lie in O, and O ranges over all open sets in M, defines a basis for a topology on $C(p, q)$. Thus, for example, a sequence γ_i of curves approaches a curve γ if the γ_i are eventually contained in any open set in M containing γ, which corresponds to the intuitive notion of curves "approaching" another curve. It should be pointed out, however, that our topology is not very useful in the presence of closed causal curves. [In particular, $C(p, q)$ is not necessarily Hausdorff in this case.] Fortunately, it is possible to arrange matters so that the topology need only be used in the absence of closed causal curves.[21]

If M has no closed causal curves, then the topological space $C(p, q)$ has a countable basis.[22]

A space–time M is said to be *globally hyperbolic*[10,11] if M has no closed causal curves and if $C(p, q)$ is compact for every p and q. We shall not derive any consequences of global hyperbolicity, since, in the following section, we prove that this condition is equivalent to the existence of a Cauchy surface.

For problems involving Cauchy surfaces, it is convenient to work with a slightly different set of curves. Let p be a point and S a subset of M, and let $C(p, S)$ denote the collection of all causal curves with future endpoint p and past endpoint in S. The set of curves in $C(p, S)$ which lie in O and have past endpoint in O', as O and O' range over all open sets in M, is a basis for a topology on $C(p, S)$. When S is a single point q, the topologies on $C(p, q)$ and $C(p, S)$ coincide.

The interest in $C(p, S)$ stems primarily from the following lemma:

Lemma 10: If S is a Cauchy surface, then $C(p, S)$ is compact.

Proof: Let γ_i be a sequence of curves in $C(p, S)$. Define a partial limit of the γ_i to be a past-directed causal curve Γ from p such that, given neighborhoods U and V of Γ and a point $q \in \Gamma$, respectively, an infinite number of γ_i remain in U at least until they reach V. Suppose a partial limit Γ has a past endpoint q such that, for some neighborhoods U and V of Γ and q, with V compact, only a finite number of the γ_i which remain in U until they reach V have a past endpoint in V. The points at which these γ_i leave V must, by compactness, have an accumulation point $q' \in V$. In this case we may extend Γ to a partial limit with past endpoint q' (see the proof of Property 4). Thus, every partial limit may be extended to a partial limit which either has no past endpoint or else is an accumulation curve of the γ_i. But S is a Cauchy surface, and so any causal curve without past endpoint

[21] Leray's topology is "reasonable" even in space–times with closed causal curves because it refers directly to the mappings which define a curve. Our topology, by referring only to the images of such mappings, is considerably easier to state and visualize, but suffers from the more limited range of applicability described above. In the absence of closed causal curves, the mapping is determined (up to monotonic changes in the parameter) by the image; then Leray's topology and our topology coincide.

[22] This follows from the fact that $C(p, q)$ is a separable metric space. To construct a countable dense set in $C(p, q)$, choose a countable dense set p_i, including the points p and q, in M [see R. Geroch, J. Math. Phys. **9**, 1739 (1968)]. Consider curves $\gamma \in C(p, q)$ which consist of geodesic segments, each of which lies in a normal neighborhood and joins two p_i. A metric on $C(p, q)$ is obtained using the "maximal distance" between curves with respect to a positive-definite metric on M.

must intersect and re-emerge from S. Such a curve cannot be a partial limit of the γ_i. Thus, every sequence in $C(p, S)$ has an accumulation curve, and so $C(p, S)$ is compact.

Lemma 10 has a number of useful consequences. In particular, it provides us with easy proofs of Properties 8 and 9 of Sec. 3.

To prove Property 8, consider a sequence of points $p_i \in I^-(p) \cap I^+[S]$. For each p_i draw a timelike curve from p to p_i to S. This sequence has a limit curve Γ with future endpoint p and past endpoint in S. Therefore, Γ is a compact set. An infinite number of p_i are contained in any neighborhood of Γ, and so the p_i have an accumulation point. [Note that we have actually proved slightly more than Property 8—namely, that any point of $\overline{I^-(p) \cap I^+[S]}$ lies on a curve in $C(p, S)$.]

To prove Property 9, consider the distance function[23] on causal curves from p to S. This function is upper semicontinuous, and so, since $C(p, S)$ is compact, it assumes its maximal value. The curve on which the distance function is a maximum is the required geodesic. (Similarly, any two points which may be connected by a timelike curve in a globally hyperbolic space-time may also be connected by a timelike geodesic.)

5. GLOBAL HYPERBOLICITY IS EQUIVALENT TO THE EXISTENCE OF A CAUCHY SURFACE

In this section we shall prove our main theorem, the equivalence of global hyperbolicity and the existence of a Cauchy surface. The proof is based on the notion of the past volume and future volume of a point. Select an arbitrary positive volume element dV on M such that the total volume of M is equal to unity[24]:

$$\int_M dV = 1.$$

The *past volume* of a point p, $V^-(p)$, is defined as the volume of the past of p:

$$V^-(p) = \int_{I^-(p)} dV > 0,$$

and similarly for the future. As these volume functions appear to be useful tools for treating certain global aspects of space-times,[25] we first derive some of their elementary properties.

An important property of the volume functions is that, in the absence of closed timelike curves, V^- (respectively V^+) is strictly increasing along each future-directed (respectively, past-directed) timelike curve. [Proof: If there are no closed timelike curves, and if $p \in I^-(q)$, then $I^-(p)$ is a proper subset of $I^-(q)$. Therefore, $V^-(p) < V^-(q)$.] In fact, that V^- be increasing along timelike curves is necessary and sufficient for the absence of closed timelike curves.[26] [Proof: If p and q lie on a closed timelike curve, then $I^-(p) = I^-(q)$, and so $V^-(p) = V^-(q)$.]

In general, V^+ and V^- will not be continuous. For example, let M be Minkowski space with the spacelike half-plane $t = 0$, $x \leq 0$ removed. Then V^- is discontinuous on the null half-plane $0 < t = -x$, and V^+ is discontinuous on the null half-plane $0 > t = x$. (There are similar examples in which only one of the volume functions is discontinuous.)

However, when M is globally hyperbolic, then both V^+ and V^- are continuous everywhere. To prove this result, let p_i be a sequence of points of M, and let p be the limit point of this sequence. We must show that $\lim V^-(p_i) = V^-(p)$. Let ΔV be a subset of M such that $\Delta V \subset I^-(q)$ for some $q \in I^-(p)$. Then, since the p_i approach p, there is some j such that $q \in I^-(p_i)$ whenever $i > j$. That is, the $I^-(p_i)$ eventually contain ΔV. Since $I^-(p)$ can be written as a union of such ΔV's, we conclude that $\underline{\lim}\, V^-(p_i) \geq V(p)$.

We use global hyperbolicity to obtain an upper limit for the $V^-(p_i)$. Let ΔV be a subset of M such that $\Delta V \subset I^+(q)$ for some q such that q cannot be joined to p by a future-directed causal curve. We show that ΔV eventually lies outside of the $I^-(p_i)$. Suppose, on the contrary, that there were an infinite number of p_i such that $\Delta V \subset I^-(p_i)$. Then $q \in I^-(p_i)$ for an infinite number of p_i. Choose any point q' to the future of p. Since the p_i approach p, the p_i are eventually in $I^-(q')$. For each i such that $p_i \in I^-(q') \cap I^+(q)$, draw a timelike curve γ_i from q' to p_i to q. Since M is globally hyperbolic, this sequence of timelike curves from q' to q has a limit curve Γ. Since the p_i approach p, this limit curve passes through p, thus contradicting our assumption that q and p could not be joined by a causal curve. Therefore, ΔV is eventually outside the $I^-(p_i)$. But the interior of $M - I^-(p)$ may be written

[23] A. Avez, Grenoble Universite Institut Fourier **13**, 105 (1963).

[24] Such a volume element always exists. Since M has a metric, it is paracompact [see R. Geroch, J. Math. Phys. **9**, 1739 (1968)] and, therefore, triangulable [see J. H. C. Whitehead, Ann. Math. **41**, 809 (1940)]. Choose a volume element on each of the 4-simplices of M such that the volume of the first simplex is $\frac{1}{2}$, that of the second $\frac{1}{4}$, etc. (No assumption of orientability is necessary.)

[25] S.W. Hawking, Proc. Roy. Soc. (London) **308A**, 433 (1969).

[26] The situation is somewhat more complicated in the case of causal curves. If V^- is strictly increasing along causal curves, then there cannot be closed causal curves. However, it is necessary to impose a condition stronger than the absence of closed causal curves—for example, the strong causality condition (see Ref. 4)—in order to insure that V^- be strictly increasing along causal curves. That is to say, strong causality implies V^- increasing along causal curves which, in turn, implies no closed causal curves. For neither of these implications is the converse true.

as the union of such ΔV's, and so $\overline{\lim} V^-(p_i) \leq V^-(p)$. We conclude that V^- is continuous on M.

The converse of this result is false: The continuity of both V^- and V^+ is not sufficient for global hyperbolicity (Examples 5 and 6).

Note that the continuity of $V^\pm$ at each point is independent of the choice of volume element. Thus we have an invariant separation of the points of any space–time M into four classes: points at which both V^+ and V^-, V^+ but not V^-, V^- but not V^+, or neither are continuous. We have proven that the collection of space–times for which all points are in the first class includes the globally hyperbolic space–times. It might be interesting to study this collection in more detail.

We now prove our theorem.

Theorem 11: A space–time M is globally hyperbolic if and only if it has a Cauchy surface.

Proof: Suppose first that M has a Cauchy surface S, so that, in particular, M has no closed causal curves. Let p and q be two points of M, and let γ_i be a sequence of past-directed causal curves from p to q.

We consider two cases. If both p and q are in $D^+(S)$, then each curve γ_i may be extended to a curve which intersects S. By Lemma 10, the resulting sequence of causal curves has an accumulation curve Γ. But since each of the γ_i contains q, so must Γ. The portion of Γ between p and q is an accumulation curve of the γ_i. If, on the other hand, $p \in D^+(S)$ and $q \in I^-[S]$, then each of the γ_i must intersect S. Thus, by Lemma 10, the curves $\gamma_i \cap D^+(S)$ have an accumulation curve, say Γ, with past endpoint $s \in S$. Choose a subsequence of the γ_i whose intersection points with S approach s. Then the curves $\gamma_i \cap D^-(S)$, for γ_i in this subsequence, have an accumulation curve Γ'. The curve Γ (respectively, Γ') has a past (respectively, future) endpoint s, and so $\Gamma \cup \Gamma'$ is an accumulation curve of the γ_i. We have shown that, for every p and q, $C(p, q)$ is compact, and therefore that M is globally hyperbolic.

To prove the converse, suppose that M is globally hyperbolic, whence V^+ and V^- are continuous. The function $\lambda = V^-/V^+$ is also continuous and strictly decreasing along each past-directed timelike curve. Let S be the set of points at which $\lambda = 1$. Since λ is strictly decreasing along timelike curves, S is achronal. We shall show that S is also a Cauchy surface. Let p be a point of M with $\lambda(p) > 1$, and let γ be any past-directed timelike curve, without a past endpoint, from p. Since λ is continuous, γ must intersect S, provided we can show that λ assumes arbitrarily small positive values on γ. Let q be a point of M, and ΔV a subset of M in $I^+(q)$. We first show that ΔV cannot be in $I^-(r)$ for every point r of γ. Suppose, on the contrary, that $q \in \bigcap_{r\in\gamma} I^-(r)$. Choose a sequence of points p_i on γ such that $p_{i+1} \in I^-(p_i)$ and such that every point of γ lies to the past of at least one p_i. For each i, draw a timelike curve which begins at p, coincides with γ to p_i, and then continues to q. Since M is globally hyperbolic, this sequence has a limit curve, Γ. The limit curve evidently contains γ. But this is impossible, for γ, if it were contained in a (compact) casual curve from p to q, would then have a past endpoint. Therefore, there must be some point r of γ such that $\Delta V \not\subset I^-(r)$. Since M may be covered by such ΔV's, we conclude that $V^-(r)$ approaches zero as r continues into the past on γ, and, therefore, that γ intersects S. We have shown that every past-directed timelike curve from p intersects S, i.e., that $p \in D^+(S)$. Similarly, if $\lambda(p) < 1$, then $p \in D^-(S)$. Therefore, S is a Cauchy surface.

The proof of Theorem 11 also provides us with a somewhat stronger version of Property 7 (Sec. 3). Consider the one-parameter family of surfaces S_λ in M, given by $\lambda = \text{const}$. Then the S_λ are achronal, and exactly one S_λ passes through each point of M. Furthermore, by the argument used in the theorem, each S_λ is a Cauchy surface. Thus, if a space–time M has a single Cauchy surface, then M may be covered by a one-parameter family of nonintersecting Cauchy surfaces. The argument of Property 7 shows that these Cauchy surfaces all have the same topology.

6. GLOBAL HYPERBOLICITY IS STABLE

In this section we shall prove that global hyperbolicity is a stable property of space–times, i.e., that arbitrary, sufficiently small variations in the metric will not destroy global hyperbolicity. This result is useful both conceptually and as a lemma for simplifying the proofs of various other properties of space–times.[7]

In order to make clear what "sufficiently small" means, we must introduce a topology on the collection $\mathfrak{G}$ of all metrics on M of signature $(+, -, -, -)$. Let $h_{ab}, h'_{ab} \in \mathfrak{G}$. We write $h_{ab} < h'_{ab}$ if every vector which is timelike or null with respect to the metric h_{ab} is strictly timelike with respect to the metric h'_{ab}. Intuitively, $h_{ab} < h'_{ab}$ means that the light cones of h'_{ab} are more "opened out" than those of h_{ab}. The set of $g_{ab} \in \mathfrak{G}$ such that $h_{ab} < g_{ab} < h'_{ab}$, as h_{ab} and h'_{ab} range over $\mathfrak{G}$, forms a basis for a topology on $\mathfrak{G}$. This topology does not distinguish between conformally related metrics: i.e., two metrics which differ only by a conformal factor are always contained in the same open

sets in $\mathfrak{G}$.[27] This feature is perhaps an advantage in discussing properties of space-times—such as global hyperbolicity—which are conformally invariant. Two metrics which are not conformally equivalent may always be separated by disjoint open neighborhoods.

The characteristic feature of our topology on $\mathfrak{G}$ is that, compared with alternative topologies,[5] it contains a great many open sets. That is to say, it is very difficult in this topology to have a sequence of metrics approach a given metric as a limit. For example, the sequence

$$g_{ab}(n) = \text{diag}\,(\mu_n, -1, -1, -1),$$

where

$$\mu_n = [1 + n^{-2}(x^2 + y^2 + z^2 + t^2)]^{-1},$$
$$n = 1, 2, \cdots,$$

does not approach Minkowski space as a limit. However, the topology is perhaps the most appropriate one for discussing questions of stability.[28]

Using our topology, we define stability as follows: A property of space-times is said to be *stable* if, given any 4-manifold M, the collection of metrics on M which have the given property forms an open set in $\mathfrak{G}$.

Theorem 12: Global hyperbolicity is stable.

Proof: Let M, g_{ab} be a globally hyperbolic space-time. Choose a normalized volume element on M, and let S denote the Cauchy surface given by $V^+ = V^-$. For each value of the parameter τ, set

$$S_\tau = \{p \in M \mid \tau = V^+(p) \leq V^-(p)\}$$

(see Fig. 8). Each S_τ is an achronal set in $D^+(S)$, and exactly one of the S_τ passes through each point of $D^+(S)$.

We first show that, for each value of τ, the closed region $C_\tau = \overline{I^-[S_\tau] \cap I^+[S]}$ between S_τ and S is compact. Let p_i be a sequence of points in C_τ. If, for every $q \in M$, there were only a finite number of $p_i \in I^-(q)$, then we should have $\lim V^+(p_i) = 0$, which contradicts the fact that $V^+(p_i) \geq \tau$ for all $p_i \in C_\tau$. Therefore, there exists a q such that $I^-(q)$ contains an infinite number of p_i. But $\overline{I^-(q) \cap I^+[S]}$ is compact, and so, since it contains an infinite number of p_i, the p_i have an accumulation point. Consequently, C_τ and also S_τ, a closed subset of C_τ, are compact sets.

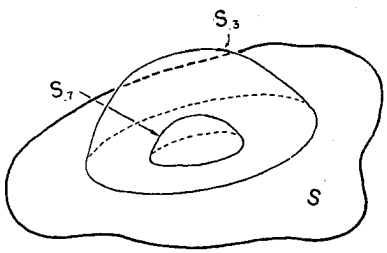


FIG. 8. The Cauchy surface S with the surfaces S_τ in $D^+(S)$. We see that the S_τ form the future boundaries of a nested family of (compact) "bubbles" residing on S. (It may happen that some S_τ are disconnected. Then, as τ decreases, new independent bubbles spring up in other regions of S. Later, as τ decreases further, these bubbles coalesce.)

We say that the family of surfaces S, S_τ are strictly spacelike, with respect to a metric h_{ab}, at a point p if there exists an open neighborhood O of p and a metric $h'_{ab} > h_{ab}$, defined in O, such that no two points of a single surface S_τ or S may be joined by a curve, timelike with respect to h'_{ab} and contained in O. The construction described at the end of the proof shows that S, S_τ may be replaced by surfaces which are strictly spacelike with respect to g_{ab}. (In fact, examples suggest that these surfaces are already strictly spacelike, but no proof is known.) With S, S_τ strictly spacelike with respect to g_{ab}, then, since M is paracompact,[29] there exists a metric $h_{ab} > g_{ab}$ such that S, S_τ are strictly spacelike with respect to h_{ab}.

Let $p \in S_{\tau_0}$, and let γ be a curve from p, past-directed and timelike with respect to h_{ab}, and without a past endpoint. We shall show that γ must intersect S. Since the S_τ are strictly spacelike with respect to h_{ab}, the parameter τ is strictly increasing into the past along γ. Consequently, since C_{τ_0} is compact and since γ has no past endpoint, γ must re-emerge from C_{τ_0}. The boundary of C_{τ_0} consists of S_{τ_0} and certain points of S. The point of emergence of γ cannot be in S_{τ_0}, for τ is strictly increasing along γ. Therefore, γ intersects S.

Using a similar construction for $D^-(S)$, we obtain a metric $h_{ab} > g_{ab}$, defined on M, with respect to which S is a Cauchy surface. Therefore, S is a Cauchy surface with respect to any metric $< h_{ab}$, an open set in $\mathfrak{G}$ which includes g_{ab}. We conclude that global hyperbolicity is stable.

Finally, we outline the construction, required for Theorem 12, which replaces achronal surfaces by strictly spacelike surfaces. Consider two Cauchy surfaces S and S' in a space-time M, and suppose that $S' \subset D^+(S)$. From Property 9 we see that, for each point p, there is a timelike geodesic from p to S whose length is at least as large as that of every other geodesic from p to S. Let $d(p, S)$ denote this length, whence

[27] It is not difficult to construct a similar, but somewhat more complicated, topology which does distinguish between conformally related metrics. The idea of applying this alternative topology to problems in relativity theory was suggested by S. W. Hawking. For a discussion of such topologies, see Ref. 5.

[28] There is another topology on $\mathfrak{G}$ which, because it contains very few open sets, is more suitable for questions involving limits [see R. Geroch, Commun. Math. Phys. 13, 180 (1969).]. It also differs from the topology above in that the manifold need not be given *a priori*.

[29] R. Geroch, J. Math. Phys. 9, 1739 (1968).

$d(p, S)$ is continuous in p. Define the Cauchy surface

$$\sum(S, S') = \{p \in M \mid d(p, S) = d(p, S'), \quad p \in D^+(S) \cap D^-(S')\}.$$

Let p be a point of $\sum$ with $d(p, S) = d_0 > 0$, γ a timelike geodesic from p to S of length d_0, and q a point on γ such that $d(q, S) = d_0 - \epsilon$. Then, for ϵ sufficiently small, q lies in a normal neighborhood O of p. The set A consists of points in O whose geodesic distance from q is ϵ is a regular, spacelike 3-surface through p. For each point $a \in A$, $d(a, S) \geq d_0$. It follows that $\sum$ is strictly spacelike at all points not included in S or S'.

It is necessary for Theorem 12 to replace S, S_τ by strictly spacelike surfaces. We replace S by the Cauchy surface $\tilde{S} = \sum(S, S')$, where S and S' are the Cauchy surfaces $V^+ = V^-$ and $V^+ = \frac{1}{2}V^-$, respectively. For the S_τ, we first note that for each τ, $W_\tau = (S_\tau \cup \tilde{S}) - I^-[S_\tau \cup \tilde{S}]$ is a Cauchy surface (see Fig. 8). Replace each S_τ by $\overline{\sum(W_{\tau 2}, W_\tau)} - \tilde{S}$.

ACKNOWLEDGMENTS

I wish to thank S. W. Hawking and especially R. Penrose for numerous discussions on the subject of the domain of dependence. The statements of Theorems 11 and 12 arose originally from remarks of R. Penrose and Y. Choquet-Bruhat, respectively.

This work was carried out under an Air Force Office of Scientific Research postdoctoral fellowship.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Mathematical Description of a System Consisting of Identical Quantum-Mechanical Particles

H. KUMMER
Queen's University, Kingston, Ontario, Canada

(Received 19 February 1968)

In this paper a rigorous description of a system consisting of identical particles is given for which the particle number is a superselection rule. A state of such a system is described by a sequence of density operators $\mathbf{D} = (D^0, D^1, \cdots, D^n, \cdots)$, where D^n acts in n-particle space and the asymptotic behavior is determined by the requirement

$$\sum_{n=0}^{\infty} k^n |D^n| < \infty \quad \text{for all} \quad k = 1, 2, \cdots,$$

where $|\ |$ denotes the trace norm. The (bounded) observables in turn are described by sequences of bounded self-adjoint operators: $\mathbf{B} = (B^0, B^1, \cdots, B^n, \cdots)$ such that

$$\sup_n k^{-n} \|B^n\| < \infty \quad \text{for some} \quad k = 1, 2, \cdots.$$

The expectation value of the observable $\mathbf{B}$ in the state $\mathbf{D}$ can be expressed as

$$\langle B \rangle_D = (\Gamma(\mathbf{B}) \mid \mathbf{D}) = (\mathbf{B} \mid L(\mathbf{D})),$$

where

$$(\mathbf{B} \mid \mathbf{D}) = \sum_{n=0}^{\infty} \mathrm{Tr}\,(B^n D^n);$$

Γ denotes the expansion operator, whereas L stands for its adjoint, the contraction operator. The analog to the n-representability problem, i.e., the so-called representability problem, is put forward and its solution is connected with the solutions of the n-representability problem for different n. Finally, a possible mathematical foundation of the BCS theory is given.

1. INTRODUCTION

In a recent paper (Kummer,[1] subsequently denoted by NRP) the n-representability problem for density operators was considered. The problem can be formulated in the following manner: Let H^1 be the Hilbert space of one particle, H^n the n-fold tensor product of H^1, and $H^{n\wedge}$ the physically relevant subspace (antisymmetric for fermions, symmetric for bosons) of H^n. Given a p-particle density operator D^p, i.e., a positive linear operator in H^p of trace 1, whose range is contained in $H^{p\wedge}$, under what circumstances does there exist an n-particle density operator D^n, such that D^p is the (n, p) contraction of D^n?

[1] H. Kummer, J. Math. Phys. 8, 2063 (1967).

$d(p, S)$ is continuous in p. Define the Cauchy surface

$$\sum(S, S') = \{p \in M \mid d(p, S) = d(p, S'), \quad p \in D^+(S) \cap D^-(S')\}.$$

Let p be a point of $\sum$ with $d(p, S) = d_0 > 0$, γ a timelike geodesic from p to S of length d_0, and q a point on γ such that $d(q, S) = d_0 - \epsilon$. Then, for ϵ sufficiently small, q lies in a normal neighborhood O of p. The set A consists of points in O whose geodesic distance from q is ϵ is a regular, spacelike 3-surface through p. For each point $a \in A$, $d(a, S) \geq d_0$. It follows that $\sum$ is strictly spacelike at all points not included in S or S'.

It is necessary for Theorem 12 to replace S, S_τ by strictly spacelike surfaces. We replace S by the Cauchy surface $\tilde{S} = \sum(S, S')$, where S and S' are the Cauchy surfaces $V^+ = V^-$ and $V^+ = \frac{1}{2}V^-$, respectively. For the S_τ, we first note that for each τ, $W_\tau = (S_\tau \cup \tilde{S}) - I^-[S_\tau \cup \tilde{S}]$ is a Cauchy surface (see Fig. 8). Replace each S_τ by $\overline{\sum(W_{\tau 2}, W_\tau)} - \tilde{S}$.

ACKNOWLEDGMENTS

I wish to thank S. W. Hawking and especially R. Penrose for numerous discussions on the subject of the domain of dependence. The statements of Theorems 11 and 12 arose originally from remarks of R. Penrose and Y. Choquet-Bruhat, respectively.

This work was carried out under an Air Force Office of Scientific Research postdoctoral fellowship.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Mathematical Description of a System Consisting of Identical Quantum-Mechanical Particles

H. KUMMER
Queen's University, Kingston, Ontario, Canada

(Received 19 February 1968)

In this paper a rigorous description of a system consisting of identical particles is given for which the particle number is a superselection rule. A state of such a system is described by a sequence of density operators $\mathbf{D} = (D^0, D^1, \cdots, D^n, \cdots)$, where D^n acts in n-particle space and the asymptotic behavior is determined by the requirement

$$\sum_{n=0}^{\infty} k^n |D^n| < \infty \quad \text{for all} \quad k = 1, 2, \cdots,$$

where $|\ |$ denotes the trace norm. The (bounded) observables in turn are described by sequences of bounded self-adjoint operators: $\mathbf{B} = (B^0, B^1, \cdots, B^n, \cdots)$ such that

$$\sup_n k^{-n} \|B^n\| < \infty \quad \text{for some} \quad k = 1, 2, \cdots.$$

The expectation value of the observable $\mathbf{B}$ in the state $\mathbf{D}$ can be expressed as

$$\langle B \rangle_D = (\Gamma(\mathbf{B}) \mid \mathbf{D}) = (\mathbf{B} \mid L(\mathbf{D})),$$

where

$$(\mathbf{B} \mid \mathbf{D}) = \sum_{n=0}^{\infty} \mathrm{Tr}\,(B^n D^n);$$

Γ denotes the expansion operator, whereas L stands for its adjoint, the contraction operator. The analog to the n-representability problem, i.e., the so-called representability problem, is put forward and its solution is connected with the solutions of the n-representability problem for different n. Finally, a possible mathematical foundation of the BCS theory is given.

1. INTRODUCTION

In a recent paper (Kummer,[1] subsequently denoted by NRP) the n-representability problem for density operators was considered. The problem can be formulated in the following manner: Let H^1 be the Hilbert space of one particle, H^n the n-fold tensor product of H^1, and $H^{n\wedge}$ the physically relevant subspace (antisymmetric for fermions, symmetric for bosons) of H^n. Given a p-particle density operator D^p, i.e., a positive linear operator in H^p of trace 1, whose range is contained in $H^{p\wedge}$, under what circumstances does there exist an n-particle density operator D^n, such that D^p is the (n, p) contraction of D^n?

[1] H. Kummer, J. Math. Phys. 8, 2063 (1967).

The (n, p) contraction L_n^p maps an n-particle density operator D^n onto the p-particle density operator D^p defined by the requirement that its matrix representation with respect to the complete orthonormal set (CONS) $\{e_i\}$ in H^p is given by

$$(e_k \mid D^p e_i) = \mathrm{Tr}\,((P_k^i \otimes I^{n-p})D^n), \tag{1.1}$$

where P_k^i denotes the operator defined by

$$P_k^i u = (e_k \mid u)e_i, \quad u \in H^p, \tag{1.2}$$

and I^{n-p} stands for the identity in H^{n-p}.

If D^p is the (n, p) contraction of some n-particle density operator, we called D^p *n-representable* and the set of all n-representable p-particle density operators we denoted by $\mathfrak{F}_n^p$. We proved that the knowledge of the exposed points of $\mathfrak{F}_n^p$ is sufficient to characterize the closure of $\mathfrak{F}_n^p$ in some suitable topology (NRP, Theorem 9). Moreover, we were able to give a characterization of the exposed points of $\mathfrak{F}_n^p$.

Let us call the adjoint mapping Γ_p^n of the (n, p) contraction operator L_n^p the (p, n) *expansion operator*. Γ_p^n maps a bounded p-particle operator B^p (i.e., a bounded self-adjoint operator in H^p whose range is contained in $H^{p\wedge}$) onto the following operator in n-particle space:

$$\Gamma_p^n(B^p) = A_n(B^p \otimes I^{n-p})A_n. \tag{1.3}$$

A_n stands for the projector of H^n onto the physically relevant subspace $H^{n\wedge}$. An operator in n-particle space of the form (3) we called a *p-particle observable* of the system of n particles.

A *p-subspace* of $H^{n\wedge}$ is any subspace of $H^{n\wedge}$ which occurs as the null space of some *positive* p-particle observable. The set $\mathfrak{L}^p(H^{n\wedge})$ of all p-subspaces of $H^{n\wedge}$ forms a lower sublattice of the set $\mathfrak{L}(H^{n\wedge})$ of all subspaces of $H^{n\wedge}$ containing the subspace [0]. A minimal element of the partially ordered set $\mathfrak{L}^p(H^{n\wedge}) - [0]$ is called a *minimal p-subspace*.

The full pre-image of an exposed point of $\mathfrak{F}_n^p$ with respect to the (n, p) contraction L_n^p consists of all n-particle density operators whose range is contained in some minimal p-subspace (NRP, Theorem 8).

In this paper we wish to investigate the analogous problem for a system consisting of an *indefinite but conserved* number of particles.

A *state* of such a system can be described by a sequence,

$$\mathbf{D} = (D^0, D^1, \cdots), \tag{1.4}$$

of density operators, where the nth component operates in n-particle space $H^{n\wedge}$ and the following normalization condition is satisfied:

$$\sum_{n=0}^{\infty} \mathrm{Tr}\,(D^n) = 1. \tag{1.5}$$

We introduce the symbol $\mathfrak{F}$ to represent the set of all sequences $\mathbf{D}$ which satisfy Eq. (5) and, in addition, fulfill certain conditions restricting the behavior of the sequence for $n \to \infty$.

On the other hand, a (bounded) *observable* can be represented by a sequence,

$$\mathbf{B} = (B^0, B^1, B^2, \cdots),$$

of bounded self-adjoint operators, where the range of the nth component is contained in $H^{n\wedge}$.

If we introduce the bilinear product

$$(\mathbf{B} \mid \mathbf{D}) = \sum_{n=0}^{\infty} \mathrm{Tr}\,(B^n D^n), \tag{1.6}$$

we can express the expectation value of the observable $\mathbf{B}$ in the state $\mathbf{D}$ by

$$\langle \mathbf{B} \rangle_{\mathbf{D}} = (\Gamma(\mathbf{B}) \mid \mathbf{D}) = (\mathbf{B} \mid L(\mathbf{D})).$$

Γ stands for the expansion operator defined componentwise by

$$\Gamma(\mathbf{B})^n = \sum_{p=0}^{n} \binom{n}{p} \Gamma_p^n(B^p),$$

and L denotes the contraction operator which can be considered as the adjoint operator of Γ with respect to the bilinear product (6). It has the following componentwise representation:

$$L(\mathbf{D})^m = \sum_{n=m}^{\infty} \binom{n}{m} L_n^m(D^n).$$

Let us introduce the projection Π^p which maps $\mathbf{B}$ and $\mathbf{D}$ onto their respective "initial segments of length p." Thus, e.g.,

$$\Pi^p(\mathbf{B})^n = B^n, \quad n \leq p,$$
$$= 0, \quad n > p.$$

A p-particle observable of our system can be defined as a bounded observable for which $\Pi^p(\mathbf{B}) = \mathbf{B}$. Its expectation value in the state $\mathbf{D}$ is given by

$$(\Gamma(\mathbf{B}) \mid \mathbf{D}) = (\Pi^p(\mathbf{B}) \mid L(\mathbf{D})) = (\mathbf{B} \mid \Pi^p \cdot L(\mathbf{D}))$$
$$= (\mathbf{B} \mid L^p(\mathbf{D})).$$

Here we have introduced the symbol L^p for the p contraction operator defined by

$$L^p = \Pi^p \circ L.$$

It replaces the (p, n) contraction operator L_n^p in the case of a fixed number (n) of particles.

The image $\mathfrak{F}^{[p]} \equiv L^p(\mathfrak{F})$ is a subset of the set

$$\mathfrak{F}^0_{(p)} \equiv \{\mathbf{D};\ \mathbf{D} = (1, D^1, \cdots, D^p, 0, \cdots)\}$$

of sequences of density operators whose zeroth component equals 1 and whose components of higher order than p vanish.

An analogous problem as in the case of a definite particle number arises: Give a characterization of $\mathfrak{F}^{[p]}$ as a subset of $\mathfrak{F}^0_{(p)}$. One could call this problem the *representability problem*. Since $\mathfrak{F}$ is convex and L^p is a linear mapping, $\mathfrak{F}^{[p]}$ is a convex subset of $\mathfrak{F}^0_{(p)}$. We shall call an element of $\mathfrak{F}^0_{(p)}$ a *potential p state* and an element of $\overline{\mathfrak{F}^{[p]}}$ an *actual p* state. (The bar refers to the closure in a suitable topology.) We shall prove that the closure $\overline{\mathfrak{F}^{[p]}}$ can be described as the closed convex hull of certain special points of $\mathfrak{F}^{[p]}$ which will turn out to be related to the exposed points of $\mathfrak{F}^p_n$ for some $n \geq p$ and the exposed points of $\mathfrak{F}^n$ for some $n \leq p$ (Theorem 6).

In the special case where $p = 1$ we obtain the well-known result that in the case of fermions a potential one-state is actual if and only if the one-component D^1 has eigenvalues not greater than one, whereas in the case of bosons every potential one-state is also actual—a statement which represents some weak form of the Pauli principle (cf. Theorem 7).

Finally, we are able to build up a rigorous theory of the BCS ensemble (Sec. 4). An ordinary BCS ensemble is associated with a "geminal," i.e., a two-particle function. The set of all BCS ensembles constitutes a connected subset of the set of all states of the many-body system. By addition of some adherence points we can extend this set to the set of all generalized BCS ensembles. Such a generalized BCS ensemble is derived from a generalized AGP function (antisymmetrized geminal power) introduced by Coleman.[2] This is a function of the form $g \wedge \cdots \wedge g \wedge [U]$, where g stands for a geminal and $[U]$ for a Slater determinant associated with the (finite-dimensional) subspace U of the one-particle Hilbert space. In Theorem 8 we give the result of a calculation of the one- and the two-density operator of such a function, a task which already has been accomplished by Coleman[3] in the case of $U = 0$ and of *finite* rank of the geminal g. However, since the extension to the case of infinite rank is nontrivial, we feel that it is in the interest of the reader of this paper if we repeat the calculation.

Theorem 9 presents an explicit form of the one- and the two-density operator of a generalized BCS ensemble, and Theorem 10 confirms a conjecture by Coleman[4] that every one-state whose one-component possesses evenly degenerate eigenvalues is the one-contraction of a generalized BCS ensemble.

2. MATHEMATICAL PRELIMINARIES

In NRP we have shown that the fundamental mathematical object that describes a quantum-mechanical system is a *dual pair of topological real linear spaces*, rather than a Hilbert space.

In conventional quantum mechanics this pair is constructed with the help of a Hilbert space, but, as we wish to demonstrate in this paper, it is not always desirable to stick to this method of construction.

We intend to give a mathematically rigorous description of a system consisting of an indefinite number of identical particles for which the particle number is an integral of motion. In this case both members of the fundamental dual pair $\langle \mathcal{O}, \varphi \rangle$ ($\mathcal{O}$ = set of observables, φ = linear space generated by the states) have to be chosen as certain linear subspaces of the direct product space:

$$\bigtimes_{n=0}^{\infty} S^n = \{\mathbf{B};\ \mathbf{B} = (B^0, B^1, \cdots),\ B^n \in S^n\}, \quad (2.1)$$

where S^n is defined (as in NRP) as the set of all bounded self-adjoint operators in H^n whose range is contained in $H^{n\wedge}$. (We introduce the definition $H^{0\wedge} \equiv \mathbf{C} \equiv$ field of complex numbers.)

We define here the members of the dual pair, which we consider to be relevant for the physical system under consideration without further justification.

Later on, in the course of the development of our scheme, we shall have the opportunity to point to the different reasons which lead us to this particular choice.

First we give the definition of the component (describing the set of observables) of the dual pair. Let $\mathcal{O}^k$ be the following subspace of the product space $\bigtimes_{n=0}^{\infty} S^n$:

$$\mathcal{O}^k = \left\{\mathbf{B} \in \bigtimes_{n=0}^{\infty} S^n;\ \sup_n k^{-n} \|B^n\| < \infty\right\}, \quad k = 1, 2, 3 \cdots. \quad (2.2)$$

(We adopt the convention $k^0 = 1$.) It is a consequence of Lemma A1 of the Appendix that $\mathcal{O}^k$, equipped with the norm

$$q_k(\mathbf{B}) = \sup_n k^{-n} \|B^n\|, \quad (2.3)$$

is a Banach space. It is clear that $q_{k+1}(\mathbf{B}) \leq q_k(\mathbf{B})$, and hence $\mathcal{O}^k \subseteq \mathcal{O}^{k+1}$. $\mathcal{O}$ is defined as the union of all the spaces $\mathcal{O}^k$:

$$\mathcal{O} = \bigcup_{k=1}^{\infty} \mathcal{O}^k. \quad (2.4)$$

[2] A. J. Coleman, "Lectures on Superconductivity," given at Queen's University, Kingston, Ontario, Canada, 1965–66.
[3] A. J. Coleman, J. Math. Phys. 6, 1425 (1965).
[4] A. J. Coleman, private communication.

A natural topology to introduce in $\mathcal{O}$ is the so-called inductive limit topology, i.e., the strongest topology such that all the imbeddings

$$j^k: \mathcal{O}^k \to \mathcal{O} \tag{2.5}$$

are continuous. A basis of neighborhoods of 0 is obtained by taking the neighborhoods in each of the $\mathcal{O}^k$'s. Since this topology is only of minor importance in the present context, we shall refrain from commenting further on it.

Next, we give the definition of the other component (space spanned by the states) of the dual pair. φ is defined as the intersection of all real linear spaces:

$$\varphi^k \equiv \{\mathbf{D}; \mathbf{D} = (D^0, D^1 \cdots), D^n \in S_T^n, \sum_n k^n |D^n| < \infty\},$$
$$k = 1, 2 \cdots. \tag{2.6}$$

Lemma A1 provides us with the result that φ^k, equipped with the norm

$$p_k(\mathbf{D}) = \sum_{n=0}^{\infty} k^n |D^n|, \quad k = 1, 2, 3 \cdots, \tag{2.7}$$

is a Banach space. Since obviously

$$p_k(\mathbf{D}) \leq p_{k+1}(\mathbf{D}) \tag{2.8}$$

for all $\mathbf{D} \in \bigtimes_{n=0}^{\infty} S_T^n$, it follows that

$$\varphi^{k+1} \subseteq \varphi^k.$$

It is easy to see that

$$\varphi = \bigcap_{k=1}^{\infty} \varphi^k \tag{2.9}$$

is a *complete* locally convex linear space under the topology generated by the sequence $\{p_k(\mathbf{D})\}$ of norms. Such a space is metrizable (cf. Robertson[5]) and is called a Fréchet space (F-space).

The topology in φ can also be described as the weakest topology such that the imbedding mappings

$$j_k: \varphi \to \varphi^k$$

are continuous.

The following injections are of some importance for the sequel:

$$i_n: S^n \to \varphi$$

defined by

$$i_n: B^n \mapsto (0, 0, \cdots, B^n, 0, \cdots), \quad B^n \in S^n, \tag{2.10}$$

and the projections

$$pr_n: \mathcal{O} \to S^n$$

defined by

$$pr_n: \mathbf{B} \mapsto B^n, \quad \mathbf{B} \in \mathcal{O}. \tag{2.11}$$

[5] A. P. Robertson and W. J. Robertson, "Topological Vector Spaces," Cambridge Tracts in Mathematics and Mathematical Physics, No. 53 (1964).

Since

$$p_k(i_n(D^n)) = k^n |D^n|, \quad D^n \in S_T^n, \tag{2.12}$$

and

$$|pr_n(\mathbf{D})| = |D^n| \leq k^{-n} |\mathbf{D}|, \tag{2.13}$$

it follows that the restrictions of i_n to S_T^n and pr_n to φ are continuous maps.

For the image $i_n(S^n) \subset \mathcal{O}$, we choose the notation $\mathcal{O}^{(n)}$, whereas $\mathcal{O}_{(n)}$ will designate *the initial segment of length n* of $\mathcal{O}$ defined by

$$\mathcal{O}_{(n)} \equiv \{\mathbf{B} \in \mathcal{O}; pr_m(\mathbf{B}) = 0, \forall m > n\}.$$

In a similar way we denote by $\varphi^{(n)}$ the image $i_n(S_T^n)$, and by $\varphi_{(n)}$ the initial segment of length n defined by

$$\varphi_{(n)} \equiv \{\mathbf{D} \in \varphi; pr_m(\mathbf{D}) = 0, \forall m > n\}.$$

From (12) it follows that $\varphi^{(n)}$ and $\varphi_{(n)}$ are closed subspaces of φ.

Lemma 1: $\langle \mathcal{O}, \varphi \rangle$ is a dual pair of real linear spaces with respect to the bilinear form:

$$(\mathbf{B}|\mathbf{D}) = \sum_{n=0}^{\infty} \mathrm{Tr}\,(B^n D^n), \quad \mathbf{B} \in \mathcal{O}, \mathbf{D} \in \varphi \tag{2.14}$$

(cf. definition A1 of the Appendix).

Proof: First of all, we show that the series on the right side converges absolutely. From NRP (Lemma 4) we know that

$$\mathrm{Tr}\,(B^n D^n) \leq \|B^n\|\, |D^n|.$$

It follows that

$$|(\mathbf{B} \mid \mathbf{D})| \leq \sum_{n=0}^{\infty} |\mathrm{Tr}\,(B^n D^n)| \leq \sum_{n=0}^{\infty} \|B^n\|\, |D^n|$$
$$= \sum_{n=0}^{\infty} k^{-n}\, \|B^n\|\, k^n\, |D^n| \leq q_k(\mathbf{B}) p_k(\mathbf{D}).$$

The last expression is finite if k is chosen to be sufficiently large.

Now let

$$(\mathbf{B} \mid \mathbf{D}) = 0, \quad \forall \mathbf{B} \in \mathcal{O}.$$

In particular, this expression has to vanish for $\mathbf{B} \in S^{(n)}$. This leads to

$$\mathrm{Tr}\,(B^n D^n) = 0, \quad \forall B^n \in S^n. \tag{2.15}$$

But since $\langle S^n, S_T^n \rangle$ represents a dual pair with respect to the bilinear expression (14) (namely, the fundamental dual pair associated with a system of n identical particles), it follows that $D^n = 0$. Since n was chosen arbitrarily, it is immediate that $\mathbf{D} = 0$. In a similar way one shows that

$$(\mathbf{B} \mid \mathbf{D}) = 0, \quad \forall \mathbf{D} \in \varphi,$$

implies $\mathbf{B} = 0$. Q.E.D.

Remark: From this proof it is obvious that for any natural number p, $\langle \mathcal{O}_{(p)}, \varphi_{(p)} \rangle$ constitutes a dual pair of real linear spaces.

Lemma 2: The topology induced by the seminorms $\{p_k(\mathbf{D})\}$ on φ is compatible with the dual pair $\langle \mathcal{O}, \varphi \rangle$.

Proof: Let f be a continuous linear functional on φ. Since for $k = 1, 2, \cdots$,

$$\lim_{n\to\infty} p_k\left(\mathbf{D} - \sum_{p=0}^{n} i_p(D^p)\right) = \lim_{n\to\infty} \sum_{l=n+1}^{\infty} k^n |D^n| = 0,$$

it follows that

$$f(\mathbf{D}) = \lim_{n\to\infty} \sum_{p=0}^{n} f(i_p(D^p)) = \sum_{p=0}^{\infty} f_p(D^p),$$

where

$$f_p = f \circ i_p.$$

Since i_p and f are both continuous, f_p is a continuous linear functional on S_T^p. From the results contained in the excellent book of Schatten's[6] (which we have mentioned in NRP), it follows that f_p can be written as

$$f_p(D^p) = \mathrm{Tr}\,(B^p D^p)$$

for some operator $B^p \in S^p$.

Hence f has the form

$$f(\mathbf{D}) = \sum_{p=0}^{\infty} \mathrm{Tr}\,(B^p D^p).$$

It remains to be shown that the sequence $\mathbf{B} = (B^0, B^1, B^2, \cdots)$ belongs to $\mathcal{O}$, i.e., that

$$\sup_p k^{-p} \|B^p\| < \infty \quad \text{for some } k.$$

Let us assume the contrary. Then to every natural number k there exists an integer $n(k)$ such that

$$k^{-n(k)} \|B^{n(k)}\| > k.$$

Since

$$\|B^n\| = \sup_{|D^n|=1} |\mathrm{Tr}\,(B^n D^n)|,$$

we can find $D_0^{n(k)}$ with $|D_0^{n(k)}| = 1$, such that

$$k^{-n(k)} \mathrm{Tr}\,(B^{n(k)} D_0^{n(k)}) > k.$$

Let us define

$$\mathbf{D}_k \equiv i_{n(k)}(k^{-n(k)} D_0^{n(k)}).$$

Then

$$\lim_{k\to\infty} f(\mathbf{D}_k) = \lim_{k\to\infty} \mathrm{Tr}\,(B^{n(k)} D_0^{n(k)}) k^{-n(k)} = \infty,$$

because of the above inequality.

On the other hand, it follows from

$$\lim_{k\to\infty} p_l(\mathbf{D}_k) = \lim_{k\to\infty} (l/k)^{n(k)} = 0, \quad l = 1, 2, \cdots,$$

that the sequence $\{\mathbf{D}_k\}$ converges to 0. In other words, it follows that f is not continuous contrary to our assumption. Q.E.D.

Lemma 1 allows us to introduce natural topologies into $\mathcal{O}$ and φ, namely, the so-called *weak topologies* associated with the dual pair (cf. Appendix, definition A3). From Lemma 2 it follows that the weak topology in φ is not stronger than the topology induced by the norms $\{p_k(\mathbf{D})\}$.

Since

$$(\mathbf{B} \mid i_n(D^n)) = \mathrm{Tr}\,(pr_n(\mathbf{B}) D^n), \quad \forall \mathbf{B} \in \mathcal{O}, \quad D^n \in S_T^n, \tag{2.16}$$

and

$$(i_n(B^n) \mid \mathbf{D}) = \mathrm{Tr}\,(B^n pr_n(\mathbf{D})), \quad \forall \mathbf{D} \in \varphi, \quad B^n \in S^n, \tag{2.17}$$

it follows from Theorem A4 of the Appendix that the mappings i_n and pr_n and their restrictions to S_T^n and φ are continuous if we think of all four members of the two dual pairs $\langle \mathcal{O}, \varphi \rangle$ and $\langle S^n, S_T^n \rangle$ as being equipped with their respective weak topology. Moreover, we easily recognize that the topologies induced by the weak topology of φ onto $\varphi_{(n)}$ and by the weak topology onto $\mathcal{O}_{(n)}$ coincide with the weak topologies associated with the dual pair: $\langle \mathcal{O}_{(n)}, \varphi_{(n)} \rangle$.

In accordance with the conventional description of a quantum-mechanical system, we have to make the following identifications: $\mathcal{O}$ is the set of all observables of the system; $\mathcal{S}$ is the set of all possible mixed states of the system, where

$$\mathcal{S} \equiv \left\{ \mathbf{D} \in \varphi;\ \forall_n,\ D^n \geq 0, \sum_{n=0}^{\infty} \mathrm{Tr}\,(D^n) = 1 \right\}. \tag{2.18}$$

The projection cone of $\mathcal{S}$ from the center 0 in φ is denoted by φ^+:

$$\varphi^+ \equiv \{\mathbf{D};\ \forall_n,\ D^n \geq 0\}.$$

φ^+ can be considered as the cone of positive elements of an order relation $\geq$ with respect to which φ forms a partially ordered vector space.

In a similar way we define the set $\mathcal{O}^+$ by

$$\mathcal{O}^+ \equiv \{\mathbf{B} \in \mathcal{O};\ \forall_n,\ B^n \geq 0\}$$

and call an observable $\mathbf{B} \in \mathcal{O}^+$ *hyperpositive*. The reason for this choice of rather a sophisticated name lies in our intention to introduce later on another (weaker) order relation into $\mathcal{O}$. It will be of greater physical significance than the one induced by $\mathcal{O}^+$. We want to reserve the name "positive observable" to designate an element of $\mathcal{O}$ which is positive with

[6] R. Schatten, *Norm Ideals of Completely Continuous Operators* (Springer-Verlag, Berlin, 1960).

respect to this second physically more important order relation.

In the sequel we adopt from NRP the notation N_A for the null space of a linear operator A and R_A for the *closure* of its range. The following theorem describes the cones $\mathcal{O}^+$ and φ^+ geometrically. (The reader should compare it with Theorem 2 of NRP.)

Theorem 1: (i) The sets $\mathcal{O}^+$ and φ^+ are the polar cones of each other (cf. definition A5 of the Appendix) and therefore weakly closed. (ii) A subset of $\mathcal{O}^+$ is exposed (cf. definition A13 of the Appendix) if and only if it is of the form

$$\mathcal{O}(\mathbf{V}) \equiv \{\mathbf{B} \in \mathcal{O}^+;\ R_{B^n} \subseteq V_n\}, \tag{2.19}$$

where $\mathbf{V}$ stands for a sequence $\mathbf{V} = (V_0, V_1, V_2 \cdots)$ of closed subspaces $V_n \subseteq H^{n\wedge}$. The exposed rays are therefore of the form $\{\rho i_n(P);\ \rho \geq 0\}$, where P denotes a one-dimensional projector in $H^{n\wedge}$ for some n. The statement remains valid after replacing $\mathcal{O}^+$ by φ^+. (iii) $\mathcal{O}^+$ is the convex closure of its exposed rays and every extreme ray is exposed. This statement is also true for φ^+.

Proof: The theorem is an almost trivial consequence of NRP Theorem 2.

(i) Let us determine the polar cone of φ^+ defined by

$$\tilde{\varphi}^+ \equiv \{\mathbf{B} \in \mathcal{O} : (\mathbf{B} \mid \mathbf{D}) \geq 0,\ \forall \mathbf{D} \in \varphi^+\}.$$

In order that $\mathbf{B} \in \tilde{\varphi}^+$, it is necessary that, for all natural numbers n and all $D^n \in S_T^{n+} \equiv \mathcal{K}_T^n$,

$$(\mathbf{B} \mid i_n(D^n)) = \operatorname{Tr}(B^n D^n) \geq 0.$$

From NRP Theorem 2 it follows that $B^n \in S^{n+} \equiv \mathcal{K}^n$ for all n, so that $\mathbf{B} \in \mathcal{O}^+$.

On the other hand, if $\mathbf{B} \in \mathcal{O}^+$, then obviously

$$(\mathbf{B} \mid \mathbf{D}) = \sum_{n=0}^{\infty} \operatorname{Tr}(B^n D^n) \geq 0$$

for all $\mathbf{D} \in \varphi^+$.

(ii) First let $\mathcal{O}^+(\mathbf{V})$ be a subset of $\mathcal{O}^+$ of the indicated type. Choose operators $D^n \in \mathcal{K}_T^n$ such that $\mathbf{D} = (D^0, D^1, D^2, \cdots)$ belongs to φ and the null space of D^n coincides with V_n: $N_{D^n} = V_n$. Now apply NRP Lemma 3 to get

$$\{\mathbf{D}\}^\perp \cap \mathcal{O}^+ = \mathcal{O}^+(\mathbf{V}).$$

[The formation of the orthogonal complement $\perp$ is understood with respect to the bilinear form $(\mathbf{B} \mid \mathbf{D})$.] Conversely, let $\{\mathbf{D}\}^\perp \cap \mathcal{O}^+$, $\mathbf{D} \in \varphi^+$, be an exposed subset of $\mathcal{O}^+$ and let V_n be the null space of D^n. Then it is again an immediate consequence of NRP Lemma 3 that

$$\{\mathbf{D}\}^\perp \cap \mathcal{O}^+ = \{\mathbf{B} \in \mathcal{O}^+;\ R_{B^n} \subseteq V_n\} = \mathcal{O}^+(\mathbf{V}).$$

A similar proof holds in the case of φ^+.

(iii) Let us denote by E the set of all elements of the form $i_n(P^n)$, where P^n is a one-dimensional projector in $H^{n\wedge}$ for some n. Since $E \subset \mathcal{O}^+$, it follows that $\tilde{E} \supset \varphi^+$, where the tilde symbolizes polarization (cf. definition A11 of the Appendix). Let $\mathbf{D} \in \tilde{E}$. Then

$$(\mathbf{D} \mid i_n(P^n)) = \operatorname{Tr}(P^n D^n) \geq 0$$

for all n and all one-dimensional projectors P^n. Hence $\mathbf{D} \in \varphi^+$ and $\tilde{E} = \varphi^+$. From the bipolar theorem (Theorem A6 of the Appendix), it now follows that the smallest closed convex cone containing E coincides with $\mathcal{O}^+$:

$$\tilde{\tilde{E}} = \tilde{\varphi}^+ = \mathcal{O}^+.$$

An analogous argument holds in the φ^+ case. Q.E.D.

As in places in NRP, we denote by W a finite-dimensional subspace of H^1 and by $W^{n\wedge}$ the subspace

$$W^{n\wedge} = \bigotimes_{i=1}^{n} W(i) \cap H^{n\wedge} \tag{2.20}$$

of the n-particle space $H^{n\wedge}$. Furthermore, we introduce the notation $S^n[W]$ for the set

$$S^n[W] \equiv \{B \in S^n;\ R_B \subset W^{n\wedge}\}$$

and

$$\varphi[W] = \{\mathbf{B} \in \varphi;\ B^n \in S^n[W],\ \forall n\}.$$

Other sets which are of interest in the sequel are

$$\mathfrak{T}[W] = \mathfrak{T} \cap \varphi[W]$$

and

$$\mathfrak{T}_{(n)}[W] = \mathfrak{T}[W] \cap \varphi_{(n)}.$$

Theorem 2: Let $W_1 \subset W_2 \subset \cdots W_k \subset \cdots H^1$ be an increasing sequence of finite-dimensional subspaces of H^1 with the property that

$$\overline{\bigcup_k W_k} = H^1.$$

Then we have for $n = 0, 1, 2, \cdots$

(i) $$\mathfrak{T}_{(n)} = \overline{\bigcup_k \mathfrak{T}_{(n)}[W_k]},$$

(ii) $$\mathfrak{T} = \overline{\bigcup_n \mathfrak{T}_{(n)}},$$

and hence

(iii) $$\mathfrak{T} = \overline{\bigcup_n \bigcup_k \mathfrak{T}_{(n)}[W_k]}.$$

The bar in (i)–(iii) denotes the closure in any topology compatible with the dual pair $\langle \mathcal{O}, \varphi \rangle$.

Remark: This theorem should be compared with Proposition 1 in NRP.

Proof: Let $\mathbf{D} = (D^0, D^1, D^2, \cdots, D^n, 0 \cdots) \in \mathfrak{S}_{(n)}$ and put for $m \geq 1$ $\hat{D}^m = D^m/\mathrm{Tr}\,(D^m)$ whenever $D^m \neq 0$. Then $\hat{D}^m \in \mathfrak{S}^m$, and it follows from NRP Proposition 1 that there exists a sequence

$$\hat{D}_l^m \in \bigcup_k \mathfrak{S}^m[W_k]$$

such that

$$\lim_{l\to\infty} |\hat{D}_l^m - \hat{D}^m| = 0.$$

Now put

$$D_l^0 = D^0, \quad \forall l,$$

$$D_l^m = \mathrm{Tr}\,(D^m)\hat{D}_l^m, \quad \text{if} \quad D^m \neq 0,$$

$$= 0, \quad \text{if} \quad D^m = 0,$$

and

$$\mathbf{D}_l = (D_l^0, D_l^1 \cdots, D_l^m, \cdots, D_l^n, 0 \cdots).$$

Then clearly

$$\mathbf{D}_l \in \mathfrak{S}_{(n)}[W_k]$$

and

$$p_k(\mathbf{D}_l - \mathbf{D}) = \sum_{m=1}^{n} k^m |D_l^m - D^m|$$

$$= \sum_{m=1}^{n} k^m \,\mathrm{Tr}\,(D^m)\,|\hat{D}_l^m - \hat{D}^m|.$$

Hence

$$\lim_{l\to\infty} p_k(\mathbf{D}_l - \mathbf{D}) = 0.$$

This proves part (i) of Theorem 2. To prove part (ii), let $\mathbf{D} \in \mathfrak{S}$ and define

$$\alpha_n \equiv \sum_{m=0}^{n} \mathrm{Tr}\,(D^m).$$

Then

$$\lim_{n\to\infty} \alpha_n = 1.$$

Let Π^n be the projection of φ onto $\varphi_{(n)}$ defined by

$$\Pi^n(\mathbf{D})^m = D^m \quad \text{for} \quad m \leq n,$$

$$\Pi^n(\mathbf{D})^m = 0 \quad \text{for} \quad m > n. \tag{2.21}$$

(It is clear that Π^n is continuous with respect to both topologies introduced in φ.) Let n_0 be the smallest integer for which $D^n \neq 0$ and for all $n \geq n_0$ put

$$\mathbf{D}_n = \alpha_n^{-1}\Pi^n(\mathbf{D}).$$

Then $\mathbf{D}_n \in \mathfrak{S}_{(n)}$ and for $n \geq n_0$ it follows that

$$p_k(\mathbf{D}_n - \mathbf{D})$$

$$= \left(\frac{1}{\alpha_n} - 1\right) \sum_{m=0}^{n} k^m |D^m| + \sum_{m=r+1}^{\infty} k^m |D^m|$$

$$= \left(\frac{1}{\alpha_n} - 1\right) p_k(\mathbf{D}) + \left(2 - \frac{1}{\alpha_n}\right) \sum_{k=n+1}^{\infty} k^m |D^m|.$$

For $n \to \infty$, both summands on the right side converge towards zero. Q.E.D.

3. EXPANSION AND CONTRACTION OPERATORS

We define an operator Γ in $\mathcal{O}$ which is the analog of the (p, n) expansion operator defined in NRP.

Definition 3.1: The expansion operator is defined by the formulas

$$\Gamma(\mathbf{B})^n = \sum_{j=0}^{n} \binom{n}{j} \Gamma_j^n(B^j), \quad n = 0, 1, \cdots, \tag{3.1}$$

where $B^j \equiv pr_j(\mathbf{B})$ and Γ_p^n stands for the (p, n) expansion operator in NRP defined by

$$\Gamma_p^n(B^p) = A_n(B^p \otimes I^{n-p})A_n. \tag{3.2}$$

That Γ carries an element $\mathbf{B} \in \mathcal{O}$ again into an element of $\mathcal{O}$ is stated in the following lemma:

Lemma 4: If $\mathbf{B} \in \mathcal{O}^k$, then $\Gamma(\mathbf{B})$ belongs to $\mathcal{O}^{k+1}$.

Proof: Let $\mathbf{B} \in \mathcal{O}^k$. Then

$$\|\Gamma(B)_n\| \leq \sum_{j=0}^{n} \binom{n}{j} \|B^j\| = \sum_{j=0}^{n} \binom{n}{j} k^j \|B^j\| \, k^{-j}$$

$$\leq q_k(\mathbf{B}) \cdot \sum_{j=0}^{n} \binom{n}{j} k^j = (k+1)^n q_k(\mathbf{B}).$$

Hence

$$[\|\Gamma(B)^n\|/(k+1)^n] \leq q_k(\mathbf{B}), \quad n = 0, 1, 2, \cdots.$$

Q.E.D.

Remark: It is clear that Lemma 4 is a consequence of our particular choice of the dual pair. For assume that

$$\mathcal{O} = \bigcup_k \mathcal{O}^k$$

with

$$\mathcal{O}^k = \left\{\mathbf{B} \in \bigtimes_{n=0}^{\infty} S^n;\ \sup_n [\|B^n\|/g(n, k)] < \infty\right\}$$

for some function $g(n, k)$. Assume that $\mathbf{B} \in \mathcal{O}^k$. Then, by the same argument as in the proof of Lemma 4,

$$\|\Gamma(B)^n\| \leq \sup_n \frac{\|B^n\|}{g(n, k)} \sum_{j=0}^{n} \binom{n}{j} g(n, j).$$

Thus we have certainly $\Gamma(\mathbf{B}) \in \mathcal{O}^{k+1}$ if

$$g(n, k+1) = \sum_{j=0}^{n} \binom{n}{j} g(n, j). \tag{3.3}$$

Together with the assumption $g(n, 1) \equiv 1$, recursion formula (3) leads automatically to $g(n, k) \equiv k^n$.

Proposition 1: The expansion operator Γ is the adjoint mapping of the contraction operator $L: \varphi \to \varphi$ defined by the equations

$$L(\mathbf{D})^m = \sum_{n=m}^{\infty} \binom{n}{m} L_n^m(D^n), \tag{3.4}$$

where L_n^m stands for the (n, m) contraction operator defined in NRP (Theorem 3).

Proof: First of all, we show that the series (3.4) is convergent inside S_T^m. For this purpose it is sufficient to prove that the series

$$\sum_{n=m}^{\infty} \binom{n}{m} |L_n^m(D^n)|$$

is convergent. But this is shown by the following sequence of inequalities:

$$\begin{aligned}\sum_{n=m}^{\infty} \binom{n}{m} |L_n^m(D^n)| &= \sum_{n=m}^{\infty} \binom{n}{m} |D^n| \\ &= \sum_{n=m}^{\infty} (m+1)^{-n} \binom{n}{m} (m+1)^n |D^n| \\ &\leq \sup_n \left[(m+1)^{-n} \binom{n}{m} \right] \cdot p_{m+1}(\mathbf{D}) \\ &\leq \frac{1}{m^m} p_{m+1}(\mathbf{D}).\end{aligned}$$

Herein we made use of the equality

$$|L_n^m(D^n)| = \operatorname{Tr}(L_n^m(D^n)) = \operatorname{Tr}(D^n) = |D^n|.$$

Moreover, we prove that $L(\mathbf{D}) \in \varphi$. Indeed we have

$$\begin{aligned}p_k(L(\mathbf{D})) &= \sum_{m=0}^{\infty} k^m \left| \sum_{n=m}^{\infty} \binom{n}{m} L_n^m(D^n) \right| \\ &= \sum_{m=0}^{\infty} \sum_{n=m}^{\infty} k^m \binom{n}{m} |D^n| \\ &= \sum_{n=0}^{\infty} |D^n| \sum_{m=0}^{n} k^m \binom{n}{m} \\ &= \sum_{n=0}^{\infty} (k+1)^n |D^n| \equiv p_{k+1}(\mathbf{D}).\end{aligned}$$

Another consequence of this equality is that L is continuous with respect to the topology generated by the norms $p_k(\mathbf{D})$.

Since the mapping Γ_j^n is adjoint to L_n^j (NRP Theorem 4), we obtain

$$\begin{aligned}(\Gamma(\mathbf{B}) \mid \mathbf{D}) &= \sum_{n=0}^{\infty} \operatorname{Tr}(\Gamma(\mathbf{B})^n D^n) \\ &= \sum_{n=0}^{\infty} \sum_{j=0}^{n} \binom{n}{j} \operatorname{Tr} \binom{n}{j} \operatorname{Tr}(\Gamma_j^n(B^j) D^n) \\ &= \sum_{j=0}^{\infty} \sum_{n=j}^{\infty} \binom{n}{j} \operatorname{Tr}(B^j L_n^j(D^n)) = (\mathbf{B} \mid L(\mathbf{D})).\end{aligned}$$

Q.E.D.

Theorem 3: The mappings L and Γ have the following properties:

(i) They are linear and order preserving (i.e., they leave the respective cones $\mathfrak{O}^+$ and φ^+ invariant).

(ii) They are bijective. The inverse mapping of Γ can be represented as

$$\Gamma^{-1}(\mathbf{B})^n = \sum_{j=0}^{n} (-1)^{n-j} \binom{n}{j} \Gamma_j^n(B^j), \tag{3.5}$$

whereas the inverse mapping of L can be defined by

$$L^{-1}(\mathbf{D})^m = \sum_{n=m}^{\infty} (-1)^{n-m} \binom{n}{m} L_n^m(D^n). \tag{3.6}$$

(iii) L and L^{-1} are continuous with respect to the topology generated by the seminorms $\{p_k(\mathbf{D})\}$.

(iv) They are topological mappings with respect to the weak topologies.

(v) They are operator automorphisms with respect to the group of unitary automorphisms of the one particle Hilbert space H^1.

Proof:

(i) This statement is an immediate consequence of the definition of the mappings L and Γ and the corresponding properties of the (n, p) contraction and (p, n) expansion operator as expressed in NRP Theorem 5(i).

(ii) That the mappings Γ^{-1} and L^{-1} as defined by (3.5) and (3.6) may be considered as operators in $\mathfrak{O}$ and φ can be proved by analogy with the case of Γ and L. That they are inverse operators of Γ and L is an easy consequence of the formula

$$\sum_{m=n}^{p} (-1)^{m-n} \binom{p}{m} \binom{m}{n} = \delta_{pn}, \tag{3.7}$$

where δ_{pn} is the Kronecker δ symbol. The validity of (3.7) can be shown in the following way: Applying twice the binomial theorem to $x^p = (x - 1 + 1)^p$, one obtains

$$\begin{aligned}x^p = (x-1+1)^p &= \sum_{m=0}^{p} \binom{p}{m} (x-1)^m \\ &= \sum_{m=0}^{p} \sum_{n=0}^{m} \binom{p}{m} \binom{m}{n} (-1)^{m-n} x^n \\ &= \sum_{n=0}^{p} \left(\sum_{m=n}^{p} \binom{p}{m} \binom{m}{n} (-1)^{m-n} \right) x^n.\end{aligned}$$

Comparing the coefficients of x^n $(n = 0, 1, \cdots, p)$ on both sides of this equation leads to the desired result.

(iii) This has already been proved for the mapping L. The proof for L^{-1} is completely analogous.

(iv) This statement is a consequence of Theorem A4 of the Appendix.

(v) Let $\mathfrak{U}^1$ be the group of unitary automorphisms of H^1 and $\Lambda^n(u)$, $u \in \mathfrak{U}^1$, its representation in $H^{n\wedge}$. Then, by

$$pr_n(u\mathbf{B}) =_{\cdot} \Lambda^n(u) pr_n(\mathbf{B}) \Lambda^n(u)^*, \tag{3.8}$$

we define a representation of $\mathfrak{U}^1$ in $\mathcal{O}$. Here u restricted to φ defines a homeomorphism with respect to the $\{p_k\}$ topology. Now Theorem A4 of the Appendix implies that u defines homeomorphisms with respect to the weak topologies in $\mathcal{O}$ and in φ.

The mappings Γ and L commute with this representation. This can be deduced immediately from their definitions and also from the fact stated in NRP [Theorem 5(v)] that the (p, n) expansion operator and the (n, p) contraction operator are operator homeomorphisms with respect to $\mathfrak{U}^1$. Q.E.D.

Now we are ready to indicate how the expectation value of the observable $\mathbf{B}$ in the state $\mathbf{D}$ can be calculated. It is given by the expression

$$\langle \mathbf{B} \rangle_{\mathbf{D}} = (\Gamma(\mathbf{B}) \mid \mathbf{D}) = (\mathbf{B} \mid L(\mathbf{D})). \tag{3.9}$$

We introduce a second-order relation in $\mathcal{O}$ by calling an observable $\mathbf{B}$ *positive* iff $\Gamma(\mathbf{B})$ is hyperpositive.

The set of all positive observables we denote by $\mathcal{O}_+$. Since Γ leaves the set of hyperpositive elements invariant [Theorem 3(i)], the new order relation is not stronger than the old one: $\mathcal{O}^+ \subseteq \mathcal{O}_+$. In fact, in general, it is strictly weaker, as is demonstrated by the following example.

Assume we are dealing with a system of fermions and consider the observable $\mathbf{B} = (\beta, B, 0, \cdots)$, where β is a strictly positive real number and B is a one-particle operator ($B \in S^1$) with a discrete spectrum having the additional property that the sum over any subset of eigenvalues exceeds $-\beta$. If B has any negative eigenvalues at all, then $\mathbf{B}$ is not hyperpositive, but it is positive since $\Gamma(\mathbf{B})$ is hyperpositive.

Equation (3.9) suggests that the states of a system of particles should be identified with the set $\hat{\mathfrak{F}} \equiv L(\mathfrak{F})$, rather than with $\mathfrak{F}$.

Proposition 2: $\hat{\mathfrak{F}}$ coincides with the intersection of the polar cone of all positive observables and the closed hyperplane of those sequences $\mathbf{D}$ whose zeroth component equals one. In symbols,

$$\hat{\mathfrak{F}} = \tilde{\mathcal{O}}_+ \cap \{\mathbf{D}; pr_0(\mathbf{D}) = 1\}. \tag{3.10}$$

Proof: We have the following chain of logically equivalent statements:

$$\begin{aligned}
\mathbf{B} \in \tilde{\hat{\mathfrak{F}}} &\Leftrightarrow (\mathbf{B} \mid \boldsymbol{\Delta}) \geq 0, \qquad \forall \boldsymbol{\Delta} \in \hat{\mathfrak{F}}, \\
&\Leftrightarrow (\mathbf{B} \mid L(\mathbf{D})) \geq 0, \quad \forall \mathbf{D} \in \mathfrak{F}, \\
&\Leftrightarrow (\Gamma(\mathbf{B}) \mid \mathbf{D}) \geq 0, \quad \forall \mathbf{D} \in \mathfrak{F}, \\
&\Leftrightarrow \Gamma(\mathbf{B}) \in \mathcal{O}^+, \\
&\Leftrightarrow \mathbf{B} \in \mathcal{O}_+.
\end{aligned}$$

Now the assertion of Proposition 2 is an immediate consequence of the bipolar theorem (Theorem A6 of the Appendix). Q.E.D.

Since L is a linear homeomorphism, it follows from Theorem 1 that $\hat{\mathfrak{F}}$ is the closed convex hull of the set of its exposed points which are the images of the exposed points of $\mathfrak{F}$ under L.

Let us summarize our insight in a theorem:

Theorem 4:

(i) The observables of a system consisting of an indefinite but conserved number of particles can be identified with the set $\mathcal{O}$.

(ii) The set of all states of the system can be identified with the intersection $\hat{\mathfrak{F}}$ of the polar cone $\hat{\mathcal{O}}_+$ of the set of all positive observables with the closed hyperplane

$$\varphi^0 \equiv \{\mathbf{D} \in \varphi; pr_0(\mathbf{D}) = 1\} \subseteq \varphi. \tag{3.11}$$

The exposed points of $\hat{\mathfrak{F}}$ have the form

$$\boldsymbol{\Delta} = \left(1, nL_n^1(P^n), \cdots, \binom{n}{m} L_n^m(P^n), \cdots, P^n, 0 \cdots\right), \tag{3.12}$$

where n is any nonnegative integer and P^n stands for any one-dimensional projector in $H^{n\wedge}$. These correspond to the pure states of the system.

(iii) The expectation value of the observable $\mathbf{B}$ in the state $\boldsymbol{\Delta}$ is given by

$$(\mathbf{B} \mid \boldsymbol{\Delta}) = \sum_{n=0}^{\infty} \mathrm{Tr}\,(B^n \Delta^n). \tag{3.13}$$

Remark: Since L is a bijection, the amount of information carried by $\boldsymbol{\Delta} = L(\mathbf{D})$ is equivalent to the amount of information carried by $\mathbf{D}$. In particular, it is sufficient to predict the time development of the system.

Definition 3.2: An observable belonging to $\mathcal{O}_{(p)}$ is called a *p-particle observable.*

Examples: The particle number N defined by

$$\mathbf{N} = (0, I, 0, \cdots), \quad I = \text{identity in } H^1, \tag{3.14a}$$

is a (hyperpositive) one-particle observable. Its square

$$\mathbf{N}^2 = (0, I, 2A_2, 0 \cdots) \tag{3.14b}$$

is a two-particle observable.

Let $\mathbf{B}$ be a p-particle observable. Let $\mathbf{\Delta} \in \hat{\mathfrak{F}}$ be a state. Then the expectation value of $\mathbf{B}$ in the state $\mathbf{\Delta}$ is given by

$$(\mathbf{B} \mid \mathbf{\Delta}) = (\mathbf{B} \mid \Pi^p(\mathbf{\Delta})), \tag{3.15}$$

where Π^p denotes the projection of φ onto $\varphi_{(p)}$ defined by (2.21).

In other words, to calculate the expectation value of a p-particle observable in the state $\mathbf{\Delta}$, we need only the first p components of the state. We denote by $\mathfrak{F}^{[p]}$ the set

$$\mathfrak{F}^{[p]} \equiv \Pi^p(\hat{\mathfrak{F}}) = \Pi^p \circ L(\mathfrak{F}) = L^p(\mathfrak{F})$$

and call an element of $\overline{\mathfrak{F}^{[p]}}$ an actual p-state.

Furthermore let us call a system of particles *p-reducible* if all the *experimentally accessible* observables of the system belong to $\mathcal{O}_{(p)}$.

In the case of a p-reducible system of particles, we are never in the position to know its state fully (not even in the sense of conventional quantum statistics), but only its p-state. But, according to the conventional scheme of quantum mechanics, the information contained in the p-state of a system is *not* sufficient to predict its future. Therefore we need a method to complete a p-state of a system to a full state in an unambiguous way.

In the present paper we assume the existence of such a method in order to be able to identify the set of states of a p-reducible system with $\overline{\mathfrak{F}^{[p]}}$.

Giles,[7] in a paper about classical statistical mechanics, has proposed such a method whose fundamental idea is indicated briefly as follows: The remaining components should be chosen in such a way as to minimize the amount of information about the system, i.e., to maximize the functional

$$S(\mathbf{D}) = -\sum_{n=0}^{\infty} \frac{1}{n!} \operatorname{Tr}(D^n \ln D^n) \tag{3.16}$$

under the constraints

$$L^p(\mathbf{D})^m = \Delta_0^m, \quad m = 0, 1, \cdots, p \quad (\Delta_0^0 = 1). \tag{3.17}$$

Without going into the details of this method, the only fact we wish to utilize here is that there does exist such a method which enables us to identify the states of a p-reducible system of particles with the set $\overline{\mathfrak{F}^{[p]}}$.

At this point a similar question arises as in the case of a definite number n of particles, where it takes the form of the n-representability problem. Let us denote by $\mathfrak{F}^0_{(p)}$ the set

$$\mathfrak{F}^0_{(p)} \equiv \{\mathbf{D} \in \varphi^+_{(p)};\ pr_0(\mathbf{D}) = 1\} \tag{3.18}$$

and let us call an element of $\mathfrak{F}^0_{(p)}$ a *potential p-state.* $\mathfrak{F}^0_{(p)}$ is a convex closed subset of $\mathfrak{F}^0_{(p)}$ and

$$\mathfrak{F}^{[p]} \subseteq \overline{\mathfrak{F}^{[p]}} \subseteq \mathfrak{F}^0_{(p)}. \tag{3.19}$$

Now the analogous problem to the n-representability problem takes the following form: Characterize $\mathfrak{F}^{[p]}$ or, equivalently, its closure $\overline{\mathfrak{F}^{[p]}}$ as a subset of $\mathfrak{F}^0_{(p)}$. Or, in another formulation, give a general criterion to decide if a given potential p-state is an actual p-state or not. We call this problem the *representability problem.*

Theorem 5:

(i) The observables of a p-reducible system of particles whose number is indeterminate but conserved in time can be identified with the set $\mathcal{O}_{(p)}$.

(ii) The set of all states of the system $\overline{\mathfrak{F}^{[p]}}$ can be represented as the intersection of the polar cone $\tilde{\mathcal{O}}_{(p)+}$ of the set $\mathcal{O}_{(p)+} \equiv \mathcal{O}_{(p)} \cap \mathcal{O}_+$ of all positive p-particle observables with the hyperplane

$$\varphi^0 \equiv \{\mathbf{D} \in \varphi;\ pr_0(\mathbf{D}) = 1\}.$$

The exposed points of $\mathfrak{F}^{[p]}$ are *either* of the form

$$L^p \circ i_p(D^p) \equiv \left(1, nL^1_p(D^p), \cdots, \binom{n}{m} L^m_p(D^p), \cdots, \binom{n}{p} D^p, 0, \cdots\right), \tag{3.20}$$

where D^p is exposed in $\mathfrak{F}^p_n$ for some $n \geq p$, *or* of the form

$$L^p \circ i_n(P^n) \equiv (1, nL^1_n(P^n), \cdots, \binom{n}{m} L^m_n(P^n), \cdots, P^n, 0, \cdots) \tag{3.21}$$

for some $n \leq p$, where P^n denotes a one-dimensional projector in $H^{n\wedge}$.

(iii) The expectation value of the observable $\mathbf{B}$ in the state $\mathbf{\Delta}$ is given by

$$(\mathbf{B} \mid \mathbf{\Delta}) = \sum_{m=0}^{p} \operatorname{Tr}(B^m \Delta^m).$$

Before we state the proof of this theorem, let us introduce the notion of a *special point.*

[7] R. Giles, unpublished.

Definition 3.3: A point of $\mathfrak{F}^{[p]}$ of type (3.20) or (3.21) is called a *special point*. The integer n occurring in Eqs. (3.20) and (3.21) is called *the order* of the special point. (Thus every special point of order $m \leq p$ is exposed in $\hat{\mathfrak{F}}$.)

Proof of Theorem 5:

(i) This is a mere repetition of the definition of a p-reducible system of particles.

(ii) Let us determine the polar cone of $\mathfrak{F}^{[p]}$ with respect to the dual pair $\langle \mathcal{O}_{(p)}, \mathcal{P}_{(p)} \rangle$. We have the following chain of equivalent statements:

$$\begin{aligned} \mathbf{B} \in \tilde{\mathfrak{F}}^{[p]} &\Leftrightarrow (\mathbf{B} \mid \mathbf{\Delta}) \geq 0, \qquad \forall \mathbf{\Delta} \in \mathfrak{F}^{[p]}, \\ &\Leftrightarrow (\mathbf{B} \mid L^p(\mathbf{D})) \geq 0, \quad \forall \mathbf{D} \in \mathfrak{F}, \\ &\Leftrightarrow (\Gamma_p(\mathbf{B}) \mid \mathbf{D}) \geq 0, \quad \forall \mathbf{D} \in \mathfrak{F}, \\ &\Leftrightarrow \Gamma_p(\mathbf{B}) \geq 0. \end{aligned}$$

L^p stands for $\Pi^p \circ L$, where Π^p is the projection defined by Eq. (2.21), and Γ_p denotes the restriction of Γ onto $\mathcal{O}_{(p)}$. Thus the following equality holds:

$$\tilde{\mathfrak{F}}^{(p)} = \mathcal{O}_{(p)+} \equiv \mathcal{O}_{(p)} \cap \mathcal{O}_+;$$

and the first part of the assertion is a consequence of the bipolar theorem (Theorem A6).

Assume $\mathbf{\Delta} \in \mathfrak{F}^{[p]}$ to be exposed. Then the full pre-image of $\mathbf{\Delta}$ in $\mathfrak{F}$ with respect to L^p is an exposed set of $\mathfrak{F}$. Indeed if

$$\{\mathbf{\Delta}\} = \{\mathbf{B}\}^\perp \cap \mathfrak{F}^{[p]}$$

for some $\mathbf{B} \in \mathcal{O}_{(p)+}$, then

$$(L^p)^{-1}(\mathbf{\Delta}) = \{\Gamma_p(\mathbf{B})\}^\perp \cap \mathfrak{F},$$

as one may easily confirm.

$\Gamma_p(\mathbf{B})$ defines a sequence of subspaces of the many-particle spaces $H^{m\wedge}$:

$$\mathbf{V} = (V_0, V_1, \cdots, V_m, \cdots),$$

where V_m denotes the null space of $\Gamma_p(\mathbf{B})^m$ in $H^{m\wedge}$. The pre-image $(L^p)^{-1}(\mathbf{\Delta})$ of $\mathbf{\Delta}$ consists of all states $\mathbf{D} \in \mathfrak{F}$ with the property that $\forall_m$, $R_{D^m} \subseteq V_m$ (cf. NRP Lemma 3).

Now let n be the smallest number m such that $V_m \neq \{0\}$:

$$n \equiv \min_m \{m \mid V_m \neq \{0\}\}.$$

We have to distinguish two cases:

(a) $n > p$. Choose $D^n \in \mathfrak{F}^n$ such that $R_{D^n} \subseteq V_n$. Then

$$\mathbf{\Delta} = L^p \circ i_n(D^n)$$

is of type (20) with $D^p \equiv L^p_n(D^n)$. Since in this case

$$\Gamma_p(B)^n = \Gamma^n_p\left(\sum_{k=0}^{p} \binom{n}{k} \Gamma^p_k(B^k)\right),$$

V_n is a p-subspace of $H^{n\wedge}$, and, from NRP Proposition 4, it follows that D^p is an exposed point of $\mathfrak{F}^p_n$.

(b) $n < p$. In this case let P^n be the projector onto a one-dimensional subspace of V_n. Then

$$\mathbf{\Delta} = L^p \circ i_n(P^n)$$

is indeed of type (21). Q.E.D.

Similar remarks as in NRP can be made at this stage. Thus, for instance, Theorem 5(ii) describes the projection cone of $\overline{\mathfrak{F}^{[p]}}$ from the center 0 as the intersection of an infinity of closed half-spaces. Notice that in the fermion case $\mathbf{B} = (1, I - 2P^1, 0, \cdots)$ belongs to $\tilde{\mathfrak{F}}^{[1]}$, where P^1 stands for any one-dimensional projector in H^1. Indeed we have

$$\Gamma_1(\mathbf{B}) = A_n + n\Gamma^n_1(I - 2P^1) \geq 0, \quad \forall n. \tag{3.22}$$

Therefore, as a necessary condition for representability in this case, we get

$$2 - 2 \operatorname{Tr}(P^1 D^1) \geq 0$$

or

$$\|D^1\| = \sup_{P^1 \in E^1} \operatorname{Tr}(P^1 D^1) \leq 1.$$

(The notation E^1 for the set of all one-dimensional projectors in H^1 is adopted as in NRP.) Thus a necessary condition that in the case of fermions a given potential one-state $(1, D^1, 0, \cdots)$ is actual can be expressed as $\|D^1\| \leq 1$. Later on (Theorem 7) we shall prove that this condition is also sufficient.

Finally, let us state the corresponding theorem to Proposition 2 of NRP:

Proposition 3 (Garrod and Percus[8]):

(i) Let β be an arbitrary complex number and B a bounded one-particle operator. Then the observable

$$\mathbf{C} \equiv (|\beta|^2, \beta B^* + \beta^* B + BB^*, A_2(B \otimes B^* + B^* \otimes B), 0, \cdots)$$

belongs to the dual cone of $\mathfrak{F}^{[2]}$.

(ii) A necessary condition that $\mathbf{D} = (1, D^1, D^2, 0, \cdots)$ is representable is that the sesquilinear form

$$H_{\mathbf{D}}(A, B) \equiv \operatorname{Tr}(AB^* D^1) + 2 \operatorname{Tr}((A \otimes B^*) D^2) - \operatorname{Tr}(AD^1) \operatorname{Tr}(B^* D^1) \tag{3.23}$$

defined on the space $B(H^1)$ of all bounded operators in one-particle space H^1 is positive semidefinite.

[8] C. Garrod and J. K. Percus, J. Math. Phys. **5**, 1756 (1964).

Proof:
(i)

$$pr_n(\Gamma_2(\mathbf{C})) = |\beta|^2 A_n + n\Gamma_1^n(\beta^*\beta + \beta B^* + BB^*) + \binom{n}{2}\Gamma_2^n(A_2(B \otimes B^* + B^* \otimes B)) = A_n\left(\beta + \sum_{i=1}^{n} B(i)\right)\left(\beta^* + \sum_{i=1}^{n} B^*(i)\right) \geq 0.$$

(ii) From (i) it follows that, for any complex number β, the expression

$$|\beta|^2 + \beta^* \operatorname{Tr}(D^1B) + \beta \operatorname{Tr}(D^1B^*) + \operatorname{Tr}(D^1BB^*) + 2 \operatorname{Tr}((B \otimes B^*)D^2) = |\beta + \operatorname{Tr}(BD^1)|^2 - |\operatorname{Tr}(BD^1)|^2 + \operatorname{Tr}(BB^*D^1) + 2 \operatorname{Tr}((B \otimes B^*)D^2)$$

is nonnegative. Choose $\beta = -\operatorname{Tr}(BD^1)$ to get the desired result. Q.E.D.

Remark: Theorem 5 states that every exposed point of $\mathfrak{F}^{[p]}$ is of special type. One could ask if, conversely, every special point of $\mathfrak{F}^{[p]}$ is exposed. The examples for $p = 1$ show that, in general, this is not the case. While in the fermion case every special point is also exposed (as we shall prove), in the boson case none of the special points with the exception of the point $(1, 0, \cdots)$ is exposed. In fact, in the case of bosons, $\mathfrak{F}^{[1]}$ is a convex cone with $(1, 0, \cdots)$ as its vertex (cf. Definition A7 and Theorem 7).

However, the question is not of great relevance since what we really are looking for is not the set of all exposed points of $\mathfrak{F}^{[p]}$ but rather a subset of points of $\mathfrak{F}^{[p]}$ which is first of all as "small" as possible and secondly has the property that its convex closure is all of $\overline{\mathfrak{F}^{[p]}}$. The following theorem, which is the analog of NRP Theorem 9, shows that the special points satisfy this last requirement.

Theorem 6: Let $W_1 \subset W_2 \subset \cdots \subset H^1$ be an increasing sequence of finite-dimensional subspaces of H^1 such that

$$\overline{\bigcup_{i=1}^{\infty} W_i} = H^1;$$

and for any finite-dimensional subspace W of H^1 and any natural number n, let $E_{(n)}^p[W]$ denote the set of all special points whose order is not greater than n and which belong to $\mathfrak{F}_{(p)}[W]$. Then

$$\overline{\mathfrak{F}^{[p]}} = \operatorname{conv} \overline{\bigcup_n \bigcup_k E_{(n)}^p[W_k]}. \tag{3.24}$$

If an element $\mathbf{D}^p \in \mathfrak{F}_{(p)}$ is said to be of *finite one-rank* iff $\mathbf{D}^p \in \mathfrak{F}_{(p)}[W]$ for some finite-dimensional subspace W of H^1, we can also formulate Theorem 6 as follows:

Corollary to Theorem 6: $\overline{\mathfrak{F}^{[p]}}$ is the closed convex hull of all special points of finite one-rank.

Proof: First of all, it follows from NRP Theorem 2(ii) and the continuity of L^p that

$$\overline{\mathfrak{F}^{[p]}} = \overline{L^p(\mathfrak{F})} = \overline{\bigcup_n \bigcup_k L^p(\mathfrak{F}_{(n)}[W_k])}.$$

Since $\mathfrak{F}_{(n)}[W_k]$ is a compact subset of $\varphi_{(n)}[W_k] \equiv \varphi[W_k] \cap \varphi_{(n)}$, the continuity of L^p implies that $L^p(\mathfrak{F}_{(n)}[W_k])$ is a compact subset of $\varphi_{(p)}[W_k]$. According to the theorem of Klee (cf. Theorem A8), $L^p(\mathfrak{F}_{(n)}[W_k])$ is the convex hull of its exposed points.

To complete the proof we have to show that every exposed point of $L^p(\mathfrak{F}_{(n)}[W])$ is of special type. Arguments similar to those used in the proof of Theorem 5 show that for any closed subspace W of H^1 the exposed points of $L^p(\mathfrak{F}_{(n)}[W])$ are either of type $L^p \circ i_m(D^m)$ $p < m \leq n$, where $L_m^p(D^m)$ is an exposed point of $\mathfrak{F}_m^p[W]$, or of type $L^p \circ i_m(P^m)$, $m \leq p$, where P^m is a projector onto a one-dimensional subspace of $W^{m\wedge}$. Since (cf. proof of Theorem 9 of NRP) any point which is exposed in $\mathfrak{F}_n^p[W]$ is also exposed in $\mathfrak{F}_n^p$, we conclude that every exposed point of $L^p(\mathfrak{F}_{(p)}[W])$ is either of type (20) or (21), i.e., special. Q.E.D.

Finally, let us make an application of this theorem to the case $p = 1$ and fermions. In this case the special points are of the form $\mathbf{\Delta} = (1, P_U, 0, \cdots)$, where U is a finite-dimensional subspace of H^1 and P_U the corresponding projector.

We first show that $\mathbf{\Delta}$ is exposed at $\mathfrak{F}^{[1]}$. Indeed, let $\mathbf{B}$ be the observable

$$\mathbf{B} = (1, I - (1 + (1/n))P_U, 0, \cdots).$$

Then

$$\Gamma(\mathbf{B})^m = (m + 1)A_m - (m/n)(n + 1)\Gamma_1^m(P_U).$$

The smallest eigenvalue λ_m of $\Gamma(\mathbf{B})^m$ is nondegenerate and it is given by

$$\lambda_m = \frac{n - m}{n}, \quad \text{for} \quad m \leq n,$$
$$= m - n, \quad \text{for} \quad m \geq n.$$

Hence $\Gamma(\mathbf{B}) \geq 0$. Moreover, for $m \neq n$, $\Gamma(\mathbf{B})^m$ is positive definite. On the other hand, $\Gamma(\mathbf{B})^n$ possesses the space generated by the "Slater determinant":

$$[U] = (n!)^{\frac{1}{2}} A_n(e_1 \otimes \cdots \otimes e_n),$$

($\{e_i\}_{i=1}^n$ = CONS in U) as its nullspace. (Notice that the Slater determinant $[U]$ is determined by the subspaces $[U]$ only up to a phase factor.) It follows that

$$(\mathbf{B} \mid \mathbf{\Delta}) = 0, \quad \mathbf{\Delta} \in \mathcal{I}^{[1]},$$

if and only if $\mathbf{\Delta} = L^1 \circ i_n(P_{[U]}) = (1, P_U, 0, \cdots)$.

In the case of bosons, the special points consist of integer multiples nP of one-dimensional projectors.

These results together with Theorem 6 immediately imply the following:

Theorem 7:

(i) Fermions: A potential one-state $(1, D, 0, \cdots)$ is an actual one-state if and only if $\|D\| \leq 1$. Or, in other terms,

$$\overline{\mathcal{I}^{[1]}} = \{\mathbf{\Delta} \in \mathcal{I}^0_{(1)}; \|pr_1(\mathbf{\Delta})\| \leq 1\}. \tag{3.25}$$

(ii) Bosons: Every potential one-state is also actual, i.e.,

$$\overline{\mathcal{I}^{[1]}} = \mathcal{I}^0_{(1)}. \tag{3.26}$$

Proof:

(ii) is an obvious consequence of Theorem 6 and some subsequent remarks.

(i) We denote by τ the set

$$\tau \equiv \{D \in \mathcal{K}^1_T; \|D\| \leq 1\}, \tag{3.27}$$

where similarly, as in NRP $\mathcal{K}^1_T$, stands for the set S^{1+}_T of all positive trace operators in one-particle space. Since τ is a closed convex set and contains all the finite-dimensional projectors, it follows from Theorem 6 that

$$\overline{\mathcal{I}^{[1]}} \subseteq \{\mathbf{\Delta} \in \varphi^+_{(1)}; pr_1(\mathbf{\Delta}) \in \tau\} \equiv \hat{\tau}. \tag{3.28}$$

Now for any finite-dimensional subspace $W \subset H^1$, we define

$$\tau[W] \equiv \tau \cap S^1[W] \quad \text{and} \quad \hat{\tau}[W] \equiv \hat{\tau} \cap \varphi_{(1)}[W].$$

Then $\tau[W]$ is a compact subset of $S^1[W]$. What are the extreme points of $\tau[W]$? We assert that they coincide with the projectors onto subspaces of W. Indeed, let $U \subseteq W$ be such a subspace and P_U the corresponding projector. Assume that $P_U = \alpha A + (1 - \alpha)B$, $0 < \alpha < 1$, and $A, B \in \tau[W]$. Since $U = \{R_A, R_B\}$, where $\{\ \}$ = linear hull, it follows that $R_A, R_B \subseteq U$. Furthermore, for any $x \in U$ with $\|x\| = 1$ we have

$$1 = \alpha(x \mid Ax) + (1 - \alpha)(x \mid Bx)$$

and hence

$$(x \mid Ax) = (x \mid Bx) = 1.$$

From this result one easily deduces $A = B = P_U$, which means that P_U is extreme in $\tau[W]$.

Conversely, let $A \in \tau[W]$; let

$$A = \sum_{i=1}^{r} \lambda_i P_i, \quad 1 \geq \lambda_1 \geq \lambda_2 \cdots \geq \lambda_{r-1} \geq \lambda_r > 0,$$

be its spectral decomposition, and assume $\lambda_r < 1$. Furthermore, define

$$P \equiv \sum_{i=1}^{r-1} P_i$$

and let

$$0 < \epsilon \leq \min [\lambda_{r-1}, 1 - \lambda_r].$$

If we define

$$B \equiv \frac{1}{1 - \epsilon}\left[\sum_{i=1}^{r-1}(\lambda_i - \epsilon)P_i + \lambda_r P_r\right],$$

then $B \in \tau[W]$ and we obtain

$$A = \epsilon P + (1 - \epsilon)B,$$

an equation indicating that A is not extreme. Thus the extreme points of $\hat{\tau}[W]$ coincide with certain (exposed) points of $\mathcal{I}^{[1]}$ and the theorem of Krein–Milman (Theorem A9) (which in this special case can be regarded as a consequence of the theorem of Klee) implies that, for every finite-dimensional subspace $W \subseteq H^1$,

$$\hat{\tau}[W] \subseteq \mathcal{I}^{[1]}.$$

Now let $\mathbf{\Delta} = (1, D, 0, \cdots) \in \hat{\tau}$. Then, according to the spectral theorem, there exists a sequence $D_n \in \tau$ such that

$$\lim_{n \to \infty} |D_n - D| = 0$$

and $W_n \equiv R_{D_n}$ defines an increasing sequence of finite-dimensional subspaces of H^1 such that

$$\overline{\bigcup_n W_n} = R_D.$$

If we put

$$\mathbf{\Delta}_n = (1, D_n, 0, \cdots),$$

it follows that

$$\mathbf{\Delta}_n \in \hat{\tau}[W_n]$$

and

$$\lim_{n \to \infty} p_k(\mathbf{\Delta}_n - \mathbf{\Delta}) = k \lim_{n \to \infty} |D_n - D| = 0.$$

Hence

$$\mathbf{\Delta} \in \overline{\bigcup_n \hat{\tau}[W_n]} \subseteq \overline{\mathcal{I}^{[1]}}.$$

It follows that $\hat{\tau} \subseteq \overline{\mathcal{I}^{[1]}}$, and, combining this with the result expressed in Eq. (3.28), we finally obtain $\hat{\tau} = \overline{\mathcal{I}^{[1]}}$. Q.E.D.

4. THE BCS ENSEMBLE

In this section we give a description of the BCS ensemble inside the mathematical framework presented

so far. First of all we turn the set

$$\mathcal{R} = \bigcup_{m=0}^{\infty} H^{2m\wedge}$$

into a commutative monoid (cf. Chevalley[9]) by introducing the product

$$x \wedge y = A_{2(m+n)}(x \otimes y), \quad x \in H^{2m\wedge}, \quad y \in H^{2n\wedge}. \tag{4.1}$$

If $\alpha H^0 \equiv \mathbf{C}$, then, by the usual convention, $\alpha \otimes y = \alpha \cdot y$ and hence $\alpha \wedge y = \alpha \cdot y$. So that the number 1 (identity in $\mathbf{C}$) becomes the identity in $\mathcal{R}$. It is easy to recognize $\mathcal{R}$ as a commutative monoid when equipped with the product (1). Moreover, the norms in the different spaces $H^{2m\wedge}$ induce on $\mathcal{R}$ a positive submultiplicative functional:

$$\|y \wedge z\| \leq \|y\| \, \|z\|. \tag{4.2}$$

In a similar fashion one can turn the set

$$\mathcal{R}' \equiv \bigcup_{m=0}^{\infty} S_T^{2m}$$

into a commutative monoid by introducing the product

$$C \wedge D = A_{2(m+n)}(C \otimes D)A_{2(m+n)}, \qquad C \in S_T^{2m\wedge}, \quad D \in S_T^{2n\wedge}. \tag{4.3}$$

Again a similar inequality

$$|C \wedge D| \leq |C| \, |D| \tag{4.4}$$

holds, so that the norms on the different spaces S_T^{2m} induce a submultiplicative functional on $\mathcal{R}'$. The map which attaches to every element of $\mathcal{R}(\mathcal{R}')$ its order m is a homomorphism of $\mathcal{R}(\mathcal{R}')$ into the monoid consisting of the set of all natural numbers extended by 0, with the addition as the law of composition.

For every integer m we introduce a mapping

$$D\colon H^{2m\wedge} \times H^{2m\wedge} \to S_T^{2m}.$$

D is defined by the equation

$$D(x, y)z = \tfrac{1}{2}[(x \mid z)y + (y \mid z)x], \quad x, y, z \in H^{2m\wedge}. \tag{4.5}$$

$D(x, y)$ has a rank of at most two, and hence it certainly belongs to S_T^{2m}. For $D(x, x)$ we write $D(x)$. $D(x, y)$ has essentially the same properties as a real-valued symmetric bilinear form. Let us enumerate some of them:

Proposition 4:

(i) $D(x, y) = D(y, x), \quad x, y \in H^{2m\wedge}$;

(ii) $|D(x, y)| = [\|x\|^2 \, \|y\|^2 - \operatorname{Im}(x \mid y)^2]^{\frac{1}{2}} \leq \|x\| \, \|y\|$;

(iii) $D(x, ix) = 0$;

(iv) $D(\alpha, \beta) = \operatorname{Re}(\bar{\alpha}\beta), \quad \alpha, \beta \in H^0 \equiv \mathbf{C}$;

(v) $D(\alpha x) = \alpha^2 D(x)$;

(vi) $D(\alpha x, y) = D(x, \alpha y) = \alpha D(x, y), \quad \alpha \in \mathbf{R}, \quad x, y \in H^{2m\wedge}$;

(vii) $D(x + y, z) = D(x, z) + D(y, z)$;

(viii) $D(x, y) \wedge D(t, z) = \tfrac{1}{2}[D(x \wedge t, y \wedge z) + D(x \wedge z, y \wedge t)]$, $x, y \in H^{2m\wedge}$ $t, z \in H^{2m'\wedge}$;

(ix) $D(x) \wedge D(y) = D(x \wedge y)$;

(x) $D(x) = \|x\|^2 P_x$ (P_x = projector onto the ray generated by x).

Remarks:

(i) Physically, $D(x, y)$ has the meaning of the real part of the transition density operator between the states x and y.

(ii) Statement (ii) of Proposition 4 expresses the fact that $D(x, y)$ depends continuously on both arguments.

(iii) Statements (iv) and (ix) of Proposition 4 show that the map $x \to D(x)$ is a homomorphism of the monoid $\mathcal{R}$ into the monoid $\mathcal{R}'$. However, it is not an isomorphism since $D(x) = D(y)$ if and only if $x = e^{i\theta}y$ for some real θ.

Proof of Proposition 4: The only statements which are not immediately obvious are statements (ii) and (viii). Statement (ix) is an easy consequence of statement (viii). To prove statement (ii) we have to find the eigenvalues of the matrix

$$\begin{pmatrix} (y \mid x) & \|x\|^2 \\ \|y\|^2 & (x \mid y) \end{pmatrix}.$$

These are given by the expression

$$\mu_{1,2} = [\operatorname{Re}(x \mid y)]^2 \pm (\|x\|^2 \, \|y\|^2 - [\operatorname{Im}(x \mid y)]^2)^{\frac{1}{2}},$$

which is of the general form

$$\mu_{1,2} = a \pm b \quad \text{with} \quad b \geq |a|.$$

Hence

$$|D(x, y)| = \tfrac{1}{2}(|\mu_1| + |\mu_2|) = b = (\|x\|^2 \, \|y\|^2 - [\operatorname{Im}(x \mid y)]^2)^{\frac{1}{2}} \leq \|x\| \, \|y\|.$$

For (viii) let $M \equiv 2(m + m')$. Then, for any vector in H^M of the form

$$u \equiv v_1 \otimes \cdots \otimes v_M,$$

[9] C. Chevalley, *Fundamental Concepts of Algebra* (Academic Press Inc., New York, 1956), p. 3.

we have

$$
\begin{aligned}
&[D(x, y) \wedge D(t, z)]u \\
&= \sum_{s\in\Pi_M} (-1)^{\sigma(s)} A_M (D(x, y) \otimes D(t, z)) v_{\alpha_1} \otimes \cdots \otimes v_{\alpha_M} \\
&= \tfrac{1}{4} \sum_{s\in\Pi_M} (-1)^{\sigma(s)} [(x \mid v_{\alpha_1} \otimes \cdots \otimes v_{\alpha_{2m}}) y \\
&\quad + (y \mid v_{\alpha_1} \otimes \cdots \otimes v_{\alpha_{2m}}) x] \\
&\quad \times [(t \mid v_{\alpha_{2m+1}} \otimes \cdots \otimes v_{\alpha_M}) z \\
&\quad + (z \mid v_{\alpha_{2m+1}} \otimes \cdots \otimes v_{\alpha_M}) t] \\
&= \tfrac{1}{4} \sum_{s\in\Pi_M} (-1)^{\sigma(s)} [(x \otimes t \mid v_{\alpha_1} \otimes \cdots \otimes v_{\alpha_M}) y \wedge z \\
&\quad + (x \otimes z \mid v_{\alpha_1} \otimes \cdots \otimes v_{\alpha_M}) y \wedge t \\
&\quad + (y \otimes t \mid v_{\alpha_1} \otimes \cdots \otimes v_{\alpha_M}) x \wedge t \\
&\quad + (y \otimes z \mid v_{\alpha_1} \otimes \cdots \otimes v_{\alpha_M}) x \wedge t] \\
&= \tfrac{1}{2} (D(x \wedge t \mid y \wedge z) + D(x \wedge z \mid y \wedge t)) u.
\end{aligned}
$$

In this sequence of equations Π_M denotes the symmetric group of M objects, $s \equiv \binom{1 \cdots M}{\alpha_1 \cdots \alpha_M}$ represents a particular element of Π_M, and $\sigma(s)$ stands for its signature. Since the set of all elements of the form (6) (i.e., the set of all "M vectors") contains a CONS of H^M, the asserted statement (viii) follows.

In the sequel we shall often use shorthand notations like

$$x^m \equiv \underbrace{x \wedge \cdots \wedge x}_{m},$$

$$xy \equiv x \wedge y,$$

and

$$D^m(x, y) \equiv \underbrace{D(x, y) \wedge \cdots \wedge D(x, y)}_{m}.$$

Proposition 5 (cf. Coleman[10], Theorem 10.2; Zumino[11] and Blatt[12]): Let $g \in H^{2\wedge}$ and $D^1(g) \equiv L_2^1(D(g))$; then there exists a decomposition of the closure W of the range of $D^1(g)$ into a direct sum of mutually orthogonal, two-dimensional subspaces:

$$W = \bigoplus_{\sigma\in J} W_\sigma, \quad \dim W_\sigma = 2, \tag{4.6}$$

such that W_σ is an eigenspace of $D^1(g)$ (to the eigenvalue λ_σ) and g can be represented as

$$g = 2^{\frac{1}{2}} \sum_{\sigma\in J} \lambda_\sigma^{\frac{1}{2}} [\sigma], \tag{4.7}$$

where J should be identified with the (well-ordered) set of pairs of successive natural numbers:

$$J \equiv \{(12), (34), \cdots, (2k-1, 2k) \cdots\} \tag{4.8}$$

and $[\sigma]$ stands for the normalized Slater determinant generated by a particular orthonormal basis e_{2k-1}, e_{2k} of W_σ:

$$[\sigma] \equiv 2^{\frac{1}{2}} (e_{2k-1} \wedge e_{2k}).$$

(The superscript $\frac{1}{2}$ on a positive number denotes the positive square root throughout this paper.)

For the proof of this proposition we need the notion of a conjugation in a Hilbert space.

Definition 4.1: A conjugation $x \to \bar{x}$ in a (complex) Hilbert space H is an antilinear, antiunitary involution, i.e., a mapping satisfying the axioms

(i) $\overline{x + y} = \bar{x} + \bar{y}, \quad x, y \in H,$
(ii) $\overline{\alpha x} = \alpha^* \bar{x}, \quad \alpha \in \mathbf{C}, \quad x \in H,$
(iii) $(\bar{x} \mid \bar{y}) = \overline{(x \mid y)}$, antiunitary character,
(iv) $\bar{\bar{x}} = x$, involutionary character.

To see that every Hilbert space allows a conjugation, let $\{e_i\}$ be a CONS (complete orthonormal set) in H and define

$$\bar{x} = \sum_{i=1}^{\infty} \alpha_i^* e_i,$$

whenever

$$x = \sum_{i=1}^{\infty} \alpha_i e_i.$$

Let H be a Hilbert space with conjugation and $A \in B(H)$ [$B(H)$ = set of all bounded linear operators in H]. Define

$$\bar{A}x = \overline{A\bar{x}}, \quad x \in H.$$

Then obviously $\bar{A} \in B(H)$ and

(i) $\overline{A + B} = \bar{A} + \bar{B},$
(ii) $\overline{\alpha A} = \alpha^* \bar{A},$
(iii) $\overline{AB} = \bar{A}\,\bar{B},$
(iv) $\bar{\bar{A}} = A,$
(v) $\|\bar{A}\| = \|A\|.$

Thus a conjugation in H induces an antilinear, involutionary isometry on $B(H)$. Now we are ready to prove Proposition 5.

Proof of Proposition 5: Let $x \to \bar{x}$ be a conjugation in H^1. Then for every $y \in H^1$ the expression $(g \mid \bar{y} \otimes x)$ represents a continuous linear functional in H^1 and, according to the Riesz representation theorem, there exists $\hat{y}$ in H^1 such that

$$(g \mid \bar{y} \otimes x) = (\hat{y} \mid x), \quad \forall x \in H^1.$$

It is easy to verify that the map

$$G: y \to \hat{y} \tag{4.9}$$

[10] A. J. Coleman, Rev. Mod. Phys. **35**, 668 (1963).
[11] B. Zumino, J. Math. Phys. **3**, 1055 (1962).
[12] J. M. Blatt, *Theory of Superconductivity* (Academic Press Inc., New York, 1964).

is a linear operator of Hilbert–Schmidt class [$G \in \mathcal{K}(H^1)$]. Indeed for any CONS $\{e_i\}$ and $\{f_k\}$ in H^1, we have

$$\|G\|^2 = \mathrm{Tr}\,(G^*G) = \sum_{i=1}^{\infty}(Ge_i \mid Ge_i)$$
$$= \sum_{i=1}^{\infty}(Ge_i \mid f_k)(f_k \mid Ge_i) = \sum_{i=1}^{\infty}\sum_{k=1}^{\infty}(g \mid \bar{e}_i \otimes f_k)$$
$$= (\bar{e}_i \otimes f_k \mid g) = \|g\|^2. \tag{4.10}$$

Thus $g \to G$ defines a linear isometry from $H^{2\wedge}$ into $\mathcal{K}(H^1)$. Since

$$(Gx \mid y) = (g \mid \bar{x} \otimes y) = -(g \mid y \otimes \bar{x}) = -(G\bar{y} \mid \bar{x})$$
$$= (x \mid \bar{G}\bar{y}), \quad \forall x, y \in H^1,$$

it follows that

$$G^* = -\bar{G}. \tag{4.11}$$

Let $\{e_i\}$ be a CONS in H^1. Then, using (4.21), we obtain

$$(e_i \mid GG^*e_j) = (\bar{G}e_i \mid \bar{G}e_j) = (G\bar{e}_j \mid G\bar{e}_i)$$
$$= \sum_k (G\bar{e}_j \mid f_k)(f_k \mid G\bar{e}_i)$$
$$= \sum_k (g \mid e_j \otimes f_k)(e_i \otimes f_k \mid g)$$
$$= \sum_k (e_i \otimes f_k \mid D(g)(e_j \otimes f_k))$$
$$= (e_i \mid D^1(g)e_j).$$

Therefore

$$D^1(g) = GG^* = -G\bar{G}. \tag{4.12}$$

Moreover, the expression G^*G can be written as

$$G^*G = -\bar{G}G = \bar{D}^1(g).$$

It follows that the polar decompositions of G and G^* (cf. Schatten,[6] p. 4) take the form

$$G = U(\bar{D}^1(g))^{\frac{1}{2}}, \tag{4.13}$$

$$G^* = U^*(D^1(g))^{\frac{1}{2}}, \tag{4.14}$$

where U is an isometry of the closure W of the range of $D^1(g)$ onto itself. Substituting (4.13) and (4.14) into (4.11), we obtain

$$U^*[D^1(g)]^{\frac{1}{2}} = -\bar{U}[D^1(g)]^{\frac{1}{2}},$$

an equation which implies

$$U^* = -\bar{U}. \tag{4.15}$$

Hence, if $x \in W$, $\|x\| = 1$, then $\|U\bar{x}\| = 1$ and

$$(U\bar{x} \mid x) = -(\bar{x} \mid \bar{U}x) = -(U\bar{x} \mid x),$$

i.e., $U\bar{x} \perp x$. Moreover,

$$U\overline{U\bar{x}} = U\bar{U}x = -UU^*x = -x. \tag{4.16}$$

Let

$$W = \bigoplus_{i=1}^{\infty} W_i \tag{4.17}$$

be the decomposition of W into a direct sum of eigenspaces of $D^1(g)$ corresponding to *different* eigenvalues λ_i. Since $D^1(g)$ is completely continuous, the dimensions of the eigenspaces W_i are finite. W_i is invariant under the operation $x \to U\bar{x}$. Indeed, let $x \in W_i$. Then

$$D^1(g)U\bar{x} = \overline{\bar{D}^1(g)\bar{U}x} = -\overline{\bar{D}^1(g)U^*x}$$
$$= -\overline{(\bar{D}^1(g))^{\frac{1}{2}}G^*x} = -\overline{U^*GG^*x} = \overline{\bar{U}D^1(g)x}$$
$$= \lambda_i\overline{\bar{U}x} = \lambda_i U\bar{x}.$$

Moreover, since $U\bar{x} \perp x$ and $U\overline{U\bar{x}} = -x$, it follows that W_i can be decomposed into a direct sum,

$$W_i = \bigoplus_{\sigma \in J_i} W_\sigma, \tag{4.18}$$

of mutually orthogonal two-dimensional subspaces, each of which is invariant under the operation $x \to U\bar{x}$, a property which we express by the equation

$$U\overline{W}_\sigma = W_\sigma. \tag{4.19}$$

Combining (4.17) and (4.18), we get a representation of W as a direct sum of two-dimensional eigenspaces of $D^1(g)$:

$$W = \bigoplus_{\sigma \in J} W_\sigma, \quad J = \bigcup_i J_i, \tag{4.20}$$

each of which has the property expressed in Eq. (4.19).

Now let us choose in each two-dimensional subspace $W_{(2k-1,2k)}$, a normalized vector e_{2k-1}, and define

$$e_{2k} \equiv U\bar{e}_{2k-1}. \tag{4.21}$$

Then $e_{2k} \in W_{(2k-1,2k)}$, $\|e_{2k}\| = 1$, $e_{2k} \perp e_{2k-1}$, and, as a consequence of (4.12),

$$U\bar{e}_{2k} = -e_{2k-1}.$$

Finally, using the polar decomposition (4.13) of G, we obtain

$$(e_i \otimes e_j \mid g) = (e_j \mid Ge_i) = \lambda_\sigma^{\frac{1}{2}}(e_j \mid Ue_i) = 0$$
$$\text{if} \quad i \in \sigma, \quad j \in \tau, \quad \sigma \neq \tau,$$

and

$$(e_{2k-1} \otimes e_{2k} \mid g) = -(e_{2k} \otimes e_{2k-1} \mid g)$$
$$= (e_{2k} \mid Ge_{2k-1}) = \lambda_\sigma^{\frac{1}{2}}.$$

It follows that g can be written as

$$g = 2^{\frac{1}{2}}\sum_{\sigma \in J}\lambda_\sigma^{\frac{1}{2}}[\sigma],$$

where $[\sigma]$ denotes the normalized Slater determinant generated by the pair (e_{2k-1}, e_{2k}) of vectors. Q.E.D.

Next we are going to define a BCS ensemble. For this purpose we shall consider a mapping

$$\beta: H^{2\wedge} \to \mathfrak{F}. \tag{4.22}$$

However, let us first define a mapping

$$\alpha: H^{2\wedge} \to \mathcal{H}(\mathbf{C}),$$

where $\mathcal{H}(\mathbf{C})$ stands for the set of all holomorphic functions on the finite complex plane (cf. Sec. 3 of the Appendix). Let $\{\lambda_\sigma\}$ denote the sequence of eigenvalues of $D^1(g)$, corresponding to the subspaces $\{W_\sigma\}$ occurring in the decomposition (4.7) of the closure W of the range of $D^1(g)$. Then define

$$\alpha(g) \equiv f(t) \equiv \prod_{\sigma \in J} (1 + \lambda_\sigma t). \tag{4.23}$$

Since $\sum_\sigma |\lambda_\sigma| < \infty$, the infinite product in the equation of definition of f is normally convergent on every compact subset of the complex plane (cf. definitions A15 and A13). Therefore f belongs indeed to $\mathcal{H}(\mathbf{C})$. Put $C \equiv f(1) \geq 1$ and let

$$f(t) = \sum_{m=0}^{\infty} a_m t^m, \quad a_m \equiv \frac{1}{m!} f^{(m)}(0), \tag{4.24}$$

be the representation of f as a power series with center 0.

For every $g \in H^{2\wedge}$ let P_{g^m} be the projector generated by g^m. (Throughout this paper we are keeping the convention that the projector onto the trivial subspace is identified with the zero operator.) Now the mapping β is defined by

$$\begin{aligned} \beta(g)^0 &= C^{-1}, \\ \beta(g)^{2m} &= C^{-1} a_m P_{g^m}, \quad m = 1, 2 \cdots, \\ \beta(g)^{2m-1} &= 0, \qquad\quad m = 1, 2 \cdots. \end{aligned} \tag{4.25}$$

Since

$$p_k(\beta(g)) = C^{-1} \sum_{m=0}^{\infty} k^{2m} a_m = C^{-1} f(k^2) < \infty \tag{4.26}$$

and

$$\sum_{n=0}^{\infty} \mathrm{Tr}\,(\beta(g)^n) = C^{-1} \sum_{m=0}^{\infty} a_m = 1,$$

we have indeed $\beta(g) \in \mathfrak{F}$.

Definition 4.2: The state $\beta(g)$ of the many-particle system corresponding to the "geminal" g (= point of $H^{2\wedge}$) is called the BCS ensemble associated with g.

Proposition 5:

(i) The mapping $\alpha: H^{2\wedge} \to \mathcal{H}(\mathbf{C})$ is continuous.

(ii) The mapping $\beta: H^{2\wedge} \to \mathfrak{F}$ is continuous, and therefore the "BCS manifold" is connected.

Proof: The proof makes use of a series of lemmas.

Lemma 5: Let a_m^p be the value of the elementary function of p variables and of order m at the point $(\lambda_{12}, \cdots, \lambda_{(2p-1,2p)})$. Then

$$a_m = \lim_{p \to \infty} a_m^p \equiv \sum_{\sigma_1 < \cdots < \sigma_m} \lambda_{\sigma_1} \cdots \lambda_{\sigma_m}. \tag{4.27}$$

Proof of Lemma 5: Define

$$g_p(t) \equiv \prod_{k=1}^{p} (1 + t\lambda_{(2k-1,2k)}).$$

Then $g_p \to f$ in $\mathcal{H}(\mathbf{C})$ [cf. Theorem A11 (i)]. Since the mapping $f \to f'$ is continuous in $\mathcal{H}(\mathbf{C})$ [cf. Theorem A10 (iii)], it follows that

$$g_p^{(m)} \to f^{(m)} \quad \text{in } \mathcal{H}(\mathbf{C}).$$

In particular, we obtain

$$\lim_{p \to \infty} a_m^p = \frac{1}{m!} \lim_{p \to \infty} g_p^{(m)}(0) = \frac{1}{m!} f^{(m)}(0) = a_m. \quad \text{Q.E.D.}$$

Lemma 6: Let

$$c_m \equiv [2^{2m}(m!)^2/2m!]. \tag{4.28}$$

Then

(i) $\beta^{2m}(g) = c_m^{-1} C^{-1} D^m(g), \quad m = 1, 2, \cdots,$

(ii) $|D^m(g)| = c_m a_m,$

(iii) $1 \leq c_m \leq 2^m.$

Proof: From Proposition 4(ix) and 4(x) we obtain the result

$$D^m(g) = D(g^m) = \|g^m\|^2 P_{g^m}. \tag{4.29}$$

But, on the other hand, we have

$$g^m = 2^m \sum_{\sigma_1 \cdots \sigma_m} \lambda_{\sigma_1}^{\frac{1}{2}} \cdots \lambda_{\sigma_m}^{\frac{1}{2}} [e_{2k_1-1} \wedge e_{2k_1} \cdots e_{2k_m-1} \wedge e_{2k_m}].$$

Or, if we write $[\sigma_1 \cdots \sigma_m]$ for the normalized Slater determinant generated by $e_{2k_1-1} \cdots e_{2k_m}$,

$$g^m = \frac{2^m m!}{(2m!)^{\frac{1}{2}}} \sum_{\sigma_1 < \cdots < \sigma_m} \lambda_{\sigma_1}^{\frac{1}{2}} \cdots \lambda_{\sigma_m}^{\frac{1}{2}} [\sigma_1 \cdots \sigma_m]. \tag{4.30}$$

Hence

$$\|g^m\|^2 = c_m \cdot a_m \tag{4.31}$$

with c_m defined in Eq. (4.28). Combination of (4.31) with (4.29) yields

$$D^m(g) = c_m a_m P_{g^m}, \tag{4.32}$$

an equation which immediately implies Lemma 6(i) and (ii).

The result (iii) is easily proved by induction on m using the recursion formula

$$c_{m+1} = (2m + 2/2m + 1) c_m. \quad \text{Q.E.D.}$$

The following lemma generalizes and sharpens Lemma 4 of NRP.

Lemma 7: Let H be a Hilbert space and let $x,y \in H$. Let D_x and D_y be the operators in H defined by

$$D_x z = (x \mid z)x, \quad z \in H,$$
$$D_y z = (y \mid z)y.$$

Then

$$|D_x - D_y| \leq \|x + y\| \, \|x - y\|.$$

Proof: We have to find the eigenvalues of the matrix

$$\begin{pmatrix} \|x\|^2 - (y \mid x) \\ (x \mid y) - \|y\|^2 \end{pmatrix}.$$

These are given by the expression

$$\mu_{1,2} = \tfrac{1}{2}(\|x\|^2 - \|y\|^2) \pm [\tfrac{1}{4}(\|x\|^2 - \|y\|^2)^2 + \|x\|^2 \, \|x\|^2 - |(x \mid y)|^2]^{\frac{1}{2}},$$

which is of the general form

$$\mu_{1,2} = a \pm b \quad \text{with} \quad b \geq |a|.$$

Hence

$$\begin{aligned} |D_x - D_y| = |\mu_1| + |\mu_2| = 2b &= [(\|x\|^2 + \|y\|^2)^2 - 4\,|(x \mid y)|^2]^{\frac{1}{2}} \leq [(\|x\|^2 + \|y\|^2)^2 - 4[\operatorname{Re}(x \mid y)]^2]^{\frac{1}{2}} \\ &= [\|x\|^2 + \|y\|^2 - 2 \operatorname{Re}(z \mid y)]^{\frac{1}{2}} \times [\|x\|^2 + \|y\|^2 + 2 \operatorname{Re}(x \mid y)]^{\frac{1}{2}} \\ &= \|x - y\| \, \|x + y\|. \end{aligned}$$

Q.E.D.

Now let $g, \hat{g} \in H^{2\wedge}$ and let $f \equiv \alpha(g)$, $\hat{f} \equiv \alpha(\hat{g})$. Then one can easily verify that

$$D^m(g) - D^m(\hat{g}) = \sum_{i=0}^{m-1} [D^{(m-1-i)}(g) \wedge (D(g) - D(\hat{g})) \wedge D^{(i)}(\hat{g})].$$

It is a consequence of Lemma 6 that

$$|D^m(g) - D^m(\hat{g})| \leq 2^{m-1} b_{m-1} \, |D(g) - D(\hat{g})|, \quad (4.33)$$

where

$$b_m \equiv \sum_{i=1}^{m} a_{m-i} \hat{a}_i$$

is the coefficient of t^m in

$$f(t)\hat{f}(t) = \sum_{m=0}^{\infty} b_m t^m. \quad (4.34)$$

Now let $K \subseteq \mathbf{C}$ be a compact subset and

$$r = \sup_{t \in K} |t|.$$

Then, using Lemma 6 and Eq. (4.33), we obtain

$$\begin{aligned} \sup_{t \in K} |f - \hat{f}| &\leq \sum_{m=1}^{\infty} |a_m - \hat{a}_m| \, r^m \\ &\leq \sum_{m=1}^{\infty} c_m^{-1} \, \big| |D^m(g)| - |D^m(\hat{g})| \big| \, r^m \\ &\leq \sum_{m=1}^{\infty} |D^m(g) - D^m(\hat{g})| \, r^m \\ &\leq r \cdot f(2r)\hat{f}(2r) \, |D(g) - D(\hat{g})|. \end{aligned}$$

Finally, the application of Lemma 7 leads to the inequality

$$\sup_k |f - \hat{f}| \leq r f(2r)\hat{f}(2r) \, \|g + \hat{g}\| \, \|g - \hat{g}\|, \quad (4.35)$$

which expresses the asserted continuity of α. Moreover, from the estimation

$$\begin{aligned} p_k(\beta(g) - \beta(\hat{g})) &= \sum_{m=0}^{\infty} k^{2m} c_m^{-1} \, |C^{-1} D^m(g) - \hat{C}^{-1} D^m(\hat{g})| \\ &\leq \sum_{m=0}^{\infty} k^{2m} c_m^{-1} \, |C^{-1} D^m(g) - \hat{C}^{-1} D^m(g)| \\ &\quad + \sum_{m=0}^{\infty} k^{2m} \, |\hat{C}^{-1} D^m(g) - \hat{C}^{-1} D^m(\hat{g})| \\ &\leq \sum_{m=0}^{\infty} k^{2m} a_m \, |C^{-1} - \hat{C}^{-1}| \\ &\quad + k^2 \hat{C}^{-1} \sum_{m=0}^{\infty} (2k^2)^m b_m \, |D(g) - D(\hat{g})| \\ &\leq [C^{-1}\hat{C}^{-1} f(k^2) f(2) \hat{f}(2) + k^2 \hat{C}^{-1} f(2k^2) \hat{f}(2k^2)] \times \|g + \hat{g}\| \, \|g - \hat{g}\|, \end{aligned}$$

it follows that β is continuous. Q.E.D.

Next we extend the mapping β from $H^{2\wedge}$ onto a certain subset

$$\mathfrak{G} \subseteq H^{2\wedge} x \mathfrak{L}_e,$$

where $\mathfrak{L}_e$ stands for the (partially ordered) set of all subspaces of H^1 whose dimension is finite and even. $\mathfrak{G}$ is defined by

$$\mathfrak{G} \equiv \{(g, U) : U \perp R_{D^1(g)}\}.$$

With each element $U \in \mathfrak{L}_e$, we associate a normalized "Slater determinant" $[U]$. We put $[0] = 1$. Obviously $[U] \in H^{2p\wedge}$ if and only if $\dim U = 2p$. The extension of β is defined by

$$\begin{aligned} \beta(g, U)^n &= 0, \\ \beta(g, U)^{2p} &= C^{-1} P_{[U]}, \\ \beta(g, U)^{2(m+p)} &= C^{-1} a_m P_{g^m \wedge [U]}, \\ \beta(g, U)^{2(m+p)-1} &= 0, \end{aligned} \quad (4.36)$$

where $P_{g^m \wedge [U]}$ stands for the projector onto the ray generated by $g^m \wedge [U]$ and $2p \equiv \dim U$.

Definition 4.3: A state of the many-body system of the form $\beta(g, U)$, $U \neq 0$, is called a generalized BCS ensemble.

Proposition 6: Each generalized BCS ensemble $\beta(g, U)$, $U \neq 0$, is an adherence point of the set $\beta(H^{2\wedge}, 0) \equiv \beta(H^{2\wedge})$ of all (ordinary) BCS ensembles.

Proof: Let J be the set of all pairs of successive natural numbers as defined by Eq. (4.8) and let $J_0 \equiv \{(12), \cdots, (2p-1, 2p)\}$. Moreover, denote by J_1 the complement of J_0 in J. Now let $\beta(g, U)$ be a generalized BCS ensemble with $\dim U = 2p \neq 0$ and let

$$g = 2^{\frac{1}{2}} \sum_{\sigma \in J_1} \lambda_\sigma^{\frac{1}{2}} [\sigma]$$

be the representation (2) of g. Let $e_1, \cdots, e_{2p}$ be an orthonormal basis of U and let

$$h = 2^{\frac{1}{2}} \sum_{\tau \in J_0} [\tau],$$

where $[\tau]$ stands for $2^{\frac{1}{2}}(e_{2j-1} \wedge e_{2j})$ if $\tau = (2j-1, 2j)$.

Let $\{\mu_n\}$ be a sequence of positive numbers such that

$$\lim_{n \to \infty} \mu_n = \infty$$

and define

$$h_n = \mu_n h + g.$$

Then h_n is a sequence of "geminals" with the property

$$R_{D^1(h_n)} = R_{D(g)} \oplus U \quad \text{for all } n.$$

We want to show that $\beta(h_n) \to \beta(g, U)$ in φ.

Lemma 8: Denote by $\mathbf{D}(h_n)$ the element in φ with the components 0 at the odd entries and the components $c_m^{-1} D^m(h_n)$ at the even entries $2m$. Then $\mathbf{D}(h_n)$ can be represented as a polynomial of degree $2p$ in $\mu_n^{\frac{1}{2}}$:

$$\mathbf{D}(h_n) = \mathbf{K}_{2p}(\mu_n^{\frac{1}{2}})^{2p} + \mathbf{K}_{2p-1}(\mu_n^{\frac{1}{2}})^{2p-1} + \cdots + \mathbf{K}_0, \tag{4.37}$$

whose coefficients $\mathbf{K}_l$ belong to φ. In particular,

$$K_{2p}^{2m} = 0 \quad \text{for} \quad m < p, \tag{4.38}$$

$$K_{2p}^{2m} = \binom{m}{p}^2 c_m^{-1} c_p (P_{[U]} \wedge D^{m-p})(g). \tag{4.39}$$

$P_{[U]} \in S_T^{2p}$ denotes the projector onto the ray generated by $[U]$.

Proof of Lemma 8: From Proposition 4 we deduce for $m \geq p$

$$\begin{aligned} D^m(h_n) = D(h_n^m) &= D\left(\sum_{i=0}^{p} \binom{m}{i} (\mu_n^{\frac{1}{2}})^i h^i g^{m-i}\right) \\ &= \sum_{i=0}^{p} \binom{m}{i}^2 \mu_n^i D(h^i g^{m-i}) \\ &\quad + 2 \sum_{j=1}^{p} \sum_{i=0}^{j-1} (\mu_n^{\frac{1}{2}})^{i+j} \binom{m}{i} \binom{m}{j} D(h^i g^{m-i}, h^j g^{m-j}) \\ &= \sum_{i=1}^{p} \binom{m}{i}^2 \mu_n^i D(h^i g^{m-i}) \\ &\quad + 2 \sum_{l=1}^{2p-1} (\mu_n^{\frac{1}{2}}) \sum_{i=\max[l-p,0]}^{[\frac{1}{2}(l-1)]-} \binom{m}{l-i} \binom{m}{i} \\ &\qquad \times D(h^{l-i} g^{m-l+i}, h^i g^{m-i}), \end{aligned}$$

where for any real number δ $[\delta]_-$ denotes the largest integer which is not greater than δ. It follows that

$$K_0^{2m} = c_m^{-1} D^m(g)$$

and for $l = 1, 2, \cdots, 2p-1$,

$$K_l^{2m} = 2 \sum_{i=\max[l-p,0]}^{\frac{1}{2}(l-1)} K^{2m}(l, i), \quad \text{if } l \text{ is odd},$$

$$K_l^{2m} = 2 \sum_{i=\max[l-p,0]}^{\frac{1}{2}(l-2)} K^{2m}(l, i) + K^{2m}(l, l/2), \quad \text{if } l \text{ is even},$$

where

$$K^{2m}(l, i) \equiv c_m^{-1} \binom{m}{l-i} \binom{m}{l} D(h^{l-i} g^i, h^i g^{l-i}) \wedge D^{m-l}(g)$$

and finally

$$K_{2p}^{2m} = c_m^{-1} \binom{m}{p}^2 D^p(h) \wedge D^{m-p}(g). \tag{4.40}$$

To prove that the coefficients $\mathbf{K}_1, \cdots, \mathbf{K}_{2p}$ belong to φ, it is sufficient to show that, for every $l = 1, \cdots, 2p$ and $i = \max[l-p, 0], \cdots, [l/2]_-$, the sequence $\mathbf{K}(l, i)$ is an element of φ.

Indeed, for every natural number k we have

$$\begin{aligned} \sum_{m=l}^{\infty} k^{2m} |K^{2m}(l, i)| &\leq [l; i] \sum_{m=l} k^{2m} c_m^{-1} \binom{m}{l-} \binom{m}{l} c_{m-l} a_{m-l} \\ &\leq \frac{k^{2l} [l; i]}{l!\,(l-i)!\,2^{2l}} \sum_{m=0}^{\infty} \frac{(2(m+l))!}{(2m)!} k^{2m} a_m \\ &\leq \frac{(2l)!\,[l; i]}{l!\,(l-i)!} K^{2l} f(4k^2), \end{aligned}$$

where $[l; i]$ stands for $|D(h^{l-i} g^i, h^i g^{l-i})|$.

Furthermore, since $D^m(h_n)$ is of order $m < p$ in μ_n for $m < p$, it follows that $K_{2k}^{2m} = 0$ for $m < p$. Finally,

if we combine the equality

$$D^p(h) = c_p P_{[U]}$$

[cf. Proposition 4(x)] with (4.40), we obtain the expression for K_{2p}^{2m} $(m \geq p)$ asserted by Lemma 8. Q.E.D.

Now by the definition of $\beta(g, U)$ we have

$$\beta^{2m}(g, U) = 0 \quad \text{for} \quad m < p$$

and

$$\begin{aligned} C\beta^{2m}(g, U) &= a_{m-p} P_g^{m-p} \wedge_{[U]} \\ &= \|g^{m-p} \wedge [U]\|^{-2} a_{m-p}(P_{[U]} \wedge D^{m-p}(g)) \\ &= c_{m-p}^{-1} \binom{2m}{2p} (P_{[U]} \wedge D^{m-p}(g)) = K_{2p}^{2m} \end{aligned}$$

for $m > p$.

Again we apply Proposition 4(x) to obtain the above sequence of equalities. It follows that

$$C \cdot \beta(g, U) = \mathbf{K}_{2p},$$

and we are able to write

$$\begin{aligned} &p_k(\beta(h_n) - \beta(g, U)) \\ &= C^{-1}(p_k(D(h_n) \cdot (1 + \mu_n)^{-p} - \mathbf{K}_{2p})) \\ &\leq C^{-1}\left[p_k(\mathbf{K}_{2p})\left(1 - \frac{\mu_n^p}{(1 + \mu_n)^p}\right) + \sum_{i=0}^{2p-1} p_k(\mathbf{K}_j) \frac{(\mu_n^{\frac{1}{2}})^j}{(1 + \mu_n)^p} \right], \end{aligned}$$

an inequality which immediately implies

$$\lim_{n \to \infty} p_k(\beta(h_n) - \beta(g, U)) = 0.$$

This completes the proof of Proposition 6. Q.E.D.

From Proposition 6 and some elementary results of topological spaces, one can deduce the following corollary:

Corollary to Proposition 6: The set of all generalized BCS ensembles is a *connected* subset of the set φ of all states of the many-body system.

In the sequel, the following functions associated with a geminal g play a certain part:

$$\text{(i)} \quad f_\sigma \equiv \prod_{\rho \neq \sigma} (1 + \lambda_\rho t) = (1 + t\lambda_\sigma)^{-1} f,$$

$$\text{(ii)} \quad f_{\sigma\tau} \equiv \prod_{\rho \neq \sigma, \tau} (1 + \lambda_\rho t) = (1 + t\lambda_\tau)^{-1} f_\sigma. \qquad (4.41)$$

Their expansion coefficients we denote by

$$\text{(i)} \quad a_m(\hat{\sigma}) \equiv \frac{1}{m!} f_\sigma^{(m)}(0),$$

$$\text{(ii)} \quad a_m(\hat{\sigma}\hat{\tau}) \equiv \frac{1}{m!} f_{\sigma\tau}^{(m)}(0). \qquad (4.42)$$

Lemma 9:

$$\text{(i)} \quad \sum_{\sigma \in J} \lambda_\sigma a_{m-1}(\hat{\sigma}) = m a_m,$$

$$\text{(ii)} \quad a_m = a_m(\hat{\sigma}) + \lambda_\sigma a_{m-1}(\hat{\sigma}).$$

Proof:

(i) From Theorem A12 we deduce

$$\begin{aligned} f' &= \sum_{\sigma \in J} \lambda_\sigma f_\sigma \\ &= \sum_{\sigma \in J} \lambda_\sigma \sum_{m=1}^{\infty} a_{m-1}(\hat{\sigma}) t^{m-1} \\ &= \sum_{m=1}^{\infty} \left(\sum_{\sigma \in J} \lambda_\sigma a_{m-1}(\hat{\sigma}) \right) t^{m-1}. \end{aligned}$$

On the other hand, it follows from (4.24) that the power-series expansion of f' is given:

$$f' = \sum_{m=1}^{\infty} m a_m t^{m-1}.$$

The comparison of the two expansions for f' yields the desired result.

(ii) If we reformulate [4.41(i)] as

$$f(t) = (1 + \lambda_\sigma t) f_\sigma(t)$$

and write both sides of the equations as power series, we obtain

$$\begin{aligned} \sum_{m=0}^{\infty} a_m t^m &= (1 + \lambda_\sigma t) \sum_{m=1}^{\infty} a_{m-1}(\hat{\sigma}) t^{m-1} \\ &= \sum_{m=0}^{\infty} (a_m + \lambda_\sigma a_{m-1}(\hat{\sigma})) t^m. \end{aligned}$$

If we compare both expressions for f, we obtain the asserted result. Q.E.D.

The following theorem is a generalization of a theorem by Coleman[3] to a "generalized AGP function" and to *infinite* rank of the geminal g.

Theorem 8: Let U be a q-dimensional subspace of H^1. Moreover, let J_1 denote the set of all pairs of successive numbers greater than q:

$$J_1 = \{(q + 1, q + 2), \cdots, (q + 2 - 1, q + 2), \cdots\}.$$

Let $g \in H^{2\wedge}$ and let

$$W = \bigoplus_{\sigma \in J_1} W_\sigma \qquad (4.43)$$

be the representation of the closure of the range of $D^1(g)$ described in Proposition 5. Let

$$P_W = \sum_{\sigma \in J_1} P_\sigma$$

be the corresponding decomposition of the projector onto W so that $D^1(g)$ can be written as

$$D^1(g) = \sum_{\sigma \in J} \lambda_\sigma P_\sigma .$$

Furthermore, let

$$\Delta_m^i(U) \equiv L^i_{2m+q}(P_U{}^m \wedge [U]), \quad i = 1, 2.$$

Then

$$\text{(i)}\ (2m+q)\Delta_m^1(U) = a_m^{-1} \sum_{\sigma \in J_1} \lambda_\sigma a_{m-1}(\hat{\sigma}) P_\sigma + P_U$$
$$= 2m\Delta_m^1(0) + P_U .$$

(ii) Let $\{e_i\}$ be a CONS in $U \oplus W$ such that the first q elements span U and the rest is a CONS in W of the particular kind occurring in Proposition 5, and for every finite sequence $(i_1 \cdots i_p)$ of natural numbers let $[i_1 \cdots i_p]$ denote the normalized Slater determinant corresponding to the linear span of $e_{i_1} \cdots e_{i_p}$. Furthermore, let

$$b_m(ij \mid lk) \equiv ([ij] \mid \Delta_m^2(U)[lk]).$$

Then $b_m(ij \mid lk) = 0$ with the following exceptions:

$$\text{(a)}\ \binom{2m+q}{2} b(ij \mid ij) = 1, \quad i, j \in (1 \cdots q),$$

$$\text{(b)}\ \binom{2m+q}{2} a_m b(ij \mid ij) = \lambda_\sigma a_{m-1}(\hat{\sigma}),$$
$$i \in (1 \cdots q), \quad j \in \tau, \quad \tau \in J_1 ,$$

$$\text{(c)}\ \binom{2m+q}{2} a_m b_m(\sigma \mid \sigma) = \lambda_\sigma a_{m-1}(\hat{\sigma}), \quad \sigma \in J_1 ,$$

$$\text{(d)}\ \binom{2m+q}{2} a_m b_m(\sigma \mid \tau) = \lambda_\sigma^{\frac{1}{2}} \lambda_\tau^{\frac{1}{2}} a_{m-1}(\hat{\sigma}\hat{\tau}),$$
$$\sigma, \tau \in J_1 ,$$

$$\text{(e) for } m \geq 2\text{:} \binom{2m+q}{2} a_m b_m(ij \mid ij) = \lambda_\sigma \lambda_\tau a_{m-2}(\hat{\sigma}\hat{\tau}),$$
$$i \in \sigma, \quad j \in \tau, \quad \sigma \neq \tau, \quad \sigma, \tau \in J_1 .$$

The result shows that

$$\binom{2m+q}{2} \Delta_m^2(U) = \binom{2m+q}{2} \Delta_m^2(0) + P_U \wedge P_U$$
$$+ 2(2m+q)(P_U \wedge \Delta_m^1(0)).$$

Proof: Let h be the normalized vector on the ray generated by $g^m \wedge [U]$ defined by

$$h \equiv g^m \wedge [U] / \| g^m \wedge [U] \| .$$

Then h can be written as

$$h = a_m^{-\frac{1}{2}} \sum_{\sigma_1 < \cdots \sigma_m} \lambda_{\sigma_1}^{\frac{1}{2}} \cdots \lambda_{\sigma_m}^{\frac{1}{2}} [1 \cdots q, \sigma_1 \cdots \sigma_m].$$

Hence

$$([k_1, \cdots, k_{2m+q-1}] \mid h) = 0,$$

unless $(k_1 \cdots k_{2m+q-1}) = s(1 \cdots q\ \sigma_1 \cdots \sigma_m)$ for some permutation s and some $\sigma_i \in J_1$. In this case

$$([k_1 \cdots k_{2m+q-1}] \mid h) = (-1)^{\sigma(s)} \lambda_{\sigma_1}^{\frac{1}{2}} \cdots \lambda_{\sigma_m}^{\frac{1}{2}}, \quad (4.44)$$

where $\sigma(s)$ denotes the signature of s. Let us write $b_m(ij)$ for $(e_i \mid \Delta_m^1(U) e_j)$. Then

$$b_m(ij) = \sum_{i_1 \cdots i_{2m+q-1}} (e_i \otimes e_{i_1} \cdots \otimes e_{i_{2m+q-1}} \mid h)$$
$$\times (h \mid e_j \otimes e_{i_1} \cdots \otimes e_{i_{2m+q-1}})$$
$$= [(2m+q)!]^{-1} \sum_{i_1 \cdots i_{2m+q-1}} ([ii_1 \cdots i_{2m+q-1}] \mid h)$$
$$\times (h \mid [ji_1 \cdots i_{2m+q-1}])$$
$$= (2m+q)^{-1} \sum_{i_1 < \cdots < i_{2m+q-1}} ([ii_1 \cdots i_{2m+q-1}] \mid h)$$
$$\times (h \mid [ji_1 \cdots i_{2m+q-1}]). \quad (4.45)$$

Substitution of (4.44) into (4.45) yields

(a) $b_m(ij) = 0$,

(b) $i \in (1 \cdots q)(2m+q) b_m(ii)$
$$= \sum_{i_1 < \cdots i_{2m+q-1}} |([ii_1 \cdots i_{2m+q-1}] \mid h)|^2$$
$$= a_m^{-1} \sum_{\sigma_1 < \cdots \sigma_m} \lambda_{\sigma_1} \cdots \lambda_{\sigma_m} = 1,$$

(c) $i \in \sigma \in J_1$: $(2m+q) b_m(ii)$
$$= a_m^{-1} \lambda_\sigma \sum_{\sigma_1 < \cdots \sigma_{m-1}} \lambda_{\sigma_1} \cdots \lambda_{\sigma_{m-1}}$$
$$= a_m^{-1} \lambda_\sigma a_{m-1}(\hat{\sigma}).$$

(ii) In exactly the same way as we deduced Eq. (4.45), it is possible to derive

$$b_m(ij \mid kl) = \binom{2m+q}{2}^{-1}$$
$$\times \sum_{i_1 < \cdots i_{2m+q-2}} ([iji_1 \cdots i_{2m+q-2}] \mid h)$$
$$\times (h \mid [kli_1 \cdots i_{2m+q-2}]). \quad (4.46)$$

Applying the result expressed in Eq. (4.44), we obtain assertion (ii) of Theorem 8. Q.E.D.

We are now ready to calculate the first three components of $L \circ \beta(g, U)$, i.e., the contraction of the generalized BCS ensemble associated with g and U.

Theorem 9: Under the same hypothesis (with $q = 2p$) as in Theorem 8, the following statements hold. Let for $i = 1, 2$, $\Delta^i(U) = pr_i \circ L \circ \beta(g, U)$. Then

$$\text{(i)}\ \Delta^1(U) = P_U + \sum_{\sigma \in J_1} \frac{\lambda_\sigma}{1 + \lambda_\sigma} P_\sigma = \Delta^1(0) + P_U .$$

(ii) Let $\{e_i\}$ be a CONS in $U \oplus W$ such that the first $2p$ elements span U and the rest is adapted to the decomposition (4.43) of W and let $J_0 = \{(12), \cdots, (2p-1, 2p)\}$. Moreover, define

$$b(ij \mid kl) \equiv ([ij] \mid \Delta^2(0)[kl]).$$

Then $b(ij \mid kl) = 0$ with the following exceptions:

(a) $i \in \sigma, j \in \tau, \sigma, \tau \in J_0$, $b(ij \mid ij) = 1$,

(b) $i \in \sigma \in J_0, j \in \tau \in J_1$, $b(ij \mid ij) = \dfrac{\lambda_\tau}{1+\lambda_\tau}$,

(c) $\sigma \in J_1$, $\qquad b(\sigma \mid \sigma) = \dfrac{\lambda_\sigma}{1+\lambda_\sigma}$,

(d) $\sigma, \tau \in J_1, \sigma \neq \tau$, $\quad b(\sigma \mid \tau) = \dfrac{\lambda_\sigma^{\frac{1}{2}}\lambda_\tau^{\frac{1}{2}}}{(1+\lambda_\sigma)(1+\lambda_\tau)}$,

(e) $i \in \sigma, j \in \tau, \sigma < \tau, \sigma, \tau \in J_1$,

$$b(ij \mid ij) = \frac{\lambda_\sigma \lambda_\tau}{(1+\lambda_\sigma)(1+\lambda_\tau)}.$$

The result shows that

$$\Delta^2(U) = \Delta^2(0) + P_U \wedge P_U + 2(P_U \wedge \Delta^1(0)).$$

Proof:

(i) $$\Delta^1(U) = C^{-1}\sum_{m=0}^{\infty} 2(m+p) a_m \Delta_m^1(U)$$

$$= C^{-1}\sum_{m=0}^{\infty}\left(a_m P_U + \sum_{\sigma \in J_1} \lambda_\sigma a_{m-1}(\hat{\sigma}) P_\sigma\right)$$

$$= P_U + C^{-1}\sum_{\sigma \in J_1} \lambda_\sigma f_\sigma(1) P_\sigma$$

$$= P_U + \sum_{\sigma \in J_1} \frac{\lambda_\sigma}{1+\lambda_\sigma} P_\sigma.$$

To get the last equality we used part (i) of formula (4.41).

(ii) $i, j \in (1, \cdots, 2p)$:

$$b(ij \mid ij) = C^{-1}\sum_{m=0}^{\infty}\binom{2(m+p)}{2} a_m b_m(ij \mid ij) = 1;$$

$i \in (1, \cdots, 2p)$, $j \in \sigma \in J_1$:

$$b(ij \mid ij) = C^{-1}\sum_{m=0}^{\infty}\binom{2(m+p)}{2} a_m b_m(ij \mid ij)$$

$$= C^{-1}\sum_{m=1}^{\infty} \lambda_\sigma a_{m-1}(\hat{\sigma}) = C^{-1} f_\sigma(1)\lambda_\sigma = \frac{\lambda_\sigma}{1+\lambda_\sigma};$$

$\sigma \in J_1$:

$$b(\sigma \mid \sigma) = C^{-1}\sum_{m=0}^{\infty}\binom{2(m+p)}{2} a_m b_m(\sigma \mid \sigma)$$

$$= C^{-1}\sum_{m=1}^{\infty} \lambda_\sigma a_{m-1}(\hat{\sigma}) = \frac{\lambda_\sigma}{1+\lambda_\sigma};$$

$\sigma, \tau \in J_1, \sigma \neq \tau$:

$$b(\sigma \mid \tau) = C^{-1}\sum_{m=0}^{\infty}\binom{2(m+p)}{2} a_m b(\sigma \mid \tau)$$

$$= C^{-1}\lambda_\sigma^{\frac{1}{2}}\lambda_\tau^{\frac{1}{2}}\sum_{m=1}^{\infty} a_{m-1}(\hat{\sigma}\hat{\tau}) = C^{-1} f_{\sigma\tau}(1)\lambda_\sigma^{\frac{1}{2}}\lambda_\tau^{\frac{1}{2}}$$

$$= \frac{\lambda_\sigma^{\frac{1}{2}}\lambda_\tau^{\frac{1}{2}}}{(1+\lambda_\sigma)(1+\lambda_\tau)};$$

$i \in \sigma$, $j \in \tau$, $\sigma < \tau$, $\sigma, \tau \in J_1$:

$$b(ij \mid ij) = C^{-1}\sum_{m=0}^{\infty}\binom{2(m+p)}{2} a_m b_m(ij \mid ij)$$

$$= C^{-1}\lambda_\sigma\lambda_\tau\sum_{m=2}^{\infty} a_{m-2}(\hat{\sigma}\hat{\tau}) = C^{-1}\lambda_\sigma\lambda_\tau f_{\sigma\tau}(1)$$

$$= \frac{\lambda_\sigma}{1+\lambda_\sigma}\frac{\lambda_\tau}{1+\lambda_\tau}.$$ Q.E.D.

Theorem 10: Let $\mathbf{\Delta} \equiv (1, \Delta^1, 0, \cdots)$ be an actual one-state with the property that Δ^1 has evenly degenerate eigenvalues. Then there exists a generalized BCS ensemble $\beta(g, U)$ such that

$$L^1(\beta(g, U)) = \mathbf{\Delta}.$$

Proof: Let $W \equiv R_{\Delta^1}$ and let

$$W = \bigoplus_{\sigma \in J} W_\sigma$$

be a representation of W as a direct sum of two-dimensional subspaces which is compatible with the decomposition into a direct sum of eigenspaces of Δ^1.

Let P_σ be the projector corresponding to W_σ and let

$$\Delta^1 = \sum_{\sigma \in J} \mu_\sigma P_\sigma, \quad \mu_\sigma \leq \mu_\tau \quad \text{for} \quad \sigma \geq \tau,$$

where J is defined by Eq. (4.21). Since $\mathbf{\Delta} \in \overline{\mathfrak{P}^{[1]}}$, it follows that $\|\Delta^1\| \leq 1$, and hence $0 < \mu_\sigma \leq 1$ for all $\sigma \in J$. Since, on the other hand, $\mathrm{Tr}\,(\Delta^1) = \sum_{\sigma \in J} \mu_\sigma < \infty$, only a finite number of μ_σ's can be equal to 1:

$$\mu_\sigma = 1 \quad \text{for} \quad J_0 \equiv \{(12), \cdots, (2p-1, 2p)\}$$

and

$$0 < \mu_\sigma < 1 \quad \text{for} \quad \sigma \in J - J_0 \equiv J_1.$$

Define

$$U \equiv \bigoplus_{\sigma \in J_0} W_\sigma,$$

$$g \equiv 2^{\frac{1}{2}}\sum_{\sigma \in J_1}\left(\frac{\mu_\sigma}{1-\mu_\sigma}\right)^{\frac{1}{2}}[\sigma],$$

where $[\sigma]$ stands for some normalized Slater determinant associated with W_σ.

Now it is an immediate consequence of Theorem 9(i) that

$$L^1(\beta(g, U)) = \mathbf{\Delta}.$$ Q.E.D.

According to Theorem 9, the eigenvalues n_σ of $\Delta^1(U) \equiv pr_1 \circ L \circ \beta(g, U)$ are given by

$$n_\sigma = 1, \quad \sigma \in J_0,$$
$$n_\sigma = \lambda_\sigma/1 + \lambda_\sigma, \quad \sigma \in J_1. \tag{4.47}$$

What is their physical meaning? To answer this question let us look at the observable

$$N_\sigma \equiv (0, P, 0, \cdots),$$

where P is the projector onto a pure state contained in the subspace W_σ which corresponds to the projector P_σ, i.e., a one-dimensional projector for which $PP_\sigma = P$. It is clear that

$$(\Gamma(N_\sigma) \mid \beta(g, U)) = n_\sigma.$$

Thus n_σ can be interpreted as the expectation value for the occupation number of any pure state contained in W_σ if the total system is in the state $\beta(g, U)$.

Using the concept of the occupation number defined by (4.47), the content of Theorem 9 can simply be expressed as

(i) $$\Delta^1(U) = \sum_{\sigma \in J} n_\sigma P_\sigma,$$

(ii) $b(ij \mid kl) = 0$ except *for*: (4.48)

$\sigma \in J$, $\quad b(\sigma \mid \sigma) = n_\sigma$,

$\sigma \neq \tau$, $\sigma, \tau \in J$, $\quad b(\sigma \mid \tau) = n_\sigma^{\frac{1}{2}} n_\tau^{\frac{1}{2}} \times (1 - n_\sigma)^{\frac{1}{2}}(1 - n_\tau)^{\frac{1}{2}}$,

$i \in \sigma$, $j \in \tau$, $\sigma < \tau$, $\quad b(ij \mid ij) = n_\sigma n_\tau$.

$\sigma, \tau \in J$,

Hence the expectation value of a two-particle observable

$$\mathbf{B} = (B^0, B^1, B^2, \cdots)$$

in the state $\beta(g, U)$ can be written as

$$\langle \mathbf{B} \rangle \equiv (\Gamma(\mathbf{B}) \mid \beta(g, U)) = B^0 + \sum_{\sigma \in J} n_\sigma (\mathrm{Tr}\,(B^1 P_\sigma) + b_{\sigma\sigma}) + 2 \sum_{\sigma < \tau} n_\sigma^{\frac{1}{2}} n_\tau^{\frac{1}{2}} (1 - n_\sigma)^{\frac{1}{2}}(1 - n_\tau)^{\frac{1}{2}} b_{\sigma\tau} + \sum_{\sigma < \tau} n_\sigma n_\tau b_\sigma^\tau,$$

where

$$b_{\sigma\tau} \equiv ([\sigma] \mid B^2[\tau]), \quad \sigma, \tau \in J,$$
$$b_\sigma^\tau \equiv \sum_{i \in \sigma} \sum_{j \in \tau} ([ij] \mid B^2[ij]), \quad \sigma \neq \tau, \quad \sigma, \tau \in J.$$

For example, for the observable

$$\mathbf{N}^2 = (0, I, 2A_2, 0, \cdots),$$

we obtain

$$\langle \mathbf{N}^2 \rangle = 4 \sum_{\sigma \in J} n_\sigma + 8 \sum_{\sigma < \tau} n_\sigma n_\tau = \langle \mathbf{N} \rangle^2 + 4 \sum_{\sigma \in J} n_\sigma (1 - n_\sigma). \tag{4.49}$$

Thus the dispersion of $\mathbf{N}$ is given by

$$\langle \mathbf{N}^2 \rangle - \langle \mathbf{N} \rangle^2 = 4 \sum_{\sigma \in J} n_\sigma (1 - n_\sigma). \tag{4.50}$$

It is minimal ($= 0$) if $g = 0$, i.e., if the (generalized) BCS ensemble degenerates to a Slater determinant. On the other hand, it is clear that, by an appropriate choice of the n_σ's, the dispersion can be made arbitrarily large.

Let W be a subspace of H^1 whose dimension is finite and even. We associate with W the following subset of the set $\beta(\mathfrak{G})$ of all generalized BCS ensembles:

$$\theta_W \equiv \{\mathbf{D} \in \beta(\mathfrak{G});\ R_{L(\mathbf{D})^1} \subseteq W\}.$$

Then the dispersion of $\mathbf{N}$ takes a maximum value on θ_W for $\beta(g_e, 0)$, where g_e denotes the *extreme* geminal associated with W (cf. Coleman[3]), i.e., the geminal for which $D^1(g)$ coincides with the projector onto W. The value of the maximum equals $\frac{1}{2}$ dim W.

If the value of $\langle \mathbf{N} \rangle$ is prescribed, the following situation holds: For practical purposes we may assume that $\langle \mathbf{N} \rangle$ is an integer.

Let $\{W_r\}$ be an increasing sequence of even-dimensional subspaces of H^1 such that

$$\overline{\bigcup_r W_r} = H^1.$$

Let dim $W_r = r \geq \langle \mathbf{N} \rangle$ and let g_e^r be the extreme geminal associated with W_r. Furthermore, let $\mu_r \equiv (r - \langle \mathbf{N} \rangle)^{-1} \langle \mathbf{N} \rangle$. Then the expectation value of $\mathbf{N}$ for the BCS ensemble $\beta(\mu_r^{\frac{1}{2}} g_e^r, 0)$ is equal to $\langle \mathbf{N} \rangle$ and the dispersion is given by

$$2\langle \mathbf{N} \rangle (1 - r^{-1} \langle \mathbf{N} \rangle).$$

From this we conclude that the lowest upper bound of the dispersion for given expectation value $\langle \mathbf{N} \rangle$ of the particle number is equal to $2\langle \mathbf{N} \rangle$ and that it is not attained for any BCS ensemble.

For the minimal value we have to distinguish two cases: (a) $\langle \mathbf{N} \rangle$ = even number. Then the minimal dispersion is 0 and it is taken for $\beta(0, U)$ with dim $U = \langle \mathbf{N} \rangle$. (b) $\langle \mathbf{N} \rangle$ = odd number. Then the minimal dispersion is 1 and is taken for $\beta([\sigma], U)$, where dim $U = \langle \mathbf{N} \rangle - 1$ and $[\sigma]$ is a Slater determinant associated with a two-dimensional subspace of H^1 orthogonal to U.

In general, we may conclude from the particular form of expression (4.50) for the dispersion of $\mathbf{N}$ that the latter is relatively small for states in which the occupation numbers are either close to 0 or close to 1.

ACKNOWLEDGMENTS

The author is indebted to Professor A. J. Coleman for his generous hospitality at Queen's University and

also for his lectures on superconductivity which stimulated the writing of this paper. Moreover, the author would like to thank Professor R. Giles for his advice and encouragement, and his colleague David Gregory for helpful discussions.

The author was supported by the Canadian National Research Council in Ottawa. He would like to express his gratitude to this institution.

APPENDIX

1. A Lemma about Vector Spaces of Sequences

Lemma A1: If $\{\mathfrak{B}_n\}$ is a sequence of real or complex Banach spaces with the respective norms $\{\| \ \|_n\}$, then the sets

$$\mathfrak{B} = \{\mathbf{B}; \mathbf{B} = (B^0, B^1, \cdots)B^n \in \mathfrak{B}_n \sum \beta_n \|B^n\|_n < \infty\}$$

and

$$\mathfrak{B}' = \{\mathbf{B}; \mathbf{B} = (B^0, B^1, \cdots)B^n \in \mathfrak{B}_n \times \sup_n \beta_n \|B^n\|_n < \infty\},$$

where $\{\beta_n\}$ stands for a sequence of strictly positive numbers, are vector spaces under the componentwise addition and multiplication with a scalar. Equipped with the norms

$$\| \ \| \equiv \sum_n \beta_n \|B^n\|_n$$

and

$$\| \ \|' \equiv \sup_n \beta_n \|B^n\|_n,$$

they are again Banach spaces.

Proof: Let $\{\mathbf{B}_l\}$ be a Cauchy sequence in $\mathfrak{B}$; then, for any $\epsilon > 0$, there exists a natural number $N(\epsilon)$ such that

$$\|\mathbf{B}_l - \mathbf{B}_m\| < \epsilon, \quad l, m \geq N(\epsilon).$$

We have for any integer $s \geq 0$

$$\sum_{n=0}^{s} \beta_n \|B_l^n - B_m^n\|_n < \epsilon, \quad l, m \geq N(\epsilon), \qquad \text{(A1)}$$

and hence for any n, $\{B_m^n\}_{m=1}^{\infty}$ is a Cauchy sequence in $\mathfrak{B}_n$. Since by assumption $\mathfrak{B}_n$ is a Banach space, there exists $C^n \in \mathfrak{B}_n$ such that

$$\lim_{m \to \infty} B_m^n = C^n.$$

Going in (A1) to the limit $m \to \infty$, we obtain

$$\sum_{n=0}^{s} \beta_n \|B_l^n - C^n\|_n \leq \epsilon.$$

Now forming the limit $s \to \infty$ we get

$$\|\mathbf{B}_l - \mathbf{C}\| \leq \epsilon, \quad l \geq N(\epsilon).$$

Hence $\mathbf{C} \in \mathfrak{B}$ and $\lim_{l \to \infty} \mathbf{B}_l = \mathbf{C}$. A similar proof holds for the second statement of the lemma. Q.E.D.

2. Some Notions and Theorems about Real Linear Spaces and Convex Sets[5]

This part of the Appendix is a reprint of the Appendix to NRP with the only difference that we have added the Krein–Milman theorem (Theorem A9).

Definition A1 (Ref. 5, p. 31): A dual pair (of real linear spaces) is a pair of real linear spaces $\langle E_1, E_2\rangle$ such that to every pair $(x, u) \in E_1 \times E_2$ there corresponds a real number denoted by xu, such that:

(D1) xu is a bilinear form;
(D2′) if $xu = 0$ for all $u \in E_2$, then $x = 0$;
(D2″) if $xu = 0$ for all $x \in E_1$, then $u = 0$.

Definition A2: A topology in E_1 is called *compatible with the dual pair* $\langle E_1, E_2\rangle$ if it is locally convex and if the topological dual space of E_1 coincides with E_2.

Theorem A1 (Ref. 5, p. 34, Proposition 8): Let $\langle E_1, E_2\rangle$ be a dual pair and let M be a subset of E_1. The operation of taking the convex closure of the set M is independent of the topology as long as it is compatible with the dual pair.

Definition A3 (Ref. 5, p. 32): Let $\langle E_1, E_2\rangle$ be a dual pair. The weakest topology on E_1 compatible with the dual pair coincides with the weakest topology under which the set of all linear functionals

$$f_u(x) = xu, \quad u \in E_2,$$

are continuous. It is called the *weak topology* in E_1.

Definition A4: Let $\langle E_1, E_2\rangle$ and $\langle F_1, F_2\rangle$ be two dual pairs. Let A be a map from E_1 into F_1. Then the expression $(Ax)v$ represents a linear form in E_1 for fixed $v \in F_2$. In other words, there exists an element $v^1 \in E_1^*$ such that $(Ax)v = xv'$. The map $A' : v \to v'$ is called the *adjoint map* of A. If A is linear, A' is linear.

Theorem A4 (Ref. 5, p. 38): Let $\langle E_1, E_2\rangle$ and $\langle F_1, F_2\rangle$ be two dual pairs. A linear map A from E_1 into F_1 is weakly continuous if and only if $v' = A'v$ E_2 for every $v \in F_2$.

Definition A5 (Ref. 5, p. 4): Let E be a real linear space. A subset C of E is called *convex* if whenever $x, y \in C$, the whole segment

$$[x, y] = \{z \in E; z = \alpha x + (1 - \alpha)y \ 0 \leq \alpha \leq 1\}$$

is a subset of C.

Definition A6 (Valentine,[13] p. 137): A subset C' of a convex set C is called *extreme*, if whenever $[x, y] \in C$ shares an inner point with C', then in fact $[x, y] \in C'$. An extreme subset of C consisting of a single point is called an *extreme point*.

Definition A7 (Ref. 13, p. 27): A convex subset $\mathcal{K}$ of a real linear space E is called a *convex cone* with vertex a if whenever $x \in \mathcal{K}$, then the half-ray

$$r(x) = \{zE;\ z = a + \alpha x\ \alpha \geq 0\}$$

is a subset of $\mathcal{K}$. A convex cone is called *pointed* if $\mathcal{K} - a$ does not contain a one-dimensional subspace of E.

Definition A8 (cf. Köthe,[14] p. 187): Let C be a convex subset of a real linear space E not containing $a \in E$. The projection cone of C from the center a is the smallest convex cone with vertex a containing C.

Definition A9: Let $\langle E_1, E_2\rangle$ be a dual pair and let M be a subset of E_1. The subspace of E_2 defined by

$$M = \{u \in E_2;\ xu = 0,\ x \in M\}$$

is called the *orthogonal complement* of M.

Theorem A5 (special case of the bipolar theorem): (i) M is a closed subspace of E_2; (ii) M is the smallest closed subspace containing M.

Definition A10 (Ref. 14, p. 246): Let $\langle E_1, E_2\rangle$ be a dual pair and let M be a subset of E_1. Then the subset

$$\tilde{M} = \{u \in E_2;\ xu \geq 0,\ x \in M\}$$

is called the *polar cone* of M. ($M_1 \subseteq M_2$ implies $\tilde{M}_1 \supseteq \tilde{M}_2$.)

Theorem A6 (bipolar theorem; Ref. 5, p. 36; cf. also Halperin[15]):

(i) $\tilde{M}$ is a closed convex cone with vertex 0 contained in E_2;

(ii) $\tilde{\tilde{M}}$ is the smallest closed convex cone with vertex 0 containing M.

Definition A11: Let $\langle E_1, E_2\rangle$ be a dual pair and $a \in E_2$ ($a \neq 0$). Then the orthogonal complement $\{a\}^\perp$ is a closed subspace of codimension 1 in E_1 [cf. Theorem A5(i)]. Let $b \in E_1$. Then the set $b + \{a\}^\perp$ is called a *closed hyperplane* in E_1.

Definition A12 (cf. Ref. 14, p. 340): Let $\langle E_1, E_2\rangle$ be a dual pair and $\mathcal{K}$ a closed convex cone in E_1 with vertex 0 and let $a \in \tilde{\mathcal{K}} - \{0\}$. Then $\{a\}^\perp$ is a *supporting hyperplane* of $\mathcal{K}$. The intersection $\{a\}^\perp$ is called an *exposed subset* of $\mathcal{K}$. A point of $\mathcal{K}$ belonging to an exposed subset is called a *supporting point* of $\mathcal{K}$. A ray $r(x) = \{zE;\ z = \alpha x\ \alpha \geq 0\}$ being at the same time an exposed subset of $\mathcal{K}$ is called an *exposed ray* of $\mathcal{K}$.

Remark: If $\mathcal{K}$ is the projection cone of a closed convex subset C of a closed hyperplane not containing 0 and if $a \in \tilde{\mathcal{K}}$, then the set $\{a\}^\perp \cap C$ is called an *exposed subset* of C provided it is not empty. An exposed subset of C consisting of a single point is called an *exposed point*.

Theorem A7 (cf. NRP Appendix, Theorem A7): Every exposed subset of a closed convex cone $\mathcal{K} \subseteq E_1$ is closed, convex, and extreme.

Theorem A8 (Theorem of Klee; cf. Ref. 14, p. 340): Let C be a compact convex subset of a normed linear space. C is the closed convex hull of all the exposed points of C.

Theorem A9 (Krein–Milman; cf. Ref. 5, p. 138): If E is a separated convex space and C is a convex compact subset of E_1, then C is the closed convex envelope of the set of its extremal points. (For a normed space, Theorem A9 is a consequence of Theorem A8 since, according to Theorem A7, the set of exposed points of C is contained in the set of extreme points.)

3. Some Notions and Theorems about Infinite Sequences, Series, and Products of Holomorphic Functions[16]

Let $\mathbf{C}$ be the complex plane and $D \subseteq \mathbf{C}$ an open subset.

Let $\mathcal{C}(D)$ be the linear space of all continuous complex valued functions on D, and $\mathcal{H}(D)$ the subspace of all holomorphic functions in $\mathcal{C}(D)$. We equip $\mathcal{C}(D)$ with the topology of *compact convergence*. It is generated by the following sequence of seminorms (cf. Ref. 5, p. 20):

$$p_n(f) = \sup_{t \in A_n} |f(t)|, \quad f \in \mathcal{C}(D),$$

[13] F. A. Valentine, *Convex Sets* (McGraw-Hill Book Co., New York, 1964).

[14] G. Köthe, *Topologische lineare Räume* (Springer-Verlag, Berlin, 1960).

[15] I. Halperin: Trans. Roy. Soc. Canada XLVII, Ser. III, Sec. 3, 1 (1953).

[16] H. Cartan, *Théorie élémentaire des fonctions analytiques d'une ou plusieurs variables complexes* (Hermann et Cie, Paris, 1961).

where

$$A_n = \{t \in \mathbf{C};\ |t| \leq n,\ d(t, \sim D) \geq n^{-1}\}.$$

($\sim D$ denotes the complement of D in $\mathbf{C}$.)

A sequence $\{f_n\}$ is convergent in the topology of compact convergence if it *converges uniformly on every compact subset* $K \subseteq D$.

Theorem A10 (cf. Ref. 16, p. 150):

(i) $\mathcal{C}(D)$ is complete;

(ii) $\mathcal{K}(D)$ is a closed subspace of $\mathcal{C}(D)$;

(iii) the mapping which maps $f \in \mathcal{K}(D)$ onto its derivative $f' \in \mathcal{K}(D)$ is continuous.

Definition A13 (cf. Ref. 16, p. 144): A series $\sum_n f_n$ of functions $f_n \in \mathcal{C}(D)$ *converges normally on every compact subset* if, for every compact subset $K \subseteq D$, the series of the restrictions f_n/K is dominated by a convergent series of positive numbers. It is clear that if the series $\sum_n f_n$ converges normally on every compact subset, then the sequence of its partial sums converges uniformly on every compact subset.

Definition A14 (cf. Ref. 16, p. 160): The infinite product $\prod_n f_n$ of functions $f_n \in \mathcal{C}(D)$ is called *normally convergent on every compact subset* if the series $\sum u_n$ with $u_n \equiv f_n - 1$ converges normally on every compact subset.

Theorem A11 (cf. Ref. 16, p. 161): Let $\{f_n\}$ be a sequence of elements of $\mathcal{K}(D)$ such that the product $\prod_n f_n$ is normally convergent on every compact subset.

Then

(i) The sequence of partial products constitutes a sequence of holomorphic functions which is uniformly convergent on every compact subset. Hence its limit $f \equiv \prod f_n$ belongs to $\mathcal{K}(D)$ [Theorem A10 (ii)].

(ii) For every natural number p the formula

$$f = f_1 f_2 \cdots f_p \prod_{n>p} f_n$$

is valid.

Theorem A12 (cf. Ref. 16, p. 161): With the hypothesis of Theorem A11 the following statement holds: The series of holomorphic functions $\sum_n f/f_n \cdot f_n$ is normally convergent on every compact subset and its limit is equal to the derivative f'.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Evaluation of a Unitary Integral. II. Mandelstam Iteration*

JAMES DILLEY†
Theoretical Physics Institute, Department of Physics, University of Alberta, Edmonton, Canada

(Received 12 May 1969)

The methods of a previous communication are used to derive a general expression for the first Mandelstam iteration of inputs of the form

$$A_t(s,t) = \sum_j g_j(s,t)\delta(t - t_{Rj}),$$

where the sum is over terms representing t-channel resonances in narrow-width approximation. The coefficients $g_j(s,t)$ can be either analytic in s, as for an expansion in t-channel partial waves, or meromorphic in s, as in a Padé- or Schlessinger-type expansion.

In a previous communication,[1] we derived a general procedure for evaluating integrals of the type

$$I = \int_{s_+}^{\infty} \rho(s) f(s)\, ds, \tag{1}$$

$$\rho(s) = s^{-1}[(s - s_+)(s - s_-)]^{\frac{1}{2}}, \tag{2}$$

where $s_+ > s_-$ and $f(s)$ is an arbitrary meromorphic function such that $s^{1+\epsilon} f(s) \to 0$, as $|s| \to \infty$; we applied it to several situations involving partial wave amplitudes. Here, we wish to point out that the same methods can also be used to evaluate the first Mandelstam iteration for an input form commonly used in strong interaction calculations:

$$A_t(s,t) = \sum_j g_j(s,t)\delta(t - t_{Rj}), \tag{3}$$

where the sum is over terms representing t-channel resonances at $t = t_{Rj}$ in narrow-width approximation. The sum in (3) could come either from an expansion in t-channel partial waves, in which case g_j would be analytic in s, or resonance forms continued analytically by Padé approximants or the methods of Schlessinger,[2] in which case g_j would be meromorphic in s.

The central ideas of the iteration were first described by Mandelstam[3] and have been recently reviewed by Collins and Johnson.[4] The double spectral function is of the form[5]

$$A_{st} = \left(\frac{s - 4m^2}{s}\right)^{\frac{1}{2}} \times \int_{4m^2}^{\infty}\int_{4m^2}^{\infty} \frac{dt_1\, dt_2 A_t(s,t_1)A_t^*(s,t_2)\theta(t - t_+)}{\nu[K(t,t_1,t_2;s)]^{\frac{1}{2}}}, \tag{4}$$

$$K(t,t_1,t_2;s) = t^2 + t_1^2 + t_2^2 - 2(tt_1 + tt_2 + t_1t_2) - tt_1t_2/\nu \tag{5a}$$

$$= (t - t_+)(t - t_-), \tag{5b}$$

$$t_\pm = t_1 + t_2 + \frac{t_1t_2}{2\nu} \pm 2\left(t_1t_2\left(1 + \frac{t_1}{4\nu}\right)\left(1 + \frac{t_2}{4\nu}\right)\right)^{\frac{1}{2}}, \tag{6}$$

$$s = 4(\nu + m^2). \tag{7}$$

When (3) is inserted in (4), a series of integrals is obtained having effective minimum values at $t_1 = t_{Rl}$ and $t_2 = t_{Rj}$. Because of the factor $\theta(t - t_+)$, these terms are zero unless

$$t > t_{lj} = t_{Rl} + t_{Rj} + \frac{t_{Rl}t_{Rj}}{2\nu} + 2\left[t_{Rl}t_{Rj}\left(1 + \frac{t_{Rl}}{4\nu}\right)\left(1 + \frac{t_{Rj}}{4\nu}\right)\right]^{\frac{1}{2}} \xrightarrow[s\to\infty]{} [(t_{Rl})^{\frac{1}{2}} + (t_{Rj})^{\frac{1}{2}}]^2 \tag{8}$$

or, in s,

$$s > s_{lj} = 4m^2 + 4t_{Rl}t_{Rj}t/[(t - t_{Rl} - t_{Rj})^2 - 4t_{Rl}t_{Rj}]. \tag{9}$$

Using (5a) and (9), one finds

$$4\nu K(t,t_{Rl},t_{Rj};s) = [t^2 + t_{Rl}^2 + t_{Rj}^2 - 2(tt_{Rl} + tt_{Rj} + t_{Rl}t_{Rj})](s - s_{lj}) \tag{10}$$

and the double spectral function obtained can be written in the form

$$A_{st} = \sum_{l,j} \frac{4g_l(s,t_{Rl})g_j(s,t_{Rj})\theta(s - s_{lj})}{[(t - t_{Rl} - t_{Rj})^2 - 4t_{Rl}t_{Rj}]^{\frac{1}{2}}[s(s - s_{lj})]^{\frac{1}{2}}}. \tag{11}$$

Then, for $t > [(t_{Rl})^{\frac{1}{2}} + (t_{Rj})^{\frac{1}{2}}]^2$, we find that $A_t(s,t)$ has new contributions from (4), obtained

* Supported in part by the National Research Council of Canada.
† Present address: Department of Physics, Ohio University, Athens, Ohio 45701.
[1] J. Dilley, J. Math. Phys. **8**, 2022 (1967).
[2] L. Schlessinger, Phys. Rev. **167**, 1411 (1968).
[3] S. Mandelstam, Phys. Rev. **122**, 1344 (1958).
[4] P. D. B. Collins and R. C. Johnson, Phys. Rev. **169**, 1222 (1968).
[5] For simplicity, we limit ourselves here to the equal-mass case and ignore the other single spectral function A_u, but the method is applicable in general.

from the dispersion relation

$$A_t(s,t) = \frac{1}{\pi}\int \frac{A_{st}(s',t)}{s'-s}\,ds' \tag{12}$$

$$= \sum_{l,j} \frac{4}{[(t-t_{Rl}-t_{Rj})^2 - 4t_{Rl}t_{Rj}]^{\frac{1}{2}}} \times \frac{1}{\pi}\int_{s_{lj}}^{\infty} \frac{g_l(s',t_{Rl})g_j(s',t_{Rj})\,ds'}{(s'-s)[s'(s'-s_{lj})]^{\frac{1}{2}}}. \tag{13}$$

The integral in (13) is of the same general type as Eq. (1) and can be evaluated using the methods of Ref. 1. One need only construct a function $G(s, s_{lj})$ having only the right-hand cut of $[s(s-s_{lj})]^{-\frac{1}{2}}$ and proceed as before. The required function is

$$G(s,s_{lj}) = \frac{1}{[s(s-s_{lj})]^{\frac{1}{2}}} + \frac{1}{\pi i}\frac{1}{[s(s-s_{lj})]^{\frac{1}{2}}} \ln \frac{(s)^{\frac{1}{2}} - (s-s_{lj})^{\frac{1}{2}}}{(s)^{\frac{1}{2}} + (s-s_{lj})^{\frac{1}{2}}}. \tag{14}$$

If g_j are analytic in s, this leads to

$$\int_{s_{lj}}^{\infty} \frac{ds' g_l(s',t_{Rl})g_j(s',t_{Rj})}{(s'-s)[s'(s'-s_{lj})]^{\frac{1}{2}}} = \pi i G(s,s_{lj})g_l(s,t_{Rl})g_j(s,t_{Rj}) \tag{15}$$

or

$$A_t(s,t) = \sum_{l,j} \frac{4g_l(s,t_{Rl})g_j(s,t_{Rj})iG(s,s_{lj})\theta\{t-[(t_{Rl})^{\frac{1}{2}}+(t_{Rj})^{\frac{1}{2}}]^2\}}{[(t-t_{Rl}-t_{Rj})^2 - 4t_{Rl}t_{Rj}]^{\frac{1}{2}}}, \tag{16}$$

provided that the asymptotic requirement

$$G(s,s_{lj})g_l(s,t_{Rl})g_j(s,t_{Rj}) \xrightarrow[|s|\to\infty]{} 0 \tag{17}$$

is satisfied. Explicit forms for G with real s are

$$\pi i G(s,s_{lj}) = \frac{1}{[s(s-s_{lj})]^{\frac{1}{2}}} \ln \frac{(s_{lj}-s)^{\frac{1}{2}} - (-s)^{\frac{1}{2}}}{(s_{lj}-s)^{\frac{1}{2}} + (-s)^{\frac{1}{2}}}, \quad s<0,$$

$$= \frac{-\pi}{[s(s_{lj}-s)]^{\frac{1}{2}}}\left(1 + \frac{1}{\pi}\arg\frac{(s)^{\frac{1}{2}} - i(s_{lj}-s)^{\frac{1}{2}}}{(s)^{\frac{1}{2}} + i(s_{lj}-s)^{\frac{1}{2}}}\right), \quad 0<s<s_{lj},$$

$$= \frac{1}{[s(s-s_{lj})]^{\frac{1}{2}}}\left[\ln\left(\frac{(s)^{\frac{1}{2}} - (s-s_{lj})^{\frac{1}{2}}}{(s)^{\frac{1}{2}} + (s-s_{lj})^{\frac{1}{2}}}\right) \pm i\pi\right], \quad s>s_{lj}, \tag{18}$$

where all the radicals have a positive sign and the $\pm$ are for $s = s \pm i\epsilon$ on the cut $s > s_{lj}$.

If the condition (17) does not hold, then subtractions are required in the expression for $A(s,t)$, but the integral can be handled in the same way. For example, if one subtraction is required, we write

$$A_t(s,t) = A_t(s_0,t) + \frac{s-s_0}{\pi}\int_{s_{lj}}^{\infty} \frac{A_{st}(s',t)\,ds'}{(s'-s)(s'-s_0)}$$

$$= A_t(s_0,t) + \sum_{l,j} \frac{4(s-s_0)}{[(t-t_{Rl}-t_{Rj})^2 - tt_{Rl}t_{Rj}]^{\frac{1}{2}}} \times \frac{1}{\pi}\int_{s_{lj}}^{\infty} \frac{g_l(s',t_{Rl})g_j(s',t_{Rj})\,ds'}{(s'-s)(s'-s_0)[s'(s'-s_{lj})]^{\frac{1}{2}}}, \tag{19}$$

which can be evaluated as before, since there is only an additional term coming from the pole at s_0:

$$A_t(s,t) = \sum_{l,j}\left(\frac{4g_l(s,t_{Rl})g_j(s,t_{Rj})iG(s,s_{lj})}{[(t-t_{Rl}-t_{Rj})^2 - 4t_{Rl}t_{Rj}]^{\frac{1}{2}}} + \frac{4g_l(s_0,t_R)g_j(s,t_{Rj})iG(s_0,s_{lj})}{[(t-t_{Rl}-t_{Rj})^2 - 4t_{Rl}t_{Rj}]^{\frac{1}{2}}}\right) + A_t(s_0,t). \tag{20}$$

The first term in (20) is just the unsubtracted solution (16) and still holds up to factors which are independent of s.

If $g_j(s,t)$ is meromorphic instead of analytic in s, then the same procedure for evaluating (13) can be used except that new terms appear on the right-hand side of (15) coming from the poles of g.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Evaluation of Some Fermi-Dirac Integrals*

A. WASSERMAN
Oregon State University, Corvallis, Oregon 97731
AND
T. J. BUCKHOLTZ AND H. E. DEWITT
Lawrence Radiation Laboratory, University of California, Livermore, California 94550

(Received 20 November 1968; Revised Manuscript Received 6 June 1969)

The evaluation of Fermi–Dirac integrals is discussed for cases in which the Sommerfeld method fails. Such cases occur when the integrand has a singularity at the Fermi surface and when the integrand is a rapidly oscillating function. As examples, the first-order exchange integral for electrons and the free-energy integral of the noninteracting electron gas in a magnetic field are evaluated. The method uses a contour-integral representation of the Fermi function (previously mentioned by Dingle), supplemented by Mittag–Leffler type expansions.

1. INTRODUCTION

Integrals of the Fermi–Dirac form,

$$F(\alpha) = \int_0^\infty dz \, \frac{G(z)}{e^{z-\alpha} + 1}, \tag{1}$$

occur throughout the physics of metals and semiconductors,[1] as well as in formulations of the many-fermion problem,[2] and appear in both equilibrium and nonequilibrium calculations. For $\alpha < 0$, integrals of this type which are encountered in physical problems can usually be evaluated by elementary methods. In cases with (i) high degeneracy ($\alpha \gg 0$) and (ii) $G(z)$ slowly varying near $z = \alpha$ and possessing a Taylor series expansion with a reasonable radius of convergence about that point, it is frequently possible to use the Sommerfeld method[3] to obtain an asymptotic expansion in ascending powers of $(1/\alpha)$. When these conditions are satisfied, the result of the Sommerfeld method may be expressed as[4]

$$F(\alpha) = \csc(\pi D) G(z), \tag{2}$$

where the operator $\pi \csc \pi D$ indicates the Laurent expansion of the cosecant about zero with

$$D = \frac{d}{dz}\bigg|_{z=\alpha} \quad \text{and} \quad \frac{1}{D} = \int_0^\alpha dz. \tag{3}$$

Occasionally, a calculation at nonzero temperature requires the evaluation of a Fermi–Dirac integral, for which the integrand does not satisfy both conditions (i) and (ii). For example, $G(z)$ may oscillate rapidly or may have an inopportunely situated branch point. In such circumstances ordinary methods may fail. A useful technique involves converting the real integral into a complex integral by substituting the representation for the Fermi function:

$$f^-(z,\alpha) = (e^{z-\alpha} + 1)^{-1} = \frac{1}{2\pi i} \int_{b-i\infty}^{b+i\infty} \pi \csc(\pi t) e^{t(\alpha - z)} \, dt, \tag{4}$$

where $0 < b < 1$.

This representation is contained in an earlier analysis of the Fermi–Dirac integral[5] and has been reiterated and used by others in specific cases.[6–8] It has also been shown[9] that this representation possesses enough algebraic flexibility to permit the reduction of Fermi integrals to tabulated Laplace transforms and their inverses or, at worst, to a tractable exercise in residue theory. In some cases "exact" evaluation in terms of tabulated mathematical functions is possible; for example, the familiar integral

$$F_\nu(\alpha) = \frac{1}{\Gamma(\nu + 1)} \int_0^\infty \frac{z^\nu}{e^{a-\alpha} + 1} \, dz \tag{5}$$

can be expressed in terms of confluent hypergeometric functions.[5,9] This common integral, of course, satisfied condition (ii) above and can, therefore, be evaluated in the high-degeneracy limit by Eq. (2).

The principal intent of this paper is to evaluate two integrals which are of physical interest and to illustrate several techniques that may be employed on integrals for which the Sommerfeld approximation does not apply. The first of these is the first-order exchange integral from the many-fermion problem, which will

* Work performed under the auspices of the U.S. Atomic Energy Commission.

[1] A. H. Wilson, *Theory of Metals* (Cambridge University Press, Cambridge, England, 1960).
[2] C. Bloch and C. De Dominicis, Nucl. Phys. **7**, 459 (1958); **10**, 181 (1959); **10**, 509 (1959).
[3] A. H. Sommerfeld and H. Bethe, in *Handbuch der Physik, Vol. 24*, S. Flügge, Ed. (Springer-Verlag, Berlin, 1933), p. 346.
[4] H. E. DeWitt (unpublished notes).
[5] R. Dingle, Appl. Sci. Res. **B6**, 225 (1956).
[6] J. M. Luttinger, Phys. Rev. **121**, 1251 (1961).
[7] R. J. Swenson, Phys. Letters **26A**, 632 (1968).
[8] A. Wasserman and H. E. DeWitt, J. Phys. Chem. Solids **29**, 2113 (1968).
[9] A. Wasserman, Phys. Letters **27A**, 360 (1968).

be treated in the high-degeneracy limit. The second, which can be evaluated "exactly," is the free-energy integral of the noninteracting electron gas in a magnetic field.

2. THE EXCHANGE INTEGRAL

The exchange integral with the Coulomb interaction frequently appears in the literature as

$$\mathfrak{I}_{ex}(\alpha) = \frac{1}{\pi^3}\int d\mathbf{x}\int d\mathbf{y}\frac{f^-(x^2)f^-(y^2)}{|\mathbf{x}-\mathbf{y}|^2} \tag{6}$$

with

$$x = |\mathbf{x}|, \quad y = |\mathbf{y}|,$$

and has previously been evaluated in part by recasting it as[10]

$$\mathfrak{I}_{ex}(\alpha) = 2\int_{-\infty}^{\alpha} d\alpha'[F_{-\frac{1}{2}}(\alpha')]^2. \tag{7}$$

We can begin the evaluation of $\mathfrak{I}_{ex}(\alpha)$ along the lines suggested in Ref. 9 by performing the obvious elementary integrations to obtain

$$\begin{aligned}\mathfrak{I}_{ex}(\alpha) &= \frac{8}{\pi}\int_0^\infty dx x f^-(x^2)\int_0^\infty dy y f^-(y^2)\log\left|\frac{x+y}{x-y}\right| \\ &= \frac{8}{\pi}\int_0^\infty dx x f^-(x) J(x,\alpha).\end{aligned} \tag{8}$$

Here we encounter a branch point which makes the Sommerfeld approximation inadequate. The y integration is straightforwardly approached by isolating that term which contains the annoying branch point. Thus, we have

$$\begin{aligned}J(x,\alpha) &= 2\int_0^\infty dy y f^-(y^2)\log|x+y| \\ &\quad -\int_0^\infty dy y f^-(y^2)\log|x^2-y^2| \\ &= J_1 - J_2.\end{aligned} \tag{9}$$

The first term of Eq. (9) is (at high degeneracy), by Eq. (2),

$$\begin{aligned}J_1 &= \tfrac{1}{2}\alpha\log\alpha + \alpha(\xi - \tfrac{1}{2}) + \alpha\xi^2\log\xi + \alpha(1-\xi^2) \\ &\quad \times \log(1+\xi) + \pi^2/12\alpha(\xi+1) + O(1/\alpha^3),\end{aligned} \tag{10}$$

where

$$\xi = x/\alpha^{\frac{1}{2}}.$$

After a change of integration variable, the representation Eq. (4) is inserted into J_2 to give

$$\begin{aligned}J_2 &= \tfrac{1}{2}\alpha\int_0^\infty \frac{dz}{e^{\alpha(z-1)}+1}\log|\xi^2 - z| \\ &= \frac{\alpha}{4\pi i}\int_{b-i\infty}^{b+i\infty} ds\pi\csc(\pi s)e^{\alpha s}\int_0^\infty e^{-\alpha z s}\log|\xi^2 - z|\,dz.\end{aligned} \tag{11}$$

[10] H. E. DeWitt, J. Nucl. Energy C2, 27 (1961).

The z integral is a tabulated Laplace transform[11] and brings us to

$$J_2 = \frac{1}{4\pi i}\int_{b-i\infty}^{b+i\infty} ds\pi\frac{\csc\pi s}{s}e^{\alpha s}[\log\xi^2 - e^{-\alpha\xi^2 s}E_i^*(\alpha\xi^2 s)], \tag{12}$$

where $E_i^*(z)$ is the exponential integral. The first term in J_2 is equal to $\alpha\log\xi[1+\log(1+e^{-\alpha})]$. The second term is handled by introducing a Mittag–Leffler expansion[12]

$$\pi\csc\pi s = \frac{1}{s} + 2s\sum_{n=1}^{\infty}\frac{(-1)^n}{s^2-n^2}. \tag{13}$$

The explicit use of a Mittag–Leffler expansion for this and other meromorphic functions is a key algebraic step which frequently permits evaluation of Fermi–Dirac integrals in terms of tabulated mathematical functions.

We now have

$$\begin{aligned}J_2 = \alpha\Bigg(&\log\xi\{1+\log[1+\exp(-\alpha)]\} \\ &-\frac{1}{4\pi i}\int_{b-i\infty}^{b+i\infty} ds\frac{\exp\alpha s(1-\xi^2)}{s^2}E_i^*(\alpha\xi^2 s) \\ &-\frac{1}{2\pi i}\sum_{n=1}^{\infty}(-1)^n\int_{b-i\infty}^{b+i\infty} ds\frac{\exp\alpha s(1-\xi^2)}{(s^2-n^2)}E_i^*(\alpha\xi^2 s)\Bigg).\end{aligned} \tag{14}$$

In the first of these integrals let

$$s^{-1} = \int_0^\infty e^{-su}\,du, \quad \mathrm{Re}\,s > 0.$$

The s integration becomes a tabulated inverse Laplace transform,[11,13] the subsequent u integration being elementary. We obtain

$$\begin{aligned}&\frac{1}{4\pi i}\int_{b-i\infty}^{b+i\infty} ds\frac{\exp\alpha s(1-\xi^2)}{s^2}E_i^*(\alpha\xi^2 s) \\ &\qquad = \frac{1}{2}\left[1 + (\xi^2-1)\log\frac{|1-\xi^2|}{\xi^2}\right].\end{aligned} \tag{15}$$

The last integral of Eq. (14) is

$$\begin{aligned}&\frac{1}{2\pi i}\int_{b-i\infty}^{b+i\infty} ds\frac{\exp\alpha s(1-\xi^2)}{(s^2-n^2)}E_i^*(\alpha\xi^2 s) \\ &\quad = \frac{1}{2n}\Bigg[\frac{1}{2\pi i}\int_{b-i\infty}^{b+i\infty} ds\frac{\exp\alpha s(1-\xi^2)}{(s+n)}E_i^*(\alpha\xi^2 s) \\ &\qquad + \int_{b-i\infty}^{b+i\infty} ds\frac{\exp\alpha s(1-\xi^2)}{(n-s)}E_i^*(\alpha\xi^2 s)\Bigg].\end{aligned} \tag{16}$$

[11] Bateman Manuscript Project, *Tables of Integral Transforms*, A. Erdelyi, Ed. (McGraw-Hill Book Company, Inc., New York, 1954), Vol. I.

[12] E.g., B. Fuchs and B. Shabat, *Functions of a Complex Variable* (Pergamon Press Ltd., Oxford, England, 1964).

[13] G. E. Roberts and H. Kaufman, *Tables of Laplace Transforms* (W. B. Saunders and Co., Philadelphia, 1966).

Because $n > \operatorname{Re} s$, we can replace

$$(n \pm s)^{-1} = \int_0^\infty du \exp[-(n \pm s)u]$$

and perform integrations by parts to reduce the complex integral to a representation of the step function. The remaining integrals (ignoring terms of order $e^{-\alpha}$) are familiar representations of the exponential integral.

Thus, with modest effort, we have

$$\begin{aligned} J(x,\alpha) = {} & \alpha\xi\left(1 + \frac{1-\xi^2}{2\xi}\log\left|\frac{1+\xi}{1-\xi}\right|\right) \\ & + \frac{\pi^2}{12\alpha}(\xi+1)^{-1} + \operatorname{sgn}(1-\xi^2)\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{2n} \\ & \times [\exp(n\alpha|1-\xi^2|)E_i(-n\alpha|1-\xi^2|) \\ & - \exp(-n\alpha|1-\xi^2|)E_i^*(n\alpha|1-\xi^2|)] \\ & + O\left(\frac{1}{\alpha^3}\right) \end{aligned} \tag{17}$$

with $\operatorname{sgn}(x) =$ sign of x. $J(x,\alpha)$ itself is of some physical interest. If we take $\xi^2 = G^2/4\mu$, where G is a vector of the reciprocal lattice and μ is the chemical potential, then $J(x,\alpha)$ appears in lowest-order electron-periodic potential correction to the free energy of a noninteracting electron gas.[8]

The remaining x integration of Eq. (8) is substantially more tedious, but is, nonetheless, a repetition of the steps taken above. Again, the separation is made between those terms for which the Sommerfeld approximation is valid and those that have branch points.

With the change of the variable integration, $z = \xi^2$, we have

$$\begin{aligned} \mathfrak{J}_{\mathrm{ex}}(\alpha) = {} & \frac{7\alpha^2}{2\pi} + \pi(1 - \tfrac{4}{3}\log 2) \\ & - \frac{2\alpha^2}{\pi}\int_0^\infty dz \frac{|1-z|\log|1-z|}{e^{\alpha(z-1)}+1} \\ & + \frac{2\alpha}{\pi}\operatorname{sgn}(1-z)\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}\int_0^\infty \frac{dz}{e^{\alpha(z-1)}+1} \\ & \times [e^{n\alpha|1-z|}E_i(-n\alpha|1-z|) - e^{-n\alpha|1-z|} \\ & \times E_i^*(n\alpha|1-z|)] + O\left(\frac{1}{\alpha^2}\right), \end{aligned} \tag{18}$$

where Eq. (2) has been applied to the former type of term and the latter is displayed explicitly. The techniques outlined above reduce the remaining collection of integrals to an easy application of residue theory.

We begin with the first of these by inserting the representation of Eq. (4) to give

$$\begin{aligned} & -\frac{2\alpha^2}{\pi}\frac{1}{2\pi i}\int_{b-i\infty}^{b+i\infty} dt\pi \csc(\pi t) \\ & \times \int_0^\infty dz(1-z)\log|1-z|\, e^{-t\alpha(z-1)}. \end{aligned} \tag{19}$$

The real integral consists of standard forms, and the contour integral that remains is

$$-\frac{2\alpha^2}{\pi}\frac{1}{2\pi i}\int_{b-i\infty}^{b+i\infty} dt\pi \frac{\csc \pi t}{(t\alpha)^2}[E_i^*(t\alpha) - e^{t\alpha}]. \tag{20}$$

Use of Eq. (13) leaves the evaluation of the integrals

$$\begin{aligned} & -\frac{2}{\pi}\frac{1}{2\pi i}\left(\int_{b-i\infty}^{b+i\infty} dt \frac{[E_i^*(t\alpha) - e^{t\alpha}]}{t^3} - 2\sum_{n=1}^{\infty}(-1)^{n+1} \right. \\ & \left. \times \int_{b-i\infty}^{b+i\infty} dt \frac{[E_i^*(t\alpha) - e^{t\alpha}]}{t(t^2-n^2)}\right) = Q_1 + Q_2. \end{aligned} \tag{21}$$

Both of these are conveniently done by deforming the contour to lie along the cut on the negative real axis with an indentation of radius ϵ about the origin. Letting $\epsilon \to 0$ and using the fact that

$$E_i^*(z) = \gamma + \log z + \sum_{n=1}^{\infty}\frac{z^n}{n!\,n}, \tag{22}$$

where $\gamma =$ Euler's constant, we get

$$Q_1 = \alpha^2/2\pi \tag{23}$$

and

$$\begin{aligned} Q_2 = {} & -\tfrac{1}{3}\pi\log\alpha - \tfrac{1}{3}\pi(\gamma - 1) \\ & - \frac{4}{\pi}\sum_{n=1}^{\infty}(-1)^{n+1}\frac{\log n}{n^2} + O(e^{-\alpha}). \end{aligned} \tag{24}$$

The last of the integrals in Eq. (18) are also easily evaluated by use of Eq. (4). We obtain real integrals that are integrated by parts, leaving

$$Q_3 = \frac{4}{\pi}\frac{1}{2\pi i}\int_{b-i\infty}^{b+i\infty} dt \sum_{n=1}^{\infty}(-1)^{n+1}\frac{\pi\csc\pi t}{(t^2-n^2)}E_i^*(\alpha t). \tag{25}$$

We can either expand $\pi \csc \pi t$ once more, or we can sum over n. Employing the latter, we get

$$\begin{aligned} Q_3 = {} & \frac{2}{\pi}\frac{1}{2\pi i}\left[\int_{b-i\infty}^{b+i\infty} dt \frac{\pi\csc\pi t}{t^2}E_i^*(\alpha t) \right. \\ & \left. - \int_{b-i\infty}^{b+i\infty} dt \frac{(\pi\csc\pi t)^2}{t}E_i^*(\alpha t)\right]; \end{aligned} \tag{26}$$

whereas, by expanding $\pi \csc \pi t$, we introduce integrals tabulated in standard works and eventually obtain

$$\begin{aligned} Q_3 = {} & -\tfrac{1}{3}\pi\log\alpha - \frac{4}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\log n - \tfrac{1}{3}\pi\gamma + \tfrac{2}{3}\pi \\ & + \frac{8}{\pi}\sum_{n=1}^{\infty}\sum_{\substack{k=1 \\ k\neq n}}^{\infty}(-1)^{n+k}\frac{\log k/n}{k^2-n^2} + O(e^{-\alpha}). \end{aligned} \tag{27}$$

If we proceed, instead, from Eq. (26) and use the Mittag-Leffler expansion

$$(\pi \csc \pi t)^2 = \frac{1}{t^2} + 2\sum_{n=1}^{\infty} \frac{(t^2+n^2)}{(t^2-n^2)^2}, \tag{28}$$

we encounter contour integrals that are readily done by well-known methods and obtain

$$Q_3 = \frac{4}{\pi}\Big[\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}(\gamma + \log\alpha + \log n) - \sum_{n=1}^{\infty}\frac{1}{n^2}(\alpha - 1 + \log\alpha + \log n)\Big]. \tag{29}$$

The slowly converging sum $\sum_{n=1}^{\infty} \log n/n^2$ can be converted (see Appendix) to a more rapidly converging form:

$$\sum_{n=1}\frac{\log n}{n^2} = 2\sum(-1)^{n+1}\frac{\log n}{n^2} + \tfrac{1}{6}\pi^2\log 2. \tag{30}$$

If we collect the various terms, the exchange integral is then

$$\mathfrak{I}_{ex}(\alpha) = C_1\alpha^2 + C_2\log\alpha + C_3 + O(\alpha^{-2}),$$

where

$$C_1 = 4/\pi, \quad C_2 = \tfrac{2}{3}\pi, \tag{31}$$

and

$$C_3 = 2\pi - \tfrac{4}{3}\pi\log 2 - \tfrac{2}{3}\pi\gamma + (4/\pi)\zeta'(2) = 2\pi - 2\pi\log 2 - \tfrac{2}{3}\pi\gamma - (8/\pi)\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\log n, \tag{32}$$

where $\zeta'(s)$ is the derivative of the Riemann zeta function and

$$\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\log n = -0.10131658.$$

The result, Eq. (31), contains explicitly the term that gives rise to the long-known logarithmic divergence of the exchange contribution to the specific heat of an electron gas.[14] Equally well known is the fact that this divergence must be eliminated by complete summation of all the ring diagrams.[15]

3. THE FREE-ENERGY INTEGRAL

The free energy of an electron gas in a magnetic field H can be written

$$F = \zeta N - \frac{m\omega}{2\pi^2\beta\hbar}\int_{-\infty}^{\infty}dk_z \times\sum_{\eta=0}^{\infty}[\ln(1+\exp\{\beta[\zeta-\epsilon(\eta,k_z)]\})] = \zeta N - \mathfrak{D}(\beta,\omega,\zeta), \tag{33}$$

where ζ is the chemical potential, N is the average number of particles in the system, $\omega = (eH/mc)$, and

$$\epsilon(\eta, k_z) = \hbar\omega(\eta + \tfrac{1}{2}) + \frac{\hbar^2k_z^2}{2m}.$$

We start with an integration by parts to bring it into the form of a Fermi-Dirac integral:

$$\mathfrak{D}(\beta,\omega,\zeta) = \frac{\hbar\omega}{2\pi^2}\int_{-\infty}^{\infty}dk_z\sum_{\eta=0}^{\infty}\frac{k_z^2}{\{\beta[\epsilon(\eta,k_z)-\zeta]\}+1}. \tag{34}$$

At this point we use the integral representation of Eq. (4) to give

$$\mathfrak{D}(\beta,\omega,\zeta) = \frac{\hbar\omega}{2\pi^2}\frac{1}{2\pi i} \times\int_{b-i\infty}^{b+i\infty}ds\pi\csc(\pi s)e^{\alpha s}\exp(-\beta\hbar\omega s/2) \times\int_{-\infty}^{\infty}dk_z k_z^2\exp(-\beta s\hbar^2k_z^2/2m) \times\sum_{\eta=0}^{\infty}\exp(-\beta s\hbar\omega\eta). \tag{35}$$

In this form the k_z integration and sum over η are trivial, and the result is

$$\mathfrak{D}(\beta,\omega,\zeta) = \frac{\delta}{2\pi i}\int_{b-i\infty}^{b+i\infty}\frac{\pi\csc\pi s}{s^{\frac{3}{2}}\sinh\lambda n}e^{\alpha s}, \tag{36}$$

where

$$\delta = \frac{\hbar\omega}{4}\left(\frac{2m}{\pi\beta\hbar^2}\right)^{\frac{3}{2}} \quad\text{and}\quad \lambda = \frac{\beta\hbar\omega}{2}.$$

The integral in Eq. (36) can be expressed as a sum of tabulated mathematical functions, as will now be shown, where the important step is again a Mittag-Leffler expansion:

$$\frac{\pi\csc\pi s}{\sinh\lambda s} = \frac{1}{\lambda s^2} + \sum_{\substack{n=-\infty\\ n\neq 0}}^{\infty}(-1)^n \times\left[\frac{1}{(s-n)\sinh\lambda n} - \frac{\pi i}{\lambda\left(s-\frac{n\pi i}{\lambda}\right)\sinh\frac{n\pi^2}{\lambda}}\right]. \tag{37}$$

[14] J. Bardeen, Phys. Rev. **50**, 1098 (1936).
[15] M. Gell-Mann, Phys. Rev. **106**, 369 (1957).

Thus,

$$\mathfrak{D}(\beta,\omega,\zeta)=\delta\Bigg[\frac{1}{2\pi\lambda i}\int_{b-i\infty}^{b+i\infty}\frac{e^{\alpha s}}{s^{\frac{1}{2}}}\,ds + \sum_{\substack{n=-\infty\\ n\neq 0}}^{\infty}(-1)^n\Bigg[\frac{1}{2\pi i\sinh\lambda n}\times\int_{b-i\infty}^{b+i\infty}ds\,\frac{e^{\alpha s}}{s^{\frac{3}{2}}(s-n)}-\frac{1}{2\lambda\sinh(n\pi^2/\lambda)}\times\int_{b-i\infty}^{b+i\infty}ds\,\frac{e^{\alpha s}}{s^{\frac{3}{2}}\left(s-\frac{n\pi i}{\lambda}\right)}\Bigg]\Bigg]. \tag{38}$$

The first integral in Eq. (38) is the reciprocal of a gamma function,

$$\frac{1}{2\pi i}\int_{b-i\infty}^{b+i\infty}ds\,\frac{e^{\alpha s}}{s^{\frac{7}{2}}}=\frac{\alpha^{\frac{5}{2}}}{\Gamma(\frac{7}{2})}; \tag{39}$$

the second is a representation of confluent hypergeometric functions,

$$\frac{1}{2\pi i}\int_{b-i\infty}^{b+i\infty}ds\,\frac{e^{\alpha s}}{s^{\frac{3}{2}}(s-n)} = -\frac{\alpha^{\frac{3}{2}}}{\Gamma(\frac{3}{2})}U(1,\tfrac{5}{2};\alpha n), \qquad n>0,$$
$$= \frac{\alpha^{\frac{3}{2}}}{\Gamma(\frac{5}{2})}e^{-\alpha n}\,{}_1F_1(\tfrac{3}{2},\tfrac{5}{2};\alpha n), \qquad n<0, \tag{40}$$

where

$$U(a,b;x) = \frac{\Gamma(1-b)}{\Gamma(1+a-b)}\,{}_1F_1(a,b;x) + \frac{\Gamma(b-1)}{\Gamma(a)}X^{1-b}\,{}_1F_1(1+a-b,2-b;x).$$

The last integral of Eq. (38) may also be written in terms of confluent hypergeometric functions[13]:

$$\frac{1}{2\pi i}\int_{b-i\infty}^{b+i\infty}ds\,\frac{e^{\alpha s}}{s^{\frac{3}{2}}\left(s-\frac{n\pi i}{\lambda}\right)}=\frac{\alpha^{\frac{3}{2}}}{\Gamma(\frac{5}{2})}\,{}_1F_1\left(1,\tfrac{5}{2};\frac{\alpha n\pi i}{\lambda}\right). \tag{41}$$

An alternative form may be obtained by replacing

$$\frac{1}{s^{\frac{3}{2}}}=\frac{1}{\Gamma(\frac{3}{2})}\int_0^\infty e^{-xs}x^{\frac{1}{2}}\,dx.$$

Integrating s then gives

$$\frac{1}{2\pi i}\int_{b-i\infty}^{b+i\infty}ds\,\frac{e^{\alpha s}}{s^{\frac{3}{2}}\left(s-\frac{n\pi i}{\lambda}\right)}=\int_0^\alpha dx\,x^{\frac{1}{2}}\exp\left[\frac{n\pi i}{\lambda}(\alpha-x)\right]. \tag{42}$$

We can now write

$$\mathfrak{D}(\beta,\omega,\zeta)=\delta\Bigg\{\frac{\alpha^{\frac{5}{2}}}{\lambda\Gamma(\frac{7}{2})}+\alpha^{\frac{3}{2}}\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{\sinh\lambda n}\times\left[\frac{U(1,\frac{5}{2};\alpha n)}{\Gamma(\frac{3}{2})}+\frac{e^{-\alpha n}}{\Gamma(\frac{5}{2})}\,{}_1F_1(\tfrac{3}{2},\tfrac{3}{2};\alpha n)\right] + 2\pi\sum_{n=1}^{\infty}\frac{(-1)^n}{\lambda\sinh(n\pi^2/\lambda)}\times\int_0^\alpha dx\,x^{\frac{1}{2}}\sin\left(\frac{n\pi}{\lambda}\right)(\alpha-x)\Bigg\}. \tag{43}$$

An integration by parts recasts the remaining integral in Eq. (43) into the Fresnel integral form

$$\int_0^\alpha dx\,x^{\frac{1}{2}}\sin\left(\frac{n\pi}{\lambda}\right)(\alpha-x) = \frac{\lambda\alpha^{\frac{1}{2}}}{n\pi}-\frac{\lambda}{2n\pi}\left(\frac{2\lambda}{n}\right)^{\frac{1}{2}}\times[\sin(n\pi\alpha/\lambda)S(n\pi\alpha/\lambda)+\cos(n\pi\alpha/\lambda)C(n\pi\alpha/\lambda)] \tag{44}$$

with the Fresnel integrals

$$C(z)=\frac{1}{(2\pi)^{\frac{1}{2}}}\int_0^z\frac{\cos t}{t^{\frac{1}{2}}}\,dt, \qquad S(z)=\frac{1}{(2\pi)^{\frac{1}{2}}}\int_0^z\frac{\sin t}{t^{\frac{1}{2}}}\,dt. \tag{45}$$

Eq. (43) may finally be written

$$\mathfrak{D}(\beta,\omega,\zeta)=\delta\Bigg\{\frac{\alpha^{\frac{5}{2}}}{\lambda\Gamma(\frac{7}{2})}+\alpha^{\frac{3}{2}}\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{\sinh\lambda n}\times\left[\frac{U(1,\frac{5}{2};\alpha n)}{\Gamma(\frac{3}{2})}+\frac{e^{-\alpha n}}{\Gamma(\frac{5}{2})}\,{}_1F_1(\tfrac{3}{2},\tfrac{5}{2};\alpha n)\right] + 2\alpha^{\frac{1}{2}}\sum_{n=1}^{\infty}\frac{(-1)^n}{n\sinh(n\pi^2/\lambda)}(2\lambda)^{\frac{1}{2}}\times\sum_{n=1}^{\infty}\frac{(-1)^n}{n^{\frac{1}{2}}\sinh(n\pi^2/\lambda)}\times[\sin(n\pi\alpha/\lambda)S(n\pi\alpha/\lambda) + \cos(n\pi\alpha/\lambda)C(n\pi\alpha/\lambda)]\Bigg\}. \tag{46}$$

In the region of high degeneracy, $\alpha \gg 1$, $\alpha \gg \lambda$, we have

$$U(1, \tfrac{5}{2}; \alpha n) \to (\alpha n)^{-1},$$

$${}_1F_1(\tfrac{3}{2}, \tfrac{5}{2}; \alpha n) \to \frac{\Gamma(\frac{5}{2})}{\Gamma(\frac{3}{2})} e^{\alpha n}(\alpha n)^{-1},$$

$$C\left(\frac{n\pi\alpha}{\lambda}\right) \to \tfrac{1}{2},$$

$$S\left(\frac{n\pi\alpha}{\lambda}\right) \to \tfrac{1}{2},$$

and Eq. (46) becomes

$$\mathfrak{D}(\lambda, \alpha) = \delta\left\{\frac{\alpha^{\frac{5}{2}}}{\lambda\Gamma(\frac{7}{2})} + 2\alpha^{\frac{1}{2}} \times \left[\frac{1}{\Gamma(\frac{3}{2})} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n \sinh \lambda n} + \sum \frac{(-1)^{n+1}}{n \sinh (n\pi^2/\lambda)}\right] + \lambda^{\frac{1}{2}} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^{\frac{3}{2}} \sinh (n\pi^2/\lambda)} \cos\left(\frac{n\pi\alpha}{\lambda} \pm \tfrac{1}{4}\pi\right)\right\}. \quad (47)$$

Even in the region of partial degeneracy, the condition $\alpha \gg \lambda$ prevails, and the oscillatory sum on the right of Eq. (47) persists.

APPENDIX

We use the fact that

$$\sum_{n=1}^{\infty} \frac{\log n}{n^2} = \frac{d}{ds} \zeta(s)\Big|_{s=2},$$

where $\zeta(s)$ is the Riemann zeta function, where

$$\zeta(s) = \frac{1}{(1 - 2^{1-s})} \frac{1}{\Gamma(s)} \int_0^{\alpha} \frac{x^{s-1}}{e^x + 1}\, dx.$$

Differentiation and integration yield

$$\sum_{n=1}^{\infty} \frac{\log n}{n^2} = 2\sum_{n=1}^{\infty} (-1)^{n+1} \frac{\log n}{n^2} + \tfrac{1}{6}\pi^2 \log 2.$$

Furthermore, by comparing Eq. (27) and Eq. (29), we arrive at the unusual summation formula

$$\sum_{\substack{k=1 \\ k \neq n}}^{\infty} \sum_{n=1}^{\infty} (-1)^k \frac{\log (k/n)}{k^2 - n^2} = -\tfrac{1}{12}\pi^2 \log 2.$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Parametrization of Crossing Symmetric Amplitudes*

RALPH Z. ROSKIES
Physics Department, Yale University, New Haven, Connecticut 06520

(Received 26 May 1969)

The crossing equations of Balachandran and Nuyts are solved to give the most general expansion of the π^0–π^0 amplitude consistent with crossing symmetry and in which the partial wave expansion is manifest. We find that crossing symmetry imposes $2L + 1$ constraints on the Lth partial wave if the lower waves are known. The simplest such constraint (in the units were $4m_\pi^2 = 1$) is

$$\int_0^1 (1 - s)(3s - 1) f_0(s)\, ds = 0,$$

where $f_0(s)$ is the s-wave π^0–π^0 amplitude.

I. INTRODUCTION

During the past several years, attempts have been made to find examples of amplitudes which embody the general principles of analyticity, unitarity, and crossing. Until recently, the emphasis seemed to be on incorporating the first two, with crossing symmetry usually mutilated; for example, the N/D formalism applied to each partial-wave amplitude. Recently, however, the importance of crossing symmetry has been emphasized, and some success[1] has been achieved in attempts to incorporate analyticity and crossing.

In a parallel development, which also emphasizes crossing symmetry, Balachandran and Nuyts[2] considered the following problem: Could one find an expansion of the amplitude $F(s, t, u)$ in terms of *known* functions of s and t with unknown coefficients but which automatically satisfied crossing symmetry and in which the partial wave expansion was obvious?

* This work (Yale Report 2726-549) is supported in part by the U.S. Atomic Energy Commission under contract AT(30-1)-2726.

[1] G. Veneziano, Nuovo Cimento **57A**, 190 (1968).
[2] A. P. Balachandran and J. Nuyts, Phys. Rev. **172**, 1821 (1968).

In the region of high degeneracy, $\alpha \gg 1$, $\alpha \gg \lambda$, we have

$$U(1, \tfrac{5}{2}; \alpha n) \to (\alpha n)^{-1},$$

$${}_1F_1(\tfrac{3}{2}, \tfrac{5}{2}; \alpha n) \to \frac{\Gamma(\frac{5}{2})}{\Gamma(\frac{3}{2})} e^{\alpha n}(\alpha n)^{-1},$$

$$C\left(\frac{n\pi\alpha}{\lambda}\right) \to \tfrac{1}{2},$$

$$S\left(\frac{n\pi\alpha}{\lambda}\right) \to \tfrac{1}{2},$$

and Eq. (46) becomes

$$\mathfrak{D}(\lambda, \alpha) = \delta\left\{\frac{\alpha^{\frac{5}{2}}}{\lambda\Gamma(\frac{7}{2})} + 2\alpha^{\frac{1}{2}} \times \left[\frac{1}{\Gamma(\frac{3}{2})} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n \sinh \lambda n} + \sum \frac{(-1)^{n+1}}{n \sinh (n\pi^2/\lambda)}\right] + \lambda^{\frac{1}{2}} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^{\frac{3}{2}} \sinh (n\pi^2/\lambda)} \cos\left(\frac{n\pi\alpha}{\lambda} \pm \tfrac{1}{4}\pi\right)\right\}. \quad (47)$$

Even in the region of partial degeneracy, the condition $\alpha \gg \lambda$ prevails, and the oscillatory sum on the right of Eq. (47) persists.

APPENDIX

We use the fact that

$$\sum_{n=1}^{\infty} \frac{\log n}{n^2} = \frac{d}{ds} \zeta(s)\Big|_{s=2},$$

where $\zeta(s)$ is the Riemann zeta function, where

$$\zeta(s) = \frac{1}{(1 - 2^{1-s})} \frac{1}{\Gamma(s)} \int_0^{\alpha} \frac{x^{s-1}}{e^x + 1}\, dx.$$

Differentiation and integration yield

$$\sum_{n=1}^{\infty} \frac{\log n}{n^2} = 2\sum_{n=1}^{\infty} (-1)^{n+1} \frac{\log n}{n^2} + \tfrac{1}{6}\pi^2 \log 2.$$

Furthermore, by comparing Eq. (27) and Eq. (29), we arrive at the unusual summation formula

$$\sum_{\substack{k=1 \\ k \neq n}}^{\infty} \sum_{n=1}^{\infty} (-1)^k \frac{\log (k/n)}{k^2 - n^2} = -\tfrac{1}{12}\pi^2 \log 2.$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Parametrization of Crossing Symmetric Amplitudes*

RALPH Z. ROSKIES
Physics Department, Yale University, New Haven, Connecticut 06520

(Received 26 May 1969)

The crossing equations of Balachandran and Nuyts are solved to give the most general expansion of the π^0–π^0 amplitude consistent with crossing symmetry and in which the partial wave expansion is manifest. We find that crossing symmetry imposes $2L + 1$ constraints on the Lth partial wave if the lower waves are known. The simplest such constraint (in the units were $4m_\pi^2 = 1$) is

$$\int_0^1 (1 - s)(3s - 1) f_0(s)\, ds = 0,$$

where $f_0(s)$ is the s-wave π^0–π^0 amplitude.

I. INTRODUCTION

During the past several years, attempts have been made to find examples of amplitudes which embody the general principles of analyticity, unitarity, and crossing. Until recently, the emphasis seemed to be on incorporating the first two, with crossing symmetry usually mutilated; for example, the N/D formalism applied to each partial-wave amplitude. Recently, however, the importance of crossing symmetry has been emphasized, and some success[1] has been achieved in attempts to incorporate analyticity and crossing.

In a parallel development, which also emphasizes crossing symmetry, Balachandran and Nuyts[2] considered the following problem: Could one find an expansion of the amplitude $F(s, t, u)$ in terms of *known* functions of s and t with unknown coefficients but which automatically satisfied crossing symmetry and in which the partial wave expansion was obvious?

* This work (Yale Report 2726-549) is supported in part by the U.S. Atomic Energy Commission under contract AT(30-1)-2726.

[1] G. Veneziano, Nuovo Cimento **57A**, 190 (1968).
[2] A. P. Balachandran and J. Nuyts, Phys. Rev. **172**, 1821 (1968).

Their result was as follows: One could write the amplitude (for spinless equal-mass scattering) as

$$F(s,t,u)=\sum_{n=0}^{\infty}\sum_{l=0}^{\infty}2(n+l+1)(2l+1)a_n^l(1-s)^l \times P_n^{(2l+1,0)}(2s-1)P_l(z), \quad (1)$$

where the mass of the scattering particle has been taken as $\frac{1}{2}$, and $P_n^{(2l+1,0)}$ is the Jacobi polynomial. We have for s, t, and z the usual formula, namely,

$$z=1+2t/(s-1). \quad (2)$$

Since $P_l(z)$ is a complete set for $-1\leq z\leq 1$ and, for fixed l, $P_n^{(2l+1,0)}(2s-1)$ is complete in the region $0\leq s\leq 1$, it is not surprising that F can be expanded as in (1). The simplification arises because, in this basis, the condition $F(s,t,u)=F(t,u,s)$ reduces to a series of *finite-dimensional* linear relations on the a_n^l, which can be formulated as follows: For each σ, define the column vector (b^σ) with components

$$(b^\sigma)_l=a_{\sigma-l}^l, \quad l=0,1,\cdots,\sigma, \quad (3)$$

and define the $(\sigma+1)\times(\sigma+1)$ matrix

$$(G^\sigma)_{ll'}=\frac{(-1)^\sigma(\sigma!)^2(2l'+1)}{(\sigma-l')!\,(\sigma+l'+1)!} \times {}_4F_3(-l',l'+1,l-\sigma,-\sigma-l-1;\ -\sigma,-\sigma,1;1). \quad (4)$$

Then the amplitude is invariant under cyclic permutations of s, t, u, if

$$G^\sigma b^\sigma=b^\sigma \quad (5)$$

as a matrix equation.

In this paper, we shall find all solutions of (5) in two sets: those involving only even l and those involving only odd l. It can be shown directly that these solutions span the space of solutions of (5), but it can also be argued more simply. Any function $F(s,t,u)$ can be written as

$$F(s,t,u)=\tfrac{1}{2}[F(s,t,u)+F(s,u,t)] + \tfrac{1}{2}[F(s,t,u)-F(s,u,t)]. \quad (6)$$

If $F(s,t,u)$ is invariant under cyclic permutations, each of the bracketed terms on the right-hand side of (6) is cyclically invariant, and the first is totally symmetric under interchange of s, t, and u and thus involves only even l. The second is totally antisymmetric and involves only odd l. The symmetric solutions will be the physically meaningful ones for π^0–π^0 scattering, for which the amplitude is totally symmetric under all permutations of s, t, u.

Having found these solutions, we can write the most general crossing symmetric solution for π^0–π^0 scattering in terms of a set of completely arbitrary constants and known functions, in which the partial-wave expansion is manifest. This yields all the conditions on the partial-wave amplitudes in the region $0\leq s\leq 1$ which are imposed by, and which imply, crossing symmetry. The result is that if the partial waves for $l<L$ are given (L even), there are $2L+1$ constraints on the Lth partial wave. In the simplest case, for $L=0$, we find

$$\int_0^1(1-s)(3s-1)f_0(s)\,ds=0,$$

where $f_0(s)$ is the s-wave amplitude. This simple relation seems to have been overlooked in previous work on the rigorous results concerning the π^0–π^0 s wave.

Our expansion still has many drawbacks. First of all, the domain $0\leq s\leq 1$, $-1\leq z\leq 1$, in which the expansion is given is the Mandelstam triangle bounded by $s=0$, $t=0$, $u=0$, which is an unphysical region for all three channels. To get information on the physical amplitudes, it is necessary to continue the expansion into the physical regions. But it is difficult to find the constraints on the unknown coefficients which assure that the total amplitude has the proper analytic behavior. Secondly, it is also difficult to formulate the constraints of unitarity in the Mandelstam triangle. So it appears that we have gained crossing symmetry at the expense of analyticity and unitarity.

On the other hand, Martin[3] has obtained some remarkable results on the partial waves of the π^0–π^0 amplitude precisely in the region $0\leq s\leq 1$ (in Martin's language, $m_\pi=1$, so that he writes $0\leq s\leq 4$, whereas we have normalized to $m_\pi=\frac{1}{2}$ so that the region is $0\leq s\leq 1$). These results follow from the analytic properties of the amplitude and from the positivity of the absorptive part. (It is interesting to note that, neglecting isospin violations, π^0–π^0 scattering never satisfies elastic unitarity because the π^+–π^- threshold coincides with the elastic one.) It is hoped that these constraints of Martin will considerably reduce the arbitrariness of the original expansion. We hope to return to this question in a future publication.

We also defer to a later publication the problem of the crossing symmetric solution of the π–π amplitudes with isospin, which involves studying the constraints imposed on amplitudes which have mixed symmetry in the variables s, t, and u.

[3] A. Martin, Nuovo Cimento **58A**, 303 (1968); **63A**, 167 (1969), and references cited therein.

II. SOLUTION OF THE CROSSING PROBLEM FOR EVEN l

A. General Approach

We shall discover the most general solution of (5), subject to

$$(b^\sigma)_l = 0, \quad l \text{ odd}, \tag{7}$$

not by directly solving these equations but by a roundabout method which is based on the following observation: The π^0–π^0 amplitude is a totally symmetric function of s, t, and u and can therefore be written as a function of

$$\begin{aligned} x &= st + su + tu, \\ y &= stu, \\ z &= s + t + u. \end{aligned} \tag{8}$$

Because of the constraint

$$s + t + u = 1, \tag{9}$$

the most general symmetric amplitude can be written as an arbitrary function of x and y. A polynomial basis for the amplitude therefore consists of the functions

$$h_{mq}(x, y) = x^m y^q. \tag{10}$$

Since such a function is totally symmetric in s, t, and u, it can be written as

$$x^m y^q = \sum_{n=0}^{\infty} \sum_{l=0}^{\infty} 2(n + l + 1)(2l + 1)(a_{mq})_n^l (1 - s)^l \times P_n^{(2l+1,0)}(2s - 1) P_l(z) \tag{11}$$

$$= \sum_{l=0}^{\infty} \sum_{\sigma=l}^{\infty} 2(\sigma + 1)(2l + 1)(b_{mq}^\sigma)_l (1 - s)^l \times P_{\sigma-l}^{(2l+1,0)}(2s - 1) P_l(z), \tag{12}$$

with

$$(b_{mq}^\sigma)_l = (a_{mq})_{\sigma-l}^l \tag{13}$$

satisfying (5) and (7). Because $x^m y^q$ is a polynomial in s and t, the sum on the right-hand side of (11) involves only a finite number of n and l. In this case, there is no question of convergence of the sum, and (12) holds for all values of s and t. We now take the limit $s \to \infty$, z finite, and retain only the dominant terms on both sides of the equation. In this limit

$$t \approx \tfrac{1}{2} s(z - 1), \tag{14}$$

so that

$$x \approx -s^2[\tfrac{1}{4}(z^2 + 3)] \tag{15}$$

and

$$y \approx -s^3[\tfrac{1}{4}(z^2 - 1)]. \tag{16}$$

Thus, the left-hand side of (11) is, asymptotically,

$$s^{2m+3q}(-1)^{m+q}[\tfrac{1}{4}(z^2 + 3)]^m[\tfrac{1}{4}(z^2 - 1)]^q. \tag{17}$$

Since $P_n^{(2l+1,0)}$ is a polynomial of degree n, the term on the right-hand side of (12) is of order s^σ. Thus only the term

$$\sigma = 2m + 3q \tag{18}$$

survives and, for this value of σ, the right-hand side is, asymptotically,

$$s^\sigma \sum_{l=0}^{\sigma} (-1)^l \frac{(2\sigma + 1)!\, 2(\sigma + 1)(2l + 1)(b_{mq}^\sigma)_l P_l(z)}{(\sigma - l)!\,(\sigma + l + 1)!}, \tag{19}$$

where we have used the asymptotic form[4]

$$P_n^{(\alpha,0)}(x) \sim \frac{(\alpha + 2n)!}{n!\,(\alpha + n)!} (\tfrac{1}{2}x)^n. \tag{20}$$

Equating (17) and (19) and solving gives

$$(b_{mq}^\sigma)_l = \frac{(\sigma - l)!\,(\sigma + l + 1)!\,(-1)^{l+m+q}}{4(\sigma + 1)(2\sigma + 1)!} \times \int_{-1}^{1} P_l(z)[\tfrac{1}{4}(z^2 + 3)]^m[\tfrac{1}{4}(z^2 - 1)]^q\, dz. \tag{21}$$

It is clear that the integral vanishes for odd l, so that the factor $(-1)^l$ can be replaced by $+1$. This expression gives one solution of Eq. (5).

B. Solutions for Given σ

For a given σ, we can generate different solutions of Eq. (5) of the type given in (21) in as many ways as we can write

$$\sigma = 2m + 3q \tag{22}$$

for nonnegative integers m and q. In this section we show that these solutions are linearly independent and that they form a complete basis for the set of solutions subject to (7).

To study the number of nonnegative solutions to (22), we note that

$$m = \tfrac{1}{2}(\sigma - 3q) \tag{23}$$

must be an integer. Thus, q is odd if σ is, and q is even if σ is.

Therefore, we write

$$q = \sigma - 2p. \tag{24}$$

Then

$$m = 3p - \sigma. \tag{25}$$

Thus, we have a solution for each integer p satisfying

$$\tfrac{1}{3}\sigma \le p \le \tfrac{1}{2}\sigma, \tag{26}$$

and we label the corresponding solutions $(b^\sigma)_l$ by $(b_p^\sigma)_l$.

[4] See, e.g., *Higher Transcendental Functions*, A. Erdélyi, Ed. (McGraw-Hill Book Co., New York, 1953), Vol. I, p. 61; *ibid.* Vol. II, p. 170.

Thus, from (21), using (24) and (25), we have

$$(b_p^\sigma)_l = \frac{(\sigma - l)!\,(\sigma + l + 1)!\,(-1)^p}{4(\sigma + 1)(2\sigma + 1)!} \times \int_{-1}^{1} P_l(z)[\tfrac{1}{4}(z^2 + 3)]^{3p-\sigma}[\tfrac{1}{4}(z^2 - 1)]^{\sigma-2p}\,dz. \tag{27}$$

From (26), it is clear that if $\sigma = (0, 1, 2, 3, 4, 5)$, then the number of solutions is $n_\sigma = (1, 0, 1, 1, 1, 1)$, respectively. If

$$\sigma = 6a + b, \tag{28}$$

then the number of solutions is $a + n_b$.

To prove the linear independence of these solutions for fixed σ, suppose

$$\sum \lambda_p (b_p^\sigma)_l = 0, \quad 0 \le l \le \sigma, \tag{29}$$

that is,

$$\sum (-1)^p \lambda_p 4^{-p} \int_{-1}^{1} P_l(z)(z^2 + 3)^{3p-\sigma}(z^2 - 1)^{\sigma-2p}\,dz = 0, \quad 0 \le l \le \sigma. \tag{30}$$

The term $(z^2 + 3)^{3p-\sigma}(z^2 - 1)^{\sigma-2p}$ is a polynomial in z^2 of degree p, and so is a polynomial of at most degree σ in z [recall Eq. (26)]. Thus, the integral certainly vanishes for $l > \sigma$ and, by the completeness of the $P_l(z)$, we have

$$\sum_p (-1)^p \lambda_p 4^{-p}(z^2 + 3)^{3p-\sigma}(z^2 - 1)^{\sigma-2p} \equiv 0, \tag{31}$$

that is,

$$\sum_p (-1)^p \lambda_p [(z^2 - 1)/(z^2 + 3)]^\sigma \times [(z^2 + 3)^3/4(z^2 - 1)^2]^p \equiv 0, \tag{32}$$

that is,

$$[(z^2 - 1)/(z^2 + 3)]^\sigma \sum \lambda_p w^p \equiv 0, \tag{33}$$

where

$$w = -(z^2 + 3)^3/4(z^2 - 1)^2, \tag{34}$$

which is possible only if $\lambda_p = 0$. Thus, the b_p^σ are linearly independent.

To prove the completeness, suppose that we are given a solution $(b'^\sigma)_l$ of (5) and (7). Then, from the analysis of Ref. 2, we have that

$$F(s, t, u) = \sum_{l=0}^{\sigma} 2(\sigma + 1)(2l + 1)(b'^\sigma)_l (1 - s)^l \times P_{\sigma-l}^{(2l+1,0)}(2s - 1)P_l(z) \tag{35}$$

is a totally symmetric function of s, t, and u. Moreover, it is a polynomial in these variables and, therefore,

$$F(s, t, u) = \sum a_{mn} x^m y^n. \tag{36}$$

Comparing asymptotic terms in s, for fixed z in (35) and (36), we find that the dominant term in (36) must be of order s^σ; i.e., the dominant terms can be written as

$$\sum a_p x^{3p-\sigma} y^{\sigma-2p}. \tag{37}$$

Comparing asymptotic forms implies that

$$(b'^\sigma)_l = \sum a_p (b_p^\sigma)_l, \tag{38}$$

which establishes the completeness.

III. THE ANTISYMMETRIC CASE

We can apply the same arguments to an amplitude $G(s, t, u)$ which is totally antisymmetric under interchanges of s, t, and u. This rests on the observation that such a function can be written as

$$G(s, t, u) = (s - t)(t - u)(u - s)H(s, t, u), \tag{39}$$

where H is totally symmetric. Moreover, if G is holomorphic, then so is H, since G vanishes if two of its arguments are equal; thus,

$$G(s, t, u)/(s - t)(t - u)(u - s)$$

is also holomorphic. The basis for antisymmetric functions is, therefore,

$$(s - t)(t - u)(u - s)x^m y^q, \tag{40}$$

where x and y are defined as in (8). Carrying through exactly the same analysis, we find solutions

$$(\tilde{b}_p^\sigma)_l = \frac{(\sigma - l)!\,(\sigma + l + 1)!\,(-1)^p}{4(\sigma + 1)(2\sigma + 1)!} \times \int_{-1}^{1} P_l(z)[\tfrac{1}{4}(z^2 + 3)]^{3p-\sigma}[\tfrac{1}{4}(z^2 - 1)]^{\sigma-2p-1} \times \tfrac{1}{4}(z^2 - 9)z\,dz \tag{41}$$

with p integral, satisfying

$$\tfrac{1}{3}\sigma \le p \le \tfrac{1}{2}(\sigma - 1). \tag{42}$$

The number of solutions for $\sigma = 6a + b$ is

$$\tilde{n} = a + \tilde{n}_b, \tag{43}$$

where

$$\tilde{n}_b = (0, 0, 0, 1, 0, 1), \quad \text{if } b = (0, 1, 2, 3, 4, 5). \tag{44}$$

IV. SOME IMPLICATIONS

The most general crossing symmetric form for the π^0–π^0 amplitude is

$$F(s, t, u) = \sum_{l=0}^{\infty} \sum_{\sigma=l}^{\infty} \sum_{p=\{\frac{1}{3}\sigma\}}^{[\frac{1}{2}\sigma]} (2l + 1)c_p^\sigma (b_p^\sigma)_l (1 - s)^l \times P_{\sigma-l}^{(2l+1,0)}(2s - 1)P_l(z), \tag{45}$$

where

$\{\frac{1}{3}\sigma\}$ is the smallest integer $\geq \frac{1}{3}\sigma$,

$[\frac{1}{2}\sigma]$ is the largest integer $\leq \frac{1}{2}\sigma$,

the $(b_p^\sigma)_l$ are given by Eq. (27),

and

the (c_p^σ) are arbitrary constants.

The partial wave amplitudes are

$$f_l(s) = \sum_{\sigma=l}^{\infty} \sum_{p=\{\frac{1}{3}\sigma\}}^{[\frac{1}{2}\sigma]} c_p^\sigma (b_p^\sigma)_l (1-s)^l P_{\sigma-l}^{(2l+1,0)}(2s-1). \quad (46)$$

First we notice that, for $\sigma = 1$, there is no integer value of p, so that the term with $\sigma = 1$ is missing. Using the orthogonality properties[4] of the functions $P_{\sigma-l}^{(2l+1,0)}$ for given l and different σ, and noting that $\sigma = 1$ occurs only in the s wave, we find[5]

$$\int_0^1 ds(1-s)P_1^{(1,0)}(2s-1)f_0(s) = 0, \quad (47)$$

that is,

$$\int_0^1 ds(1-s)(3s-1)f_0(s) = 0. \quad (48)$$

Now suppose that the s wave is given in the region $0 \leq s \leq 1$ and satisfies (48). By the completeness and orthogonality of $P_\sigma^{(1,0)}(2s-1)$, this determines

$$\sum_{\{p=\frac{1}{3}\sigma\}}^{[\frac{1}{2}\sigma]} c_p^\sigma (b_p^\sigma)_0 \quad (49)$$

for each σ. For $\sigma = 0, 2, 3, 4, 5, 7$, there is only one value of p, so that c_p^σ is completely determined for these values of σ. Thus, in expanding the d wave in terms of $(1-s)^2 P_{\sigma-2}^{(5,0)}(2s-1)$, the coefficients corresponding to the terms with $\sigma = 2, 3, 4, 5, 7$ are completely determined. Thus, crossing symmetry imposes five constraints on the d wave, if the s wave is given. [This arises because the coefficients c_p^σ in (46) are independent of l.] The simplest such constraint (for $\sigma = 2$) is

$$c_1^2 = \frac{6}{(b_1^2)_0}\int_0^1 (1-s)P_2^{(1,0)}(2s-1)f_0(s)\,ds \quad (50)$$

$$= \frac{6}{(b_1^2)_2}\int_0^1 (1-s)^3 P_0^{(5,0)}(2s-1)f_2(s)\,ds, \quad (51)$$

that is,

$$\int_0^1 (1-s)(10s^2 - 8s + 1)f_0(s)\,ds = \tfrac{5}{2}\int_0^1 (1-s)^3 f_2(s)\,ds. \quad (52)$$

As one goes to higher waves, the number of constraints increases, and it is easy to see that, given all the waves with $l < L$, there are $2L + 1$ such relations involving the Lth partial wave and the lower ones.

Notice, however, that in this formalism the s wave is arbitrary except for (48); that is, any s wave satisfying (48) can be incorporated into a crossing symmetric form. However, crossing symmetry is not the only constraint to impose on the amplitude. In fact, using analyticity and the positivity of the absorptive part of the amplitude, Martin[3] has found strong constraints on the partial wave amplitudes in the region $0 \leq s \leq 1$. We hope to apply his techniques to the manifestly crossing symmetric amplitude (45) in order to derive constraints on the coefficients c_p^σ.

ACKNOWLEDGMENT

I wish to thank R. L. Goble for discussions and for suggesting improvements in the manuscript.

[5] Dr. S. Nussinov has kindly pointed out the following simple derivation of (48). It is trivially true that

$$\iint F(s,t,u)(s+t+u-1)\,ds\,dt = 0,$$

where the integration is over the Mandelstam triangle. Because F is totally symmetric, we can replace

$$s + t + u - 1 \equiv (s - \tfrac{1}{3}) + (t - \tfrac{1}{3}) + (u - \tfrac{1}{3})$$

by $3(s - \frac{1}{3})$ in the integral. The t integral then yields the s wave multiplied by $(1 - s)$, and (48) follows.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Random Walks on Lattices with Traps

HERBERT B. ROSENSTOCK
U.S. Naval Research Laboratory, Washington, D.C. 20390

(Received 25 April 1969)

We consider random walks on simple cubic lattices containing two kinds of sites: ordinary ones and "traps" which, when stepped on, absorb the walker. We study two related problems: (a) the probability of returning to the origin and (b) the situation in which the particle can meet its end, not only by absorption at a trap, but also by a process, called spontaneous emission, which has a constant probability per step. In problem (b), we ask for the probability that emission, rather than absorption, occurs. The solution to (a) is known for 1 dimension, and given here for the 3-, 4-, $\cdots$ dimensional cases; the 2-dimensional case remains unsolved. The solution to (b) is known for the 1-, 3-, 4-, $\cdots$ dimensional cases; we give it for 2-dimensional case.

I. INTRODUCTION

A random walk (RW) on a space lattice provides a model for many situations and processes encountered in statistical mechanics, solid state theory (e.g., diffusion of electrons, excitons, energy transfer, conductivity, dislocations), as well as in probability theory and related fields in pure mathematics. Here we consider a particle performing a symmetrical RW on a lattice (of dimension to be specified later). There are two kinds of lattice sites, ordinary ones and "traps" or "absorbers"; whenever the particle steps on a "trap," it gets absorbed, and the walk ends. We consider two related problems: (a) the probability of returning to the origin (Sec. II) and (b) the situation in which the particle can meet its death, not only by absorption by a trap, but also by spontaneous emission, a process defined to have a constant probability per step (Sec. III); in that case, we ask for the probability that the walk end by emission, rather than by absorption.

The solution to problem (a) is known[1] in 1 dimension; we give it here for 3 dimensions (Sec. II). The solution to problem (b) is known for 1 dimension[2,3] and for 3 and more dimensions[4]; we fill in the gap, the 2-dimensional case (Sec. III). It is interesting that the 2-dimensional case presents the greatest difficulties for both problems. The possibility of extending the treatment to different situations is discussed also. Results are summarized in Sec. IV.

II. PROBABILITY OF RETURNING TO THE ORIGIN BEFORE STEPPING ON A TRAP

Consider a particle performing an RW in a D-dimensional simple cubic lattice. Let traps be located, at random, on a fraction q of the lattice sites. We ask for the probability that the particle return to its origin (i.e., before being trapped). This problem was proposed by Montroll and Weiss[1] and solved in 1 dimension. We solve it here for $D \geq 3$, valid for small q, using a line of reasoning essentially due to Rudemo.[4]

Let the quantity of interest be $r^{(q)}$, the probability of returning to the origin, given a density q of traps; and let $r_t^{(q)}$ equal the probability of *first* return (fr) to the origin at step t, given a density of traps q. Then,

$$r^{(q)} = \sum_t r_t^{(q)}. \tag{1}$$

If we define $\xi_q(t, w_i)$ as the probability that for a given walk w_i of length t $[1 \leq i \leq (2D)^t]$ trapping not take place (at any trap) before step t, given a density of traps q, then we can write

$$r_t^{(q)} = (2D)^{-t} \sum_{i=1}^{(2D)^t} \delta_{\mathrm{fr},t}(w_i)\xi_q(t, w_i), \tag{2}$$

where $\delta_{\mathrm{fr},t}(w_i)$, somewhat analogous to a Kronecker delta, is unity for the walks that first return to the origin at t, and zero for all others.

Now let $V(t, w_i)$ equal the number of different points visited in the t steps of walk w_i, in the absence of traps. Then, the probability that none of the points visited in t steps be a trap is

$$\xi_q(t, w_i) = (1 - q)^{V(t,w_i)}. \tag{3}$$

Substitution of (3) into (2) and that in turn into (1) then gives

$$r^{(q)} = \sum_t (2D)^{-t} \sum_i \delta_{\mathrm{fr},t}(w_i)(1 - q)^{V(t,w_i)}. \tag{4}$$

Now for any "transient" RW we have asymptotically[5]

$$V(t) \to (1 - F)t, \tag{5}$$

[1] E. W. Montroll and G. H. Weiss, J. Math. Phys. **6**, 167 (1965), Sec. VI.
[2] H. B. Rosenstock, J. Soc. Ind. Appl. Math. **9**, 169 (1961).
[3] N. Levinson, J. Soc. Ind. Appl. Math. **10**, 442 (1962).
[4] M. Rudemo, SIAM J. Appl. Math. (formerly J. Soc. Ind. Appl. Math.) **14**, 1293 (1966).
[5] L. Spitzer, *Principles of Random Walk* (Van Nostrand, Inc., New York, 1964), pp. 35–38.

where F is defined as the probability of eventual return to the origin[6] (in absence of traps). F is a number whose value for the RW under consideration is easily obtained numerically, as we shall see later in connection with Eq. (12); for $D = 3$, F is known[7] to be 0.340537. Equation (5) is not accurate for small values of t, i.e., for short walks; short walks will, however, make a negligible contribution to (4) when q is small—the physically interesting situation. For small q we substitute (5) into (4), take the $(1 - q)$ term outside the i sum, and note that

$$(2D)^{-t} \sum_{i}^{(2D)^t} \delta_{\mathrm{tr},t}(w_i)$$

is just $r_t^{(0)}$, the probability of first return to the origin at step t in the absence of traps:

$$r^{(q)} = \sum_t r_t^{(0)}(1 - q)^{(1-F)t} \tag{6}$$

or

$$r^{(q)} = \sum_t r_t^{(0)} p^t, \tag{7}$$

with

$$p = (1 - q)^{1-F}. \tag{8}$$

We recognize the right-hand side of (7) as the generating function $F(p)$ for first returns to the origin,

$$r^{(q)} = F(p). \tag{9}$$

[Observe in passing that the quantity F used in Eq. (5) is, in fact, $F(1)$.] In general, this generating function for first return is related to the generating function G for return (not necessarily first) to the origin by[8]

$$F(p) = 1 - G^{-1}(p); \tag{10}$$

hence, we finally obtain

$$r^{(q)} = 1 - G^{-1}(p). \tag{11}$$

Now G is a known function[9] expressible as

$$G(p) = (D/p)\int_0^\infty e^{-Dz/p} I_0^D(z)\, dz, \tag{12}$$

where D is the dimension of the lattice and I_0 the zero-order Bessel function of pure imaginary argument. For any given D, the evaluation of (12) can be done numerically, using Simpson's rule and the convergent series for I_0 for z smaller than some z_1, and the asymptotic one for larger z.[10] We have carried this out for $D = 3$; the dividing point was chosen as

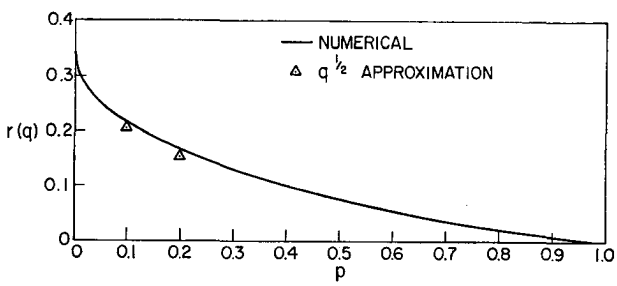


FIG. 1. Probability of return to the origin in 3 dimensions, as a function of trap density q.

$z_1 = 9$. The results are shown in the accompanying Fig. 1. A rough check on the precision of the calculation is provided by the number $F = F(1)$, which appears both in the input [in Eq. (12) via (8)] and the output ($r^{(q)}$ must equal F for physical reasons when $q = 0$). The input value was 0.340537, the output 0.340344, suggesting 3- to 4-figure accuracy. We can also find the behavior of $r^{(q)}$ for very small q from the analytic behavior of I_0; we find a negatively infinite slope, in agreement with our figure, viz.,

$$r^{(q)} \cong F - D^{\frac{3}{2}}(1 - F)^{\frac{3}{2}} 2^{-\frac{1}{2}} \pi^{-1} q^{\frac{1}{2}} + \cdots. \tag{13}$$

In the 3-dimensional case this becomes

$$r^{(q)} = 0.3405 - 0.4130 q^{\frac{1}{2}},$$

which is in good agreement (see Fig. 1) with the numerical results.

The 2-dimensional case is appreciably more difficult. To be sure, the generating function is known—merely substitute $D = 2$ into (12). But (5) does not hold for recurrent[6] RW's; instead we have[11]

$$V(t) \cong \pi t / \ln t. \tag{14}$$

If this, instead of (5), is put into (4), the result is

$$r^{(q)} = \sum_t r_t^{(0)}(1 - q)^{\pi t/\ln t} \tag{15}$$

and not simply the known generating function for the coefficients $r_t^{(0)}$. We would, therefore, have to go through the procedure of explicitly computing the expansion coefficients from the generating function before being able to evaluate (15). A straightforward computation of $r_1, r_2, r_3, \cdots$ is, of course, possible; but we have been unable to obtain an expansion for r_k asymptotically valid for large k and, therefore, we have failed to evaluate (15).

III. SPONTANEOUS EMISSION

Let $\alpha = \text{const}$ be the probability per step of spontaneous disappearance ("emission") of the walker

[6] An RW is called "transient" if $F < 1$, "recurrent" if $F = 1$. Symmetric RW's on simple cubic lattices are recurrent in 1 dimension and 2 dimensions, but transient in higher dimensions.

[7] E. W. Montroll, J. Soc. Ind. Appl. Math. 4, 241 (1965), Eq. (4.5).

[8] Ref. 7, Eq. (3.4).

[9] Ref. 7, Eq. (2.11b).

[10] H. B. Dwight, *Tables of Integrals* (Macmillan, New York, 1947), formulas 813.1 and 814.1.

[11] Ref. 1, Eq. (III.15.b).

from the lattice. (We may visualize this walker as a radioactive or otherwise unstable particle capable of decay with constant probability in time.) We ask for the probability that the walk end by spontaneous emission rather than by stepping on a trap. (Physically, this is the probability that emission be observed.) To obtain a formal solution in 2 dimensions we follow the reasoning of Rudemo[4] rather than that of earlier workers[2,3] which seems useful for 1 dimension only.

Let $P^{(q)}$ be the probability of emission if the density of traps is q, and let $P_t^{(q)}$ be the probability that this happen at step t. Then

$$P^{(q)} = \sum_t P_t^{(q)}. \tag{16}$$

Defining ξ as in Eq. (2) and $\delta_{\text{emis},t}$ as unity if emission takes place at step t and zero otherwise, we obtain, as for (2),

$$P_t^{(q)} = (2D)^{-t} \sum_{i=1}^{(2D)^t} \delta_{\text{emis},t}(w_i)\xi_q(t, w_i). \tag{17}$$

If q is small, ξ can be taken out of the i sum, since V and, hence, ξ will asymptotically become independent of w_i; $(2D)^{-1}\sum_i \delta_{\text{emis},t}$ then becomes just the probability of emission at step t in absence of traps, viz.,

$$P_t^{(0)} = (1 - \alpha)^t\alpha \tag{18}$$

(failure to emit at steps 0, 1, 2, $\cdots$, $t - 1$, followed by success at step t). Substitution of (17) into (16) then gives

$$P^{(q)} = \alpha \sum_t (1 - \alpha)^t(1 - q)^{V(t)}. \tag{19}$$

In evaluating this, we confine ourselves to the limit in which both the trap density q and the emission probability per step α are small, though their ratio remains unrestricted. In that situation, the factor α assures us that the early terms in (19) will not contribute appreciably; we can, therefore, use the asymptotic value (14) for the mean number of distinct points visited in a 2-dimensional RW throughout the range. Furthermore, since adjacent terms will not vary greatly in magnitude, we can replace the sum by an integral:

$$P^{(q)} = \alpha \int_1^\infty dt(1 - \alpha)^t(1 - q)^{\pi t/\ln t}.$$

Neglecting terms of order α^2 and q^2, we obtain finally

$$P^{(q)} = \alpha \int_1^\infty dt \exp\left[-\alpha t - (\pi qt/\ln t)\right]. \tag{20}$$

This should be evaluated for arbitrary values of the ratio α/q. We were unable to do this in closed form, but obtained separate expressions valid for small and for large α/q.

When $\alpha/q \gg 1$, the first term in the exponent is the dominating one throughout the range of integration. We, therefore, let $\alpha t = y$ and rewrite (20) as

$$P^{(q)} = \int_\alpha^\infty dy \exp\left[-y\left(1 + \frac{\pi q/\alpha}{\ln y + \ln(1/\alpha)}\right)\right]. \tag{21}$$

Here $\ln(1/\alpha)$ is a large number, much larger than y in all regions except those in which y itself is so large as to cause the leading term $\exp^{-y}$ to make the integrand negligible. The $\ln y$ term can, therefore, be neglected, and we obtain

$$\begin{aligned} P^{(q)} &= \{1 + [\pi q/\alpha \ln(1/\alpha)]\}^{-1} \\ &\quad \times \exp(-\alpha\{1 + [\pi q/\alpha \ln(1/\alpha)]\}) \\ &\cong (1 - \alpha)\{1 - [\pi q/\alpha \ln(1/\alpha)]\}. \end{aligned} \tag{22}$$

In the other extreme $\alpha/q \ll 1$, the second term in the exponential in (20) will be the larger when t is small [i.e., for values of $t < \exp(q/\alpha)$], but the first one when t is larger. However, if α/q is small enough, the crossover point will occur for t so large that the entire integrand is negligible beyond. We, therefore, set $\pi qt = z$ and write (20) as

$$P = (\alpha/\beta)\int_\beta^\infty dz e^{-g(z)}, \tag{23}$$

where

$$g(z) = (\alpha/\beta)z + z[\ln z + \ln(1/\beta)]^{-1} \tag{24}$$

with $\beta = \pi q$. Even though α/β is small, β is itself small. Hence, $\ln(1/\beta)$ is a large number, and $\ln z$ will be comparably large only for values of z so large that the entire integrand is negligible. Therefore, we expand g as follows:

$$g = \frac{z}{\ln(1/\beta)}\left(1 - \frac{\ln z}{\ln(1/\beta)}\right) + \frac{\alpha}{\beta}z$$

or, introducing $c = \ln(1/\beta)$, $w = z/c$,

$$g = w(1 + \alpha c/\beta) - (w/c)\ln cw.$$

Inserting this in (23) and expanding $\exp[-(w/c)\ln cw]$ in a Taylor series then gives

$$P^{(q)} = \frac{\alpha c}{\beta}\left[\int_{\beta/c}^\infty dw e^{-w(1+\alpha c/\beta)} + \int_{\beta/c}^\infty w\, dw e^{-w}(\ln c + \ln w)\right].$$

Now replace the lower limit by zero, thereby increasing second-order errors only. The first two integrals are then elementary, and the last one[12] is $1 - \gamma = 0.423$,

[12] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products* (Academic Press, New York, 1965), formulas 4.352.2 and 9.73.

TABLE I. Summary of solutions of problems (a) and (b).

Dimensionality D	Problem (a)—return to origin	Reference	Problem (b)—spontaneous emission	Reference
1	$q(1-q)^{-1}\ln(1/q)$	1	$1-\tau_1^{\frac{1}{2}}\int_0^\infty \exp(-\tau_1 u)\tanh u\,du$	3
			$\cong 1-\tau_1+O(\tau_1^2)$, small τ_1,	2
			$\cong 2\tau_1^{-2}+O(\tau_1^{-4})$, large τ_1, with $\tau_1=2^{\frac{1}{2}}q/\alpha^{\frac{1}{2}}$	2
2	not known		$\alpha\int_1^\infty \exp[-\alpha t-(\pi q t/\ln t)]\,dt$	0
			$=1-\alpha-[\pi\tau_2/\ln(1/\alpha)]+O(\tau_2^2)$, small τ_2,	0
			$=(\pi\tau_2)^{-1}[\ln(1/\pi q)+\ln\ln(1/\pi q)+0.423+O(1/\ln\pi q)]+O(\tau_2^{-2})$, large τ_2, with $\tau_2=q/\alpha$	0
≥ 3	$(D/P)\int_0^\infty \exp(-Dz/p)I_0^p(z)\,dz$	0	$[1+(1-F)\tau_2]^{-1}$	4
	$\cong F-D^{\frac{1}{2}}(1-F)^{\frac{1}{2}}2^{-\frac{1}{2}}\pi^{-1}q^{\frac{1}{2}}+O(q^{\frac{3}{2}})$, with $p=(1-q)^{1-F}$	0		

γ being Euler's constant. So we finally obtain

$$P^{(q)}=(\alpha/\beta)[c+\ln c+0.423+O(c^{-1})]+O((\alpha/\beta)^2). \tag{25}$$

Since the derivation involves integration over a region in which previous expansions are not formally valid, this is an asymptotic, rather than a convergent, series. Summarizing (22) and (25), we have

$$P^{(q)}=1-\alpha-[\pi q/\alpha\ln(1/\alpha)], \qquad \frac{\pi q}{\alpha}\ \text{small},$$

$$=\frac{\alpha}{\pi q}\left[\ln\frac{1}{\pi q}+\ln\ln\frac{1}{\pi q}+0.423+O\left(1\Big/\ln\left(\frac{1}{\pi q}\right)\right)\right]+O\left(\left(\frac{\alpha}{\pi q}\right)^2\right), \qquad \frac{\pi q}{\alpha}\ \text{large}. \tag{26}$$

Direct numerical evaluation of (20) gives good agreement with (26). For example, direct evaluation of (20) gives 0.00924 and 0.00121, respectively, for $(\alpha, \pi q)=(10^{-6}, 10^{-3})$ and $(10^{-8}, 10^{-4})$, whereas the second line of (26) gives 0.00926 and 0.00119 for these situations. Similarly, direct evaluation of (20) gives 0.99883 and 0.99758, respectively, for $(\alpha, \pi q)=(10^{-3}, 10^{-6})$ and $(10^{-3}, 10^{-5})$, whereas the first line of (26) gives 0.99885 and 0.99755. The general behavior of P as a function of α and πq is illustrated in Fig. 2.

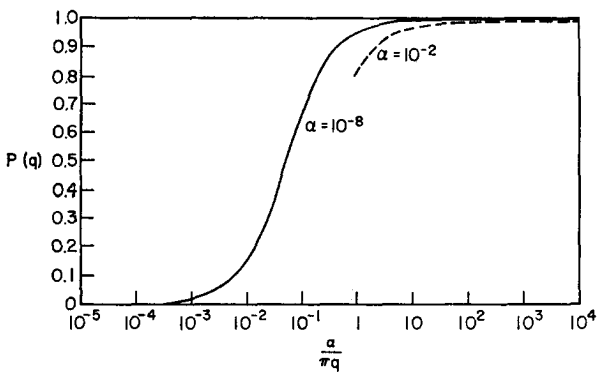


FIG. 2. Probability of spontaneous emission in 2 dimensions, for various trap densities q and spontaneous emission probabilities α per step.

REVIEW

The known results for both problems discussed in this paper are summarized in Table I, where q and α, as previously defined, are, respectively, the trap density and the emission probability per step; F, the return probability in absence of both traps and emission, is a number whose value depends on the dimensionality. The "reference" column refers to our footnotes; Ref. 0 means the present paper. The functional dependence of the calculated probabilities is seen to be quite strongly determined by the dimensionality.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Polariton–Phonon Interaction in Molecular Crystals*

C. MAVROYANNIS
Division of Pure Chemistry, National Research Council, Ottawa, Canada

(Received 2 June 1969)

A general theory is presented concerning the interaction between the polaritons and the acoustic phonons in molecular crystals. Using a one-phonon approximation to truncate the hierarchy of the Green's functions involved and disregarding mixing of different polariton bands, we derive an expression for the dielectric permeability of the crystal. The absorption coefficient for the coupled polariton–phonon spectrum is found to have an asymmetric Lorentzian lineshape even if the frequency dependence of the energy shift and spectral width is neglected. The damping function causing the asymmetry of the spectral line depends entirely on the coupling between the dressed electron subsystem and the phonon field. Expressions are developed for the energy shift and spectral width of the resonance line, and their temperature dependence is discussed. Far from the resonance peak, the frequency and temperature dependence of the absorption coefficient is established. In the transparent region of frequencies of the crystal, the expression for the absorption coefficient consists of two terms that have delta-function distributions and are peaked at different frequency regions, depending on whether or not the polariton and the phonon fields are coupled. In the limiting case where retardation can be ignored, the bare exciton–phonon interaction is discussed. The average energy of the crystal resulting from the polariton–phonon interaction at finite temperatures has been derived in a closed form.

I. INTRODUCTION

As is well known, a polariton is an elementary excitation (quasiparticle) which is a mixture of the low-lying excited states of a crystal (excitons or phonons) with the electromagnetic field (photons). In a tight-binding model of a molecular crystal, a polariton may be viewed as an electron–hole pair tightly bound to one another, dressed by the radiation field, and migrating through the crystal. This concept was first introduced by Hopfield[1] in his study of the optical properties of insulators. Since then a large number of investigations have been devoted to the properties of polaritons in molecular crystals, particularly by Russian workers. In a recent excellent review article, Ovander[2] has begun with the work of Agranovich[3] and discussed the progress that has been made in this field, with emphasis on the nonlinear optical effects in crystals; we refer to his paper[2] where references are given.

Agranovich and Konobeev[4] have found that the interaction between polaritons and acoustic phonons becomes substantial at low temperatures and determines the shape of the long-wavelength edge of the exciton absorption bands. Ovander[2,5] developed a theory for Rayleigh scattering in molecular crystals by considering the polariton–acoustic photon interaction at zero temperature. For frequencies of the incident light far from the absorption band, he derived an expression for the photon cross section, which is a generalization of the known formula for the Rayleigh scattering in gases; it contains an additional term which is due to the interaction between the electronic excitation and the acoustic phonons. Ovander suggested that this term will dominate for frequencies in the exciton absorption region. We shall indicate in due course what parts of the polariton–phonon interaction were taken into account in both mentioned studies.

In the present study we attempt to develop a general theory of the interaction between polaritons and acoustic phonons in a molecular crystal. We begin in Sec. II with an expression for the dielectric permeability derived in earlier work[6] that corresponds to the interband excitations occurring in a molecular crystal. Here we neglect terms in the expression for the dielectric permeability describing the physical process of Raman scattering because we will deal with this problem in a later publication, but the coupling of the excitation field to the acoustic branches of the phonon field is fully included. Then, by means of a canonical transformation, the exciton operators are transformed into that of polaritons and the required expression for the dielectric permeability is derived corresponding to the normal polariton waves in the crystal. In Sec. III the model Hamiltonian for the crystal is developed, which consists of the free-polariton field, the phonon field in the harmonic approximation, and the polariton–phonon interaction.

* Issued as N.R.C. No. 10816.

[1] J. J. Hopfield, Phys. Rev. **112**, 1555 (1958).

[2] L. N. Ovander, Usp. Fiz. Nauk **86**, 3 (1965) [Sov. Phys.—Usp. **8**, 337 (1965)].

[3] V. M. Agranovich, Zh. Eksp. Teor. Fiz. **37**, 430 (1959) [Sov. Phys.—JETP **10**, 307 (1960)].

[4] V. M. Agranovich and Yu. V. Konobeev, Fiz. Tverd. Tela **3**, 360 (1961) [Sov. Phys.—Solid State **3**, 260 (1961)].

[5] L. N. Ovander, Fiz. Tverd. Tela **5**, 21 (1963) [Sov. Phys.—Solid State **5**, 13 (1963)].

[6] C. Mavroyannis, J. Math. Phys. **8**, 1522 (1967).

The equations of motion in a closed form are derived in Sec. IV for the Green's functions for the polariton field and the polariton–phonon field by applying a simple decoupling approximation to the higher-order Green's functions involved. The decoupling approximation amounts to restricting ourselves to physical processes involving the emission or absorption of one phonon. Then, disregarding mixing of different polariton bands, we obtain expressions for the Green's functions of interest and consequently the expression for the dielectric permeability of the coupled polariton–phonon system. It consists mainly of two terms: one of them describes the coupled polariton–phonon spectrum, while the other is proportional to the unperturbed polariton–phonon Green's function. In the limiting case where retardation may be neglected, the corresponding expression is derived for the bare exciton–phonon spectrum. The Green's function for the phonon field is calculated in the same approximation as that used for the derivation of the polariton Green's function and the polarization operator corresponds to the physical process where two polaritons are created or annihilated simultaneously through the exchange of a phonon. The phonon occupation number is evaluated in the limit when the imaginary part of the polarization operator tends to zero.

The expression for the frequency and wave-vector-dependent absorption coefficient of the polariton–phonon system is calculated in Sec. V. It consists of two terms: the first term is proportional to a shape function, while the second has a delta-function distribution. The shape function is an asymmetric Lorentzian line, even if the frequency dependences of the energy shift and the damping functions are neglected. The asymmetry arises from the coupling between the electronic excitations as well as the intermolecular interactions with the phonon field. Expressions have been obtained for the energy shift and spectral width as a function of temperature for the case of resonance, provided that certain conditions are satisfied. It is shown that, for frequencies near resonance, the shape function consists of a superposition of symmetric and asymmetric Lorentzian lines, and, at resonance, the temperature dependence of the absorption coefficient is established. To our knowledge, the polariton resonance spectrum has not been previously discussed.

In the limiting case where the damping goes to zero, both terms in the expression for the absorption coefficient are delta shaped, but they satisfy different dispersion relations and, therefore, they are peaked at different frequency regions. The first term corresponds to the energy spectrum of one-particle excitation where the polariton and the phonon field are coupled together and the energy of excitation is just the unperturbed polariton energy renormalized to account for the interactions involved. This is the physical process for polariton scattering by acoustic lattice vibrations. The particular case where our results correspond to those obtained by Ovander[2,5] are pointed out. The second term is peaked in the frequency region where the polariton and phonon fields are not coupled; this is the frequency region of two-particle excitations. In the limit where retardation is not important, the absorption coefficient for the bare exciton–phonon is discussed.

Finally, in Sec. VI the average energy of the system is calculated by averaging the total Hamiltonian. The average energy of interaction due to the polariton–phonon interaction is obtained in a closed form. It is expressed as the difference of the excitation energy corresponding to the coupled polariton–phonon field which is temperature dependent and the unperturbed polariton energy.

II. DIELECTRIC PERMEABILITY

The dielectric permeability tensor is related to the complex electrical conductivity tensor $\sigma_{ij}(\mathbf{k}, \omega)$ by the well-known relation

$$\epsilon_{ij}(\mathbf{k}, \omega) = \delta_{ij} + (4\pi i/\omega)\sigma_{ij}(\mathbf{k}, \omega).$$

In Ref. 6, use has been made of the relation between the electrical conductivity tensor and the Fourier transform of the current–current retarded Green's function of the system, e.g., Eq. (98) of Ref. 6; then, discarding Umklapp processes, an expression for the dielectric permeability, corresponding to the interband excitations occurring in an isotropic molecular crystal, has been derived in the form[6]

$$\begin{aligned}\epsilon_{ii}(\mathbf{k}, \omega) = \epsilon_\infty &- \frac{\pi}{\omega^2}\omega_p^2(n_0 - n_\mu)\sum_\mu f_{0\mu}(\mathbf{k}, i)E_\mu(\mathbf{k})g_\mu(\mathbf{k}, \omega) \\ &- \frac{\pi\omega_p^2}{\omega^2 N^{\frac{1}{2}}}\sum_{\mathbf{q},r,\mu,\mu'}[(n_0 - n_\mu)f_{0\mu}(\mathbf{k}, i)E_\mu(\mathbf{k})]^{\frac{1}{2}} \\ &\times [(n_0 - n_{\mu'})f_{0\mu'}(\mathbf{k} - \mathbf{q}, i)E_{\mu'}(\mathbf{k} - \mathbf{q})]^{\frac{1}{2}} \\ &\times [Y_r^*(\mathbf{q})\langle\langle\tilde{b}_\mu(\mathbf{k}); \tilde{b}_{\mu'}^+(\mathbf{k} - \mathbf{q})\beta_r^+(\mathbf{q})\rangle\rangle \\ &+ Y_r(\mathbf{q})\langle\langle\tilde{b}_{\mu'}(\mathbf{k} - \mathbf{q})\beta_r(\mathbf{q}); \tilde{b}_\mu^+(\mathbf{k})\rangle\rangle] \\ &- \frac{\pi\omega_p^2}{\omega^2 N}(n_0 - n_{\mu'})\sum_{\mu',\mathbf{q},r} f_{0\mu'}(\mathbf{k} - \mathbf{q}, i)E_{\mu'}(\mathbf{k} - \mathbf{q})\,|Y_r(\mathbf{q})|^2 \\ &\times \langle\langle\tilde{b}_{\mu'}(\mathbf{k} - \mathbf{q})\beta_r(\mathbf{q}); \tilde{b}_{\mu'}^+(\mathbf{k} - \mathbf{q})\beta_r^+(\mathbf{q})\rangle\rangle, \quad (1)\end{aligned}$$

where

$$\epsilon_\infty = 1 - (n_0 - n_\mu)\omega_p^2/\omega^2, \quad \omega_p^2 = 4\pi e^2 N\sigma S/mV, \quad (2)$$

ω_p is the frequency of plasma oscillations, and $n_0 \equiv \langle \alpha^+_{\mathbf{k}0}\alpha_{\mathbf{k}0}\rangle$ and $n_\mu \equiv \langle \alpha^+_{\mathbf{k}\mu}\alpha_{\mathbf{k}\mu}\rangle$ are the occupation numbers of the holes and the electrons in the valence and μth excitation bands, respectively. The system of units with $\hbar = 1$ is used throughout. The oscillator strength $f_{0\mu}(\mathbf{k}, i)$ is defined as

$$f_{0\mu}(\mathbf{k}, i) = \left(\frac{2m}{e^2\sigma S}\right) \sum_\alpha |\mathbf{e}_{\mathbf{k}i} \cdot \mathbf{P}^\alpha_{0\mu}|^2 E_\mu(\mathbf{k}) \times |u_{\mu\alpha}(\mathbf{k}) + v_{\mu\alpha}(\mathbf{k})|^2 \exp i\mathbf{k} \cdot (\mathbf{r}^{(0)}_{n\alpha} - \mathbf{r}^{(0)}_{m\alpha}), \quad (3)$$

where m is the electron mass, e^2 denotes the square of the electronic charge divided by the static dielectric constant of the substance, $\alpha\ (= 1, 2, \cdots, \sigma)$ enumerates the molecules in the unit cell, and S and N are the total number of electrons in the unit cell and the total number of unit cells in the crystal, respectively. The compound index μ denotes the exciton band, the corresponding molecular term, and the kind of mode, transverse ($i = \mu_\perp = 1, 2$) or longitudinal ($i = \mu_\parallel = 3$). The quantity $E_\mu(\mathbf{k})$ is the energy of excitation of the μth exciton band, and $u_{\mu\alpha}(\mathbf{k})$ and $v_{\mu\alpha}(\mathbf{k})$ are the amplitudes that diagonalize the unperturbed part of the Hamiltonian arising from the direct electron–hole pair interactions in the zero approximation; explicit expressions for $E_\mu(\mathbf{k})$ and $u_{\mu\alpha}(\mathbf{k})$ and $v_{\mu\alpha}(\mathbf{k})$ are given elsewhere.[3,7] The dipole-moment operator $\mathbf{P}^\alpha_{0\mu}$ corresponds to the allowed transition $0 \to \mu$, while $\mathbf{e}_{\mathbf{k}i}$ $(= \mathbf{e}_{-\mathbf{k}i})$ and V designate the photon-polarization vector and the total volume of the crystal, respectively. The position vector of the molecule α at equilibrium on the lattice site $\mathbf{n}$ is denoted by $\mathbf{r}^{(0)}_{n\alpha}$, and $\mathbf{k}$ is the wave vector of the external electromagnetic field. In the expression (1), $g_\mu(\mathbf{k}, \omega)$ is the Fourier transform of the retarded Green's function given by

$$g_\mu(\mathbf{k}; \omega) \equiv \langle\langle \tilde{b}_\mu(\mathbf{k}); \tilde{b}^+_\mu(k)\rangle\rangle, \quad \tilde{b}_\mu(\mathbf{k}) \equiv b_\mu(-\mathbf{k}) - b^+_\mu(\mathbf{k}), \quad (4)$$

and $b^+_\mu(\mathbf{k})$ and $b_\mu(\mathbf{k})$ are the creation and annihilation operators corresponding to the bare exciton spectrum and obey Pauli statistics. The retarded Green's function $g_\mu(\mathbf{k}, t - t')$ is defined as[8]

$$g_\mu(\mathbf{k}, t - t') = -i\theta(t - t')\langle[\tilde{b}_\mu(\mathbf{k}, t), \tilde{b}^+_\mu(\mathbf{k}, t')]_-\rangle,$$

where the angular brackets denote the average over the canonical ensemble appropriate to the total Hamiltonian $\mathcal{H}$ of the system, $\theta(t)$ is the usual step function, and the operators $\tilde{b}_\mu(\mathbf{k}, t)$ and $\tilde{b}^+_\mu(\mathbf{k}, t')$ are in the Heisenberg representation. In Eqs. (1) and (4) and what follows, the time arguments of the operators have been suppressed for convenience. For further details regarding expressions (1) and (4) we refer to our previous work.[6]

We remark that in Ref. 6 the Green's functions, which appear in the expression for the dielectric permeability [Eq. (99) of Ref. 6] corresponding to interband transitions for an isotropic molecular crystal (static lattice), have been calculated by means of the total Hamiltonian consisting of the exciton Hamiltonian plus the interaction between the excitons and the electromagnetic field. Though Umklapp processes have been neglected, the derived expression for the dielectric permeability, Eqs. (103)–(108) of Ref. 6, includes "local field corrections" explicitly and is a generalized version of the well-known Lorenz–Lorentz formula. For a discussion on "local field" effects in the tight-binding model we refer to Ref. 1. In comparison with the expression $\epsilon_{ii}(\mathbf{k}, \omega)$ derived in Ref. 6, we have here disregarded in (1) quadratic terms with respect to the exciton operators corresponding to the physical process of Raman scattering, but the coupling between the bare excitons and the acoustical branch of the phonon field has been included explicitly. This is done in the usual way by expanding in power series the displacements of the molecules from their equilibrium positions and retaining only the linear terms in the expansion; then use has been made of the expansion of the molecular displacements $\Delta\mathbf{r}_{n\alpha}$ that are assumed to be small, in terms of the phonon creation and annihilation operators, i.e.,

$$\Delta\mathbf{r}_{n\alpha} = 1/N^{\frac{1}{2}} \sum_{\mathbf{q},r} \frac{Y_r(i\mathbf{q}) e^{i\mathbf{q}\cdot\mathbf{r}_{n\alpha}(0)}}{(iq)} \tilde{\beta}_r(\mathbf{q}), \quad (5)$$

with

$$Y_r(\mathbf{q}) = i(\mathbf{q} \cdot \mathbf{e}^\alpha_{\mathbf{q}r})[2M_\alpha \sigma \omega_r(\mathbf{q})]^{-\frac{1}{2}}, \quad (6)$$

where $\mathbf{e}^\alpha_{\mathbf{q}r}$ is a unit vector in the direction of polarization of the phonon, $\omega_r(\mathbf{q})$ is the frequency of the translational acoustic vibrations in the harmonic approximation, and M_α is the mass of the molecule α. The operator $\tilde{\beta}_r(\mathbf{q}) \equiv \beta_r(\mathbf{q}) + \beta^+_r(-\mathbf{q})$, where $\beta^+_r(-\mathbf{q})$ and $\beta_r(\mathbf{q})$ are the creation and annihilation operators of acoustic phonons with wave vector $\mathbf{q}$, respectively, and satisfy Bose commutation relations; the index r designates the phonon branch. The sum over $\mathbf{q}$ in (5) extends over all values, while $2\omega_r(\mathbf{q})^{-1}\tilde{\beta}_r(\mathbf{q})$ refers to the reduced vector in the first zone; this amounts to the inclusion of Umklapp processes. The Umklapp processes are not treated here explicitly, since we are mainly concerned with interband transitions, while Umklapp processes dominate only in the x-ray range of frequencies, but they may be included without major changes in the theory. The nonlinear effects due

[7] C. Mavroyannis, J. Chem. Phys. **42**, 1772 (1965).
[8] C. Mavroyannis, J. Math. Phys. **8**, 1515 (1967).

to the electronic-vibrational interaction leading to Raman scattering, which have been excluded from the expression (1), will be the subject of a later publication. Since we have made the assumption that the molecular displacements from their equilibrium positions are small, the expression (1) corresponds to weak coupling between the bare excitons and the bare acoustic phonons. The opposite case corresponding to the strong coupling between localized or trapped excitons and the lattice vibrations accompanied by a local deformation of the lattice shall not be treated in the present study. Our treatment is not applicable in this case because the displacements of the molecules from their equilibrium positions are large.

In order to study the polariton spectrum we employ the canonical transformation[3,4]

$$b_\mu(\mathbf{k}) = \sum_\rho [u_{\mu\rho}(\mathbf{k})\zeta_{\rho\lambda}(\mathbf{k}) + v^*_{\mu\rho}(-\mathbf{k})\zeta^+_{\rho\lambda}(-\mathbf{k})], \quad (7)$$

where $u_{\mu\rho}(\mathbf{k})$ and $v_{\mu\rho}(\mathbf{k})$ are the amplitudes that diagonalize the Hamiltonian for the free (static lattice) excitons interacting with the electromagnetic field, while $\zeta^+_{\rho\lambda}(\mathbf{k})$ and $\zeta_{\rho\lambda}(\mathbf{k})$ are the creation and annihilation operators for the polariton spectrum, respectively, which satisfy approximately Bose commutation relations.[3,4,9] In this approximation the Hamiltonian for the undisplaced lattice becomes diagonal, i.e.,

$$\mathcal{H}^{(0)} = \langle \mathcal{H}_0 \rangle^{(0)} + \sum_{\mathbf{k},\rho,\lambda} \omega_{\rho\lambda}(\mathbf{k})\zeta^+_{\rho\lambda}(\mathbf{k})\zeta_{\rho\lambda}(\mathbf{k}), \quad (8)$$

where $\omega_{\rho\lambda}(\mathbf{k})$ is the real energy of excitation of the polariton waves with wave vector $\mathbf{k}$ and polarization λ (= 1, 2), and is given by the ρth root of the secular equation[8]

$$c^2k^2 = \omega^2_{\rho\lambda}(\mathbf{k})\hat{\eta}^2(\mathbf{k}, \omega_{\rho\lambda}(\mathbf{k})), \quad (9)$$

with $\hat{\eta}(\mathbf{k}, \omega_{\rho\lambda}(\mathbf{k}))$ being the real part of the index of refraction of polariton waves in the crystal and which is equal to

$$\begin{aligned}\hat{\eta}^2(\mathbf{k}, \omega_{\rho\lambda}(\mathbf{k})) &= 1 + \alpha(\mathbf{k}, \omega_{\rho\lambda}(\mathbf{k})) \\ &= 1 + (n_0 - n_\mu)\omega_p^2 \sum_{\mu,\lambda} f_{0\mu}(\mathbf{k}, \lambda)/[E^2_\mu(\mathbf{k}) - \omega^2_{\rho\lambda}(\mathbf{k})].\end{aligned} \quad (10)$$

We note that in the expression (10) for the square of the index of refraction of optical waves, which satisfies the Maxwell equation (9), the "local field" effects are included.[8] The index ρ enumerates the polariton bands which correspond to the solutions of the secular equation (9). It is evident from (10) that for each value of λ there are two polariton branches, a fact that has been discussed in the literature previously.[10] In (8), $\langle \mathcal{H}^{(0)}_0 \rangle$ is the average energy resulting from the dressed electron–hole pair interactions; its explicit expression is given elsewhere.[6,8]

Substituting (5) and its complex conjugate into (1), we obtain the following expression for $\epsilon_{\lambda\lambda}(\mathbf{k}, \omega) \equiv \epsilon_\lambda(\mathbf{k}, \omega)$, corresponding to the normal polariton waves in the crystal:

$$\begin{aligned}\epsilon_\lambda(\mathbf{k}, \omega) = \epsilon_\infty &- \frac{\pi\omega_p^2}{\omega^2}(n_0 - n_\mu) \\ &\times \sum_\rho \hat{f}_{0\mu}(\mathbf{k}, \lambda)\omega_{\rho\lambda}(\mathbf{k})\langle\langle\tilde{\zeta}_{\rho\lambda}(\mathbf{k}); \tilde{\zeta}^+_{\rho\lambda}(\mathbf{k})\rangle\rangle \\ &- \frac{\pi\omega_p^2}{\omega^2 N^{\frac{1}{2}}} \sum_{\mathbf{q},r,\rho,\rho'} [(n_0 - n_\mu)\hat{f}_{0\mu}(\mathbf{k}, \lambda)\omega_{\rho\lambda}(\mathbf{k})]^{\frac{1}{2}} \\ &\times [(n_0 - n_{\mu'})\hat{f}_{0\mu'}(\mathbf{k} - \mathbf{q}, \lambda')\omega_{\rho'\lambda'}(\mathbf{k} - \mathbf{q})]^{\frac{1}{2}} \\ &\times [Y^*_r(q)\langle\langle\tilde{\zeta}_{\rho\lambda}(\mathbf{k}); \tilde{\zeta}^+_{\rho'\lambda'}(\mathbf{k} - \mathbf{q})\tilde{b}^+_r(\mathbf{q})\rangle\rangle \\ &+ Y_r(\mathbf{q})\langle\langle\tilde{\zeta}_{\rho'\lambda'}(\mathbf{k} - \mathbf{q})\tilde{\beta}_r(\mathbf{q}); \tilde{\zeta}^+_{\rho\lambda}(\mathbf{k})\rangle\rangle] \\ &- \frac{\pi\omega_p^2}{\omega^2 N}(n_0 - n_{\mu'}) \sum_{\mathbf{q},r,\rho'} \hat{f}_{0\mu'}(\mathbf{k} - \mathbf{q}, \lambda') \\ &\times \omega_{\rho'\lambda'}(\mathbf{k} - \mathbf{q})\,|Y_r(\mathbf{q})|^2 \\ &\times \langle\langle\tilde{\zeta}_{\rho'\lambda'}(\mathbf{k} - \mathbf{q})\tilde{\beta}_r(\mathbf{q}); \tilde{\zeta}^+_{\rho'\lambda'}(\mathbf{k} - \mathbf{q})\tilde{\beta}^+_r(\mathbf{q})\rangle\rangle,\end{aligned} \quad (11)$$

where $\hat{f}_{0\mu}(\mathbf{k}, \lambda)$ is defined now as

$$\begin{aligned}\hat{f}_{0\mu}(\mathbf{k}, \lambda) &= \sum_\mu f_{0\mu}(\mathbf{k}, i)\left(\frac{E_\mu(\mathbf{k})}{\omega_{\rho\lambda}(k)}\right) \times |u_{\mu\rho}(\mathbf{k}) - v_{\mu\rho}(-\mathbf{k})|^2 \\ &= \sum_\mu f_{0\mu}(\mathbf{k}, i)\left[1 - \left(\frac{d\omega^2}{d\omega^2}\hat{\eta}^2_\lambda(\mathbf{k}, \omega)\right)^{-1}_{\omega=\omega_{\rho\lambda}(\mathbf{k})}\right],\end{aligned} \quad (12)$$

and $\tilde{\zeta}_{\rho\lambda}(\mathbf{k}) \equiv \zeta_{\rho\lambda}(\mathbf{k}) + \zeta^+_{\rho\lambda}(-\mathbf{k})$. In deriving the second equation on the r.h.s. of (12), we have made use of the explicit expressions for the amplitudes $u_{\mu\rho}(\mathbf{k})$ and $v_{\mu\rho}(-\mathbf{k})$ derived in Ref. 8. The expression for $\epsilon_\lambda(\mathbf{k}, \omega)$ given by (11) can be easily reduced to the same form as that used by Agranovich and Konobeev[9] if we retain only the first two terms of the r.h.s. of (11) and then replace $\epsilon_\lambda(\mathbf{k}, \omega)$ by its inverse $\epsilon^{-1}_\lambda(\mathbf{k}, \omega)$. In the following sections we calculate the Green's functions that appear on the r.h.s. of (11) by means of a model Hamiltonian corresponding to a molecular crystal.

III. THE POLARITON-PHONON HAMILTONIAN

The Hamiltonian for a molecular crystal is taken in the form

$$\mathcal{H} = \mathcal{H}^{(0)} + \mathcal{H}_{\text{phonon}} + \mathcal{H}_{\text{int}}, \quad (13)$$

[9] V. M. Agranovich and Yu. V. Konobeev, Fiz. Tverd. Tela **5**, 2544 (1963) [Sov. Phys.—Solid State **5**, 1858 (1964)].

[10] A. A. Demidenko, Fiz. Tverd. Tela **3**, 1195 (1961) [Sov. Phys.—Solid State **3**, 869 (1961)].

where the unperturbed Hamiltonian $\mathcal{H}^{(0)}$ is given by the expression (8), while the Hamiltonian of the phonon field $\mathcal{H}_{\text{phonon}}$ in the harmonic approximation is equal to

$$\mathcal{H}_{\text{phonon}} = \sum_{\mathbf{q},r} \omega_r(\mathbf{q})(\beta_r^+(\mathbf{q})\beta_r(\mathbf{q}) + \tfrac{1}{2}). \quad (14)$$

The interaction of the polaritons with the acoustic phonons $\mathcal{H}_{\text{int}}$, linear with respect to the phonon operators, is described by the expression[5,9,11,12]

$$\begin{aligned}\mathcal{H}_{\text{int}} = &\sum_{\substack{\mathbf{k},\mathbf{q},r\\ \rho,\rho',\lambda,\lambda'}} \Phi_{\mathbf{q}r}(\mathbf{k}+\mathbf{q},\rho',\lambda';\mathbf{k},\rho,\lambda)\\ &\times \tilde{\zeta}^+_{\rho'\lambda'}(\mathbf{k}+\mathbf{q})\tilde{\zeta}_{\rho\lambda}(\mathbf{k})\tilde{\beta}_r(\mathbf{q})\\ &+ \sum_{\substack{\mathbf{k},\mathbf{q},r\\ \rho,\rho',\lambda,\lambda'}} W_{\mathbf{q}r}(\mathbf{k}+\mathbf{q},\rho',\lambda';\mathbf{k},\rho,\lambda)\\ &\times \zeta^+_{\rho'\lambda'}(\mathbf{k}+\mathbf{q})\zeta_{\rho\lambda}(\mathbf{k})\tilde{\beta}_r(\mathbf{q})\\ &+ \sum_{\substack{\mathbf{k},\mathbf{q},r\\ \rho,\rho',\lambda,\lambda'}} [\hat{W}_{\mathbf{q}r}(\mathbf{k}+\mathbf{q},\rho',\lambda';\mathbf{k},\rho,\lambda)\\ &\times \zeta^+_{\rho\lambda}(-\mathbf{k})\zeta^+_{\rho'\lambda'}(\mathbf{k}+\mathbf{q})\tilde{\beta}_r(\mathbf{q}) + \text{h.c.}], \quad (15)\end{aligned}$$

where $\Phi_{\mathbf{q}r}(\mathbf{k}+\mathbf{q},\rho'\lambda';\mathbf{k},\rho,\lambda)$ is the coupling function arising from the interaction between the electromagnetic field and the acoustical branches of the phonon field and which is equal to

$$\begin{aligned}&\Phi_{\mathbf{q}r}(\mathbf{k}+\mathbf{q},\rho',\lambda';\mathbf{k},\rho,\lambda)\\ &= \frac{i\omega_p}{2N^{\frac{1}{2}}}\sum_{\mu}[f_{0\mu}(\mathbf{k}+\mathbf{q},\lambda')E_\mu(\mathbf{k}+\mathbf{q})/c(k+q)]^{\frac{1}{2}}\\ &\times [u^*_{\mu\rho}(\mathbf{k}) - v^*_{\mu\rho}(-\mathbf{k})]\\ &\times [u_{\rho'\lambda'}(\mathbf{k}+\mathbf{q}) + v_{\rho'\lambda'}(-\mathbf{k}-\mathbf{q})]Y_r(\mathbf{q})\\ &+ (\omega_p^2/4cN^{\frac{1}{2}})[\mathbf{e}_{\mathbf{k}+\mathbf{q},\lambda'}\cdot\mathbf{e}_{\mathbf{k},\lambda}][u_{\rho\lambda}(\mathbf{k}) + v_{\rho\lambda}(-\mathbf{k})]\\ &\times [u^*_{\rho'\lambda'}(\mathbf{k}+\mathbf{q}) + v^*_{\rho'\lambda'}(-\mathbf{k}-\mathbf{q})]Y_r(\mathbf{q}). \quad (16)\end{aligned}$$

The two terms in (16) result from the expression for the electron–photon interaction Hamiltonian (in the uncoupled picture), which are linear and quadratic with respect to the vector potential, respectively.[5] In (16), $\mathbf{e}_{\mathbf{k}\lambda}$ and $\mathbf{e}_{\mathbf{k}+\mathbf{q},\lambda'}$ are the photon polarization operators with wave vectors $\mathbf{k}$ and $\mathbf{k}+\mathbf{q}$ and polarization λ and λ', respectively. The amplitudes $u_{\rho\lambda}(\mathbf{k})$ and $v_{\rho\lambda}(-\mathbf{k})$ arise from the canonical transformation from the photon operators $\alpha_\lambda(\mathbf{k})$ and $\alpha^+_\lambda(\mathbf{k})$ to the corresponding polariton operators, i.e.,

$$\alpha_\lambda(k) = \sum_\rho [u_{\rho\lambda}(\mathbf{k})\zeta_{\rho\lambda}(\mathbf{k}) + v^*_{\rho\lambda}(-\mathbf{k})\zeta^+_{\rho\lambda}(-\mathbf{k})].$$

If we make use of the expressions for the u's and v's given in Ref. 8, then the expression for $\Phi_{\mathbf{q}r}(\mathbf{k}+\mathbf{q},\rho',\lambda';\mathbf{k},\rho,\lambda)$ turns out to be

$$\begin{aligned}\Phi_{\mathbf{q}r}(\mathbf{k}+\mathbf{q},\rho',\lambda';\mathbf{k},\rho,\lambda) = &(\omega_p^2/4N^{\frac{1}{2}})[\omega_{\rho\lambda}(\mathbf{k})\omega_{\rho'\lambda'}(\mathbf{k}+\mathbf{q})]^{-\frac{1}{2}}Y_r(\mathbf{q})\\ &\times\left\{(n_0-n_\mu)\sum_\mu\left[\frac{(f_{0\mu}(\mathbf{k}+\mathbf{q},\lambda')f_{0\mu}(\mathbf{k},\lambda)E_\mu(\mathbf{k})E_\mu(\mathbf{k}+\mathbf{q})]^{\frac{1}{2}}}{\omega_{\rho\lambda}(\mathbf{k}) - E_\mu(\mathbf{k})}\right.\right.\\ &\left.\left.+\frac{[f_{0\mu}(\mathbf{k},\lambda)f_{0\mu}(\mathbf{k}+\mathbf{q},\lambda')E_\mu(\mathbf{k})E_\mu(\mathbf{k}+\mathbf{q})]^{\frac{1}{2}}}{\omega_{\rho\lambda}(\mathbf{k}) + E_\mu(\mathbf{k})}\right] + (\mathbf{e}_{\mathbf{k}+\mathbf{q},\lambda'}\cdot\mathbf{e}_{\mathbf{k},\lambda})\right\}\\ &\times\left(\frac{d\omega^2}{d\omega^2}\hat{\eta}^2(\mathbf{k},\omega)\right)^{-\frac{1}{2}}_{\omega=\omega_{\rho\lambda}(\mathbf{k})}\times\left(\frac{d\omega^2}{d\omega^2}\hat{\eta}^2(\mathbf{k}+\mathbf{q},\omega)\right)^{-\frac{1}{2}}_{\omega=\omega_{\rho'\lambda'}(\mathbf{k}+\mathbf{q})}. \quad (17)\end{aligned}$$

It is easy to see that the expression (17) for $\Phi_{\mathbf{q}r}(\mathbf{k}+\mathbf{q},\rho',\lambda';\mathbf{k},\rho,\lambda)$ is a generalization of the Mandelstam–Brillouin component[2,5] (in energy units) in the light-scattering process that is caused by the interaction with acoustic vibrations in the crystal.

The coupling function $W_{\mathbf{q}r}(\mathbf{k}+\mathbf{q},\rho',\lambda';\mathbf{k},\rho,\lambda)$ that arises from the interaction between the electronic subsystem in the excited state, the intermolecular interactions, and the acoustic vibrations, is given by

$$\begin{aligned}&W_{\mathbf{q}r}(\mathbf{k}+\mathbf{q},\rho',\lambda';\mathbf{k},\rho,\lambda)\\ &= \sum_{\mu,\mu'}[E_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k}) + F^{(1)}_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k})]\\ &\times [u^*_{\mu\rho'}(\mathbf{k}+\mathbf{q})u_{\mu'\rho}(\mathbf{k}) + v_{\mu\rho'}(-\mathbf{k}-\mathbf{q})v^*_{\mu'\rho}(\mathbf{k})]Y_r(\mathbf{q})\\ &+\sum_{\mu\mu'}F^{(2)}_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k})[v^*_{\mu\rho'}(-\mathbf{k}-\mathbf{q})u_{\mu'\rho}(\mathbf{k})\\ &+ u^*_{\mu\rho'}(\mathbf{k}+\mathbf{q})v_{\mu'\rho}(\mathbf{k})]Y_r(\mathbf{q}), \quad (18)\end{aligned}$$

where

$$\begin{aligned}&E_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k})\\ &= \sum_{f,\alpha}\Delta^\alpha_f[u^*_{f\mu}(\mathbf{k}+\mathbf{q})u_{f\mu'}(\mathbf{k}) + v_{f\mu}(-\mathbf{k}-\mathbf{q}v^*_{f\mu'}(-\mathbf{k})] \quad (19a)\end{aligned}$$

and

$$\begin{aligned}\Delta^\alpha_f = E^f_s - E^0_s + \sum_{1\neq s}[&\langle 0,f|\,V_{ss1}\,|0,f\rangle\\ &- \langle 00|\,V_{ss_1}\,|00\rangle]. \quad (19b)\end{aligned}$$

In deriving (18)–(19b), the overlap of the electronic wavefunctions has been neglected. In (19a) and (19b), $E^f_s - E^0_s$ is equal to the energy of excitation of the isolated molecule α in the fth excited state and $u_{f\mu'}(\mathbf{k})$

[11] A. S. Davydov, Phys. Status Solidi **20**, 143 (1967).
[12] A. S. Davydov and E. N. Myasnikov, Phys. Status Solidi **20**, 153 (1967).

and $v_{f\mu'}(-\mathbf{k})$ are the amplitudes that diagonalize the bare unperturbed exciton Hamiltonian. The functions $F^{(1)}_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k})$ and $F^{(2)}_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k})$ describe the coupling between the intermolecular interactions and the acoustic phonons. The expressions (18), (19a), and (19b), with the exception of the second terms on the r.h.s. of (18) and (19b), which are small, have been derived by Ovander[5] where the expression for $F^{(1)}_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k})$ is given. He has shown that the ratio of $E_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k})/F^{(1)}_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k})$ is of the order of E/V', where E is the energy of excitation of an isolated molecule and V' is the energy of interaction between the molecules; since $E \gg V'$, the function $F^{(1)}_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k})$ has been ignored in his treatment. The x component of the coupling function $F^{(1)}_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k})$ in the dipole–dipole approximation has been considered by Agranovich and Konobeev[4] in their study of the long-wavelength edge of the exciton absorption band. Explicit expressions of the coupling functions $F^{(1)}_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k})$ that result from the interaction between bare excitons with acoustic as well as with optical phonons have been recently derived by Davydov[11] for one- and three-dimensional models of molecular crystals of certain symmetry and have been used by Davydov and Myasnikov[12] for the calculation of the dielectric permeability corresponding to the bare exciton spectrum. Davydov[11] also derived a coupling function that results from the derivative of the second term on the r.h.s. of (19b) with respect to the intermolecular distance, but he has ignored in his treatment the coupling functions $E_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k})$ and $F^{(1)}_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k})$.

In discussing the coupling functions that appear in the expression (18), we may regard the coupling function $E_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k})\,Y_r(\mathbf{q})$ as that corresponding to an isolated bare exciton dressed with phonons, while the term $F^{(1)}_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k})$ results from the matrix elements causing the transfer of energy from one molecule to the other; therefore, the coupling function $F^{(1)}_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k})\,Y_r(\mathbf{q})$ accounts for the transfer of energy of the coupled bare exciton–phonon system between neighboring molecules in the crystal. Similarly, the coupling function $F^{(2)}_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k})\,Y_r(\mathbf{q})$ arises from the simultaneous creation or annihilation of two bare excitons coupled with phonons. We refer to the previously mentioned papers regarding the expression for $F^{(1)}_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k})$. As far as the expression for $F^{(2)}_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k})$ is concerned, it is obtained from that of $F^{(1)}_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k})$ by replacing the resonance matrix elements $\langle 0, f|\,V\,|f', 0\rangle$ by the matrix elements corresponding to the simultaneous excitation of two electrons, $\langle 0, 0|\,V\,|f,f'\rangle$. The bare exciton–phonon interaction due to dipole–dipole interactions between the molecules in molecular crystals at low temperatures has been elegantly discussed by Agranovich and Konobeev[13]; we refer to their paper for details. Substituting the expressions for the u's and v's into (18), we find

$$
\begin{aligned}
&W_{\mathbf{q}r}(\mathbf{k}+\mathbf{q},\rho',\lambda';\mathbf{k},\rho,\lambda)\\
&=\frac{\omega_p^2}{4N^{\frac{1}{2}}}\sum_{\mu,\mu'}[E_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k})+F^{(1)}_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k})]Y_r(\mathbf{q})\\
&\times[\omega_{\rho\lambda}(\mathbf{k})\omega_{\rho'\lambda'}(\mathbf{k}+\mathbf{q})]^{-\frac{1}{2}}\\
&\times\left\{\frac{[f_{0\mu}(\mathbf{k}+\mathbf{q},\lambda')f_{0\mu'}(\mathbf{k},\lambda)E_{\mu}(\mathbf{k}+\mathbf{q})E_{\mu'}(\mathbf{k})]^{\frac{1}{2}}}{[E_{\mu'}(\mathbf{k})-\omega_{\rho\lambda}(\mathbf{k})][E_{\mu'}(\mathbf{k}+\mathbf{q})-\omega_{\rho'\lambda'}(\mathbf{k}+\mathbf{q})]}\right.\\
&\left.+\frac{[(f_{0\mu'}(\mathbf{k},\lambda)f_{0\mu}(\mathbf{k}+\mathbf{q},\lambda')E_{\mu'}(\mathbf{k})E_{\mu}(\mathbf{k}+\mathbf{q})]^{\frac{1}{2}}}{[E_{\mu'}(\mathbf{k})+\omega_{\rho\lambda}(\mathbf{k})][E_{\mu}(\mathbf{k}+\mathbf{q})+\omega_{\rho'\lambda'}(\mathbf{k}+\mathbf{q})]}\right\}\\
&\times\left(\frac{d\omega^2}{d\omega^2}\hat{\eta}^2(\mathbf{k},\omega)\right)^{-\frac{1}{2}}_{\omega=\omega_{\rho\lambda}(\mathbf{k})}\left(\frac{d\omega^2}{d\omega^2}\hat{\eta}^2(\mathbf{k}+\mathbf{q},\omega)\right)^{-\frac{1}{2}}_{\omega=\omega_{\rho'\lambda'}(\mathbf{k}+\mathbf{q})}\\
&\times(n_0-n_{\mu'})^{\frac{1}{2}}(n_0-n_{\mu})^{\frac{1}{2}}+\frac{\omega_p^2}{4N^{\frac{1}{2}}}\\
&\times\sum_{\mu,\mu'}F^{(2)}_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k})Y_r(\mathbf{q})[\omega_{\rho\lambda}(\mathbf{k})\omega_{\rho'\lambda'}(\mathbf{k}+\mathbf{q})]^{-\frac{1}{2}}\\
&\times\left\{\frac{[f_{0\mu}(\mathbf{k}+\mathbf{q},\lambda')f_{0\mu'}(\mathbf{k},\lambda)E_{\mu}(\mathbf{k}+\mathbf{q})E_{\mu'}(\mathbf{k})]^{\frac{1}{2}}}{[E_{\mu}(\mathbf{k}+\mathbf{q})+\omega_{\rho'\lambda'}(\mathbf{k}+\mathbf{q})][E_{\mu'}(\mathbf{k})-\omega_{\rho\lambda}(\mathbf{k})]}\right.\\
&\left.+\frac{[f_{0\mu'}(\mathbf{k},\lambda)f_{0\mu}(\mathbf{k}+\mathbf{q},\lambda')E_{\mu}(\mathbf{k})E_{\mu'}(\mathbf{k}+\mathbf{q})]^{\frac{1}{2}}}{[E_{\mu}(\mathbf{k}+\mathbf{q})-\omega_{\rho'\lambda'}(\mathbf{k}+\mathbf{q})][E_{\mu'}(\mathbf{k})+\omega_{\rho\lambda}(\mathbf{k})]}\right\}\\
&\times\left(\frac{d\omega^2}{d\omega^2}\hat{\eta}^2(\mathbf{k},\omega)\right)^{-\frac{1}{2}}_{\omega=\omega_{\rho\lambda}(\mathbf{k})}\\
&\times\left(\frac{d\omega^2}{d\omega^2}\hat{\eta}^2(\mathbf{k}+\mathbf{q},\omega)\right)^{-\frac{1}{2}}_{\omega=\omega_{\rho'\lambda'}(\mathbf{k}+\mathbf{q})}\\
&\times(n_0-n_{\mu'})^{\frac{1}{2}}(n_0-n_{\mu})^{\frac{1}{2}}. \qquad (20)
\end{aligned}
$$

Similarly, the nondiagonal coupling constant $\hat{W}_{\mathbf{q}r}(\mathbf{k}+\mathbf{q},\rho',\lambda';\mathbf{k},\rho,\lambda)$ that results from the simultaneous excitation of two polaritons turns out to be

$$
\begin{aligned}
&\hat{W}_{\mathbf{q}r}(\mathbf{k}+\mathbf{q},\rho',\lambda';\mathbf{k},\rho,\lambda)\\
&=\frac{\omega_p^2}{4N^{\frac{1}{2}}}\sum_{\mu,\mu'}[E_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k})+F^{(1)}_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k})]Y_r(\mathbf{q})\\
&\quad\times[\omega_{\rho\lambda}(\mathbf{k})\,\omega_{\rho'\lambda'}(\mathbf{k}+\mathbf{q})]^{-\frac{1}{2}}\\
&\quad\times\frac{[f_{0\mu}(\mathbf{k}+\mathbf{q},\lambda')f_{0\mu'}(\mathbf{k},\lambda)E_{\mu}(\mathbf{k}+\mathbf{q})E_{\mu'}(\mathbf{k})]^{\frac{1}{2}}}{[E_{\mu}(\mathbf{k}+\mathbf{q})-\omega_{\rho'\lambda'}(\mathbf{k}+\mathbf{q})][E_{\mu'}(\mathbf{k})+\omega_{\rho\lambda}(\mathbf{k})]}\\
&\quad\times(n_0-n_{\mu})^{\frac{1}{2}}(n_0-n_{\mu'})^{\frac{1}{2}}\left(\frac{d\omega^2}{d\omega^2}\hat{\eta}^2_{\lambda}(\mathbf{k},\omega)\right)^{-\frac{1}{2}}_{\omega=\omega_{\rho\lambda}(\mathbf{k})}\\
&\quad\times\left(\frac{d\omega^2}{d\omega^2}\hat{\eta}^2_{\lambda'}(\mathbf{k}+\mathbf{q},\omega)\right)^{-\frac{1}{2}}_{\omega=\omega_{\rho'\lambda'}(\mathbf{k}+\mathbf{q})}
\end{aligned}
$$

[13] V. M. Agranovich and Yu. V. Konobeev, Fiz. Tverd. Tela 6, 831 (1964) [Sov. Phys.—Solid State 6, 644 (1964)].

$$+\frac{\omega_p^2}{8N^{\frac{1}{2}}}\sum_{\mu,\mu'}F^{(2)}_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k})Y_r(\mathbf{q})[\omega_{\rho\lambda}(\mathbf{k})\omega_{\rho'\lambda'}(\mathbf{k}+\mathbf{q})]^{-\frac{1}{2}}$$
$$\times\left\{\frac{[f_{0\mu}(\mathbf{k}+\mathbf{q},\lambda')f_{0\mu'}(\mathbf{k},\lambda)E_\mu(\mathbf{k}+\mathbf{q})E_{\mu'}(\mathbf{k})]^{\frac{1}{2}}}{[E_{\mu'}(\mathbf{k})-\omega_{\rho\lambda}(\mathbf{k})][E_\mu(\mathbf{k}+\mathbf{q})-\omega_{\rho'\lambda'}(\mathbf{k}+\mathbf{q})]}\right.$$
$$\left.+\frac{[f_{0\mu'}(\mathbf{k},\lambda)f_{0\mu}(\mathbf{k}+\mathbf{q},\lambda')E_{\mu'}(\mathbf{k})E_\mu(\mathbf{k}+\mathbf{q})]^{\frac{1}{2}}}{[E_{\mu'}(\mathbf{k})+\omega_{\rho\lambda}(\mathbf{k})][E_\mu(\mathbf{k}+\mathbf{q})+\omega_{\rho'\lambda'}(\mathbf{k}+\mathbf{q})]}\right\}$$
$$\times(n_0-n_\mu)^{\frac{1}{2}}(n_0-n_{\mu'})^{\frac{1}{2}}\left(\frac{d\omega^2}{d\omega^2}\hat{\eta}^2_\lambda(\mathbf{k},\omega)\right)^{-\frac{1}{2}}_{\omega=\omega_{\rho\lambda}(\mathbf{k})}$$
$$\times\left(\frac{d\omega^2}{d\omega^2}\hat{\eta}^2_{\lambda'}(\mathbf{k}+\mathbf{q},\omega)\right)^{-\frac{1}{2}}_{\omega=\omega_{\rho'\lambda'}(\mathbf{k}+\mathbf{q})}. \tag{21}$$

It is easy to show that the coupling constants in the expression (15) satisfy the relations

$$\Phi_{qr}(\mathbf{k}+\mathbf{q},\rho',\lambda';\mathbf{k},\rho,\lambda)=\Phi^*_{qr}(\mathbf{k},\rho,\lambda;\mathbf{k}+\mathbf{q},\rho',\lambda')$$

and

$$W_{qr}(\mathbf{k}+\mathbf{q},\rho',\lambda';\mathbf{k},\rho,\lambda)$$
$$=W^*_{qr}(\mathbf{k},\rho,\lambda;\mathbf{k}+\mathbf{q},\rho',\lambda'),$$

i.e., $\mathcal{H}_{\text{int}}$ is Hermitian. We note that in the first term on the r.h.s. of the interaction Hamiltonian (15), the diagonal and nondiagonal coupling constants with respect to the polariton operators are equal, while a comparison between the first terms on the r.h.s. of (20) and (21) yields $(\hat{W}_{qr}/W_{qr})\sim[E_{\mu'}(\mathbf{k})-\omega_{\rho\lambda}(\mathbf{k})]/[E_{\mu'}(\mathbf{k})+\omega_{\rho\lambda}(\mathbf{k})]$, and the inverse ratio results from the comparison of the corresponding last two terms. In the expression (17) for $\Phi_{qr}(\mathbf{k}+\mathbf{q},\rho',\lambda';\mathbf{k},\rho,\lambda)$, both terms are of the same order of magnitude, while a comparison between the first term in (17) with the first term in (20) leads to

$$(\Phi_{qr}/W_{qr})\sim[\omega_{\rho'\lambda'}(\mathbf{k}+\mathbf{q})-E_{\mu'}(\mathbf{k}+\mathbf{q})]/E_{\mu\mu'}(\mathbf{k}+\mathbf{q},\mathbf{k}).$$

IV. GREEN'S FUNCTIONS

Having written a model Hamiltonian in the previous section, we now evaluate the Green's functions appearing in Eq. (11) for the dielectric function.

A. Polariton–Phonon Field

If we introduce the notation

$$G(\mathbf{k},\rho,\lambda;\mathbf{k},\rho,\lambda;\omega)\equiv G(\mathbf{k},\rho,\lambda;\omega)\equiv\langle\langle\zeta_{\rho\lambda}(\mathbf{k});\tilde{\zeta}^+_{\rho\lambda}(\mathbf{k})\rangle\rangle$$

and

$$\hat{G}(-\mathbf{k},\rho,\lambda;\mathbf{k},\rho,\lambda;\omega)\equiv G(-\mathbf{k},\rho,\lambda;\omega)\equiv\langle\langle\zeta^+_{\rho\lambda}(-\mathbf{k});\tilde{\zeta}^+_{\rho\lambda}(\mathbf{k})\rangle\rangle,$$

then, using the Hamiltonian (13), we find the following equations of motion for the Green's functions $G(\mathbf{k},\rho,\lambda;\omega)$ and $\hat{G}(-\mathbf{k},\rho,\lambda;\omega)$, respectively:

$$[\omega-\omega_{\rho\lambda}(\mathbf{k})]G(\mathbf{k},\rho,\lambda;\omega)$$
$$=(1/2\pi)+\sum_{\substack{\mathbf{q},r,\rho',\lambda'\\ \rho_1,\lambda_1}}\alpha_{\rho\lambda}(\mathbf{k},\mathbf{q})[\alpha^*_{\rho_1\lambda_1}(\mathbf{k},\mathbf{q})G(\mathbf{k},\rho_1,\lambda_1;\omega)$$
$$+\hat{\alpha}^*_{\rho_1\lambda_1}(\mathbf{k},\mathbf{q})\hat{G}(-\mathbf{k},\rho_1,\lambda_1;\omega)]D^{(0)}_{qr}(\mathbf{k}-\mathbf{q},\rho',\lambda';\omega)$$
$$+\sum_{\substack{\mathbf{q},r,\rho',\lambda'\\ \rho_1,\lambda_1}}\hat{\alpha}_{\rho\lambda}(\mathbf{k},\mathbf{q})[\hat{\alpha}^*_{\rho_1\lambda_1}(\mathbf{k},\mathbf{q})G(\mathbf{k},\rho_1,\lambda_1;\omega)$$
$$+\alpha^*_{\rho_1\lambda_1}(\mathbf{k},\mathbf{q})\hat{G}(-\mathbf{k},\rho_1,\lambda_1;\omega)]$$
$$\times D^{(0)}_{qr}(\mathbf{k}-\mathbf{q},\rho',\lambda';-\omega), \tag{22}$$

$$[-\omega-\omega_{\rho\lambda}(\mathbf{k})]\hat{G}(-\mathbf{k},\rho,\lambda;\omega)$$
$$=(1/2\pi)+\sum_{\substack{\mathbf{q},r,\rho',\lambda'\\ \rho_1,\lambda_1}}\alpha_{\rho\lambda}(\mathbf{k},\mathbf{q})[\alpha^*_{\rho_1\lambda_1}(\mathbf{k},\mathbf{q})\hat{G}(-\mathbf{k},\rho_1,\lambda_1;\omega)$$
$$+\hat{\alpha}^*_{\rho_1\lambda_1}(\mathbf{k},\mathbf{q})G(\mathbf{k},\rho_1,\lambda_1;\omega)]D^{(0)}_{qr}(\mathbf{k}-\mathbf{q},\rho',\lambda';-\omega)$$
$$+\sum_{\substack{\mathbf{q},r,\rho',\lambda'\\ \rho_1,\lambda_1}}\hat{\alpha}_{\rho\lambda}(\mathbf{k},\mathbf{q})[\hat{\alpha}^*_{\rho_1\lambda_1}(\mathbf{k},\mathbf{q})\hat{G}(-\mathbf{k},\rho_1,\lambda_1;\omega)$$
$$+\alpha^*_{\rho_1\lambda_1}(\mathbf{k},\mathbf{q})G(\mathbf{k},\rho_1,\lambda_1;\omega)]D^{(0)}_{qr}(\mathbf{k}-\mathbf{q},\rho',\lambda';\omega), \tag{23}$$

where $D^{(0)}_{qr}(\mathbf{k}-\mathbf{q},\rho',\lambda';\omega)$ is the unperturbed Green's function corresponding to the polariton–phonon field and is equal to

$$D^{(0)}_{qr}(\mathbf{k}-\mathbf{q},\rho',\lambda';\omega)=\frac{1+N_r(\mathbf{q})}{\omega-\omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q})-\omega_r(\mathbf{q})}$$
$$+\frac{N_r(\mathbf{q})}{\omega-\omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q})+\omega_r(\mathbf{q})} \tag{24}$$

and $N_r(\mathbf{q})\equiv\langle\beta^+_r(\mathbf{q})\beta_r(\mathbf{q})\rangle$ is the phonon occupation number. The functions $\alpha_{\rho\lambda}(\mathbf{k},\mathbf{q})$ and $\hat{\alpha}_{\rho\lambda}(\mathbf{k},\mathbf{q})$ have been defined as

$$\alpha_{\rho\lambda}(\mathbf{k},\mathbf{q})\equiv\Phi_{qr}(\mathbf{k},\rho,\lambda;\mathbf{k}-\mathbf{q},\rho',\lambda')$$
$$+\Phi_{qr}(\mathbf{q}-\mathbf{k},\rho',\lambda';-\mathbf{k},\rho,\lambda)$$
$$+W_{qr}(\mathbf{k},\rho,\lambda;\mathbf{k}-\mathbf{q},\rho',\lambda'), \tag{25}$$

$$\hat{\alpha}_{\rho\lambda}(\mathbf{k},\mathbf{q})\equiv\Phi_{qr}(\mathbf{k},\rho,\lambda;\mathbf{k}-\mathbf{q},\rho',\lambda')$$
$$+\Phi_{qr}(\mathbf{q}-\mathbf{k},\rho',\lambda';-\mathbf{k},\rho,\lambda)$$
$$+\hat{W}_{qr}(\mathbf{k},\rho,\lambda;\mathbf{k}-\mathbf{q},\rho',\lambda')$$
$$+\hat{W}_{qr}(\mathbf{q}-\mathbf{k},\rho',\lambda';-\mathbf{k},\rho,\lambda). \tag{26}$$

In deriving the coupled equations (22) and (23), we have made use of the simple decoupling procedure for the higher-order Green's functions, i.e.,

$$\langle\langle\zeta_{\rho_1\lambda_1}(\mathbf{k}-\mathbf{q}-\mathbf{q}')\tilde{\beta}_r(\mathbf{q})\tilde{\beta}^+_r(\mathbf{q});\tilde{\zeta}^+_{\rho\lambda}(\mathbf{k})\rangle\rangle$$
$$\approx[1+2N_r(\mathbf{q})]\delta_{\mathbf{q}',-\mathbf{q}}\delta_{rr'}G(\mathbf{k},\rho_1,\lambda_1;\omega), \tag{27a}$$

$$\langle\langle\zeta_{\rho_1\lambda_1}(\mathbf{k}-\mathbf{q}-\mathbf{q}')[\beta_r(\mathbf{q})-\beta^+_r(-\mathbf{q})]\tilde{\beta}_{r'}{}^+(\mathbf{q});\tilde{\zeta}^+_{\rho\lambda}(\mathbf{k})\rangle\rangle$$
$$\approx\delta_{\mathbf{q}',-\mathbf{q}}\delta_{rr'}G(\mathbf{k},\rho_1,\lambda_1;\omega), \tag{27b}$$

$$\langle\langle\tilde{\zeta}^+_{\rho''\lambda''}(\mathbf{k}-\mathbf{q})\tilde{\zeta}_{\rho_1\lambda_1}(\mathbf{k})\zeta_{\rho'\lambda'}(\mathbf{k}-\mathbf{q});\tilde{\zeta}^+_{\rho\lambda}(\mathbf{k})\rangle\rangle$$
$$\approx\langle\zeta^+_{\rho'\lambda'}(\mathbf{k}-\mathbf{q})\zeta_{\rho'\lambda'}(\mathbf{k}-\mathbf{q})\rangle^{(0)}\delta_{\rho''\rho'}\delta_{\lambda''\lambda'}$$
$$\times[G(\mathbf{k},\rho_1,\lambda_1;\omega)+\hat{G}(-\mathbf{k},\rho_1,\lambda_1;\omega)], \tag{27c}$$

and we have set equal to zero terms proportional to $\langle\zeta^+_{\rho'\lambda'}(\mathbf{k}-\mathbf{q})\zeta_{\rho'\lambda'}(\mathbf{k}-\mathbf{q})\rangle^{(0)}$ because they give negligibly small contributions. In (27a)–(27c), the decoupling approximation has been carried out by pairing off, where possible, operators having the same time arguments. The physical process described here corresponds to the scattering of a polariton from $(\mathbf{k}, \rho, \lambda)$ to $(\mathbf{k}-\mathbf{q}, \rho', \lambda')$, with the emission of a $(\mathbf{q}, r)$ phonon or the absorption of a $(-\mathbf{q}, r)$ phonon. The remaining phonons create a self-consistent field determined by the phonon density $\langle\beta^+_r(\mathbf{q})\beta_r(\mathbf{q})\rangle$. The propagator which describes such a process is given by the expression (24).

In the same approximation, we derive the following coupled equations for the Green's functions

$$\Gamma_{\mathbf{q}r}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; \mathbf{k}-\mathbf{q}, \rho', \lambda'; \omega) \equiv \Gamma_{\mathbf{q}r}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; \omega) \equiv \langle\langle\zeta_{\rho'\lambda'}(\mathbf{k}-\mathbf{q})\tilde{\beta}_r(\mathbf{q}); \tilde{\zeta}^+_{\rho'\lambda'}(\mathbf{k}-\mathbf{q})\tilde{\beta}^+_r(\mathbf{q})\rangle\rangle$$

and

$$\hat{\Gamma}_{\mathbf{q}r}(\mathbf{q}-\mathbf{k}, \rho', \lambda'; \mathbf{k}-\mathbf{q}, \rho', \lambda'; \omega) \equiv \hat{\Gamma}_{\mathbf{q}r}(\mathbf{q}-\mathbf{k}, \rho', \lambda'; \omega) \equiv \langle\langle\zeta^+_{\rho'\lambda'}(\mathbf{q}-\mathbf{k})\tilde{\beta}_r(\mathbf{q}); \tilde{\zeta}^+_{\rho'\lambda'}(\mathbf{k}-\mathbf{q})\tilde{\beta}^+_r(\mathbf{q})\rangle\rangle,$$

respectively:

$$\begin{aligned}
&D^{(0)-1}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; \omega)\Gamma_{\mathbf{q}r}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; \omega)\\
&= (1/2\pi) + \sum_{\substack{\mathbf{q},r,\rho,\lambda\\ \rho_1,\lambda_1}} \alpha^*_{\rho\lambda}(\mathbf{k}, \mathbf{q})[\alpha_{\rho_1\lambda_1}(\mathbf{k}, \mathbf{q})\Gamma_{\mathbf{q}r}(\mathbf{k}-\mathbf{q}, \rho_1, \lambda_1; \omega)\\
&\quad + \hat{\alpha}_{\rho_1\lambda_1}(\mathbf{k}, \mathbf{q})\hat{\Gamma}_{\mathbf{q}r}(\mathbf{q}-\mathbf{k}, \rho_1, \lambda_1; \omega)]G^{(0)}_{\rho\lambda}(\mathbf{k}; \omega)\\
&\quad + \sum_{\substack{\mathbf{q},r,\rho,\lambda\\ \rho_1,\lambda_1}} \hat{\alpha}^*_{\rho\lambda}(\mathbf{k}, \mathbf{q})[\alpha_{\rho_1\lambda_1}(\mathbf{k}, \mathbf{q})\hat{\Gamma}_{\mathbf{q}r}(\mathbf{q}-\mathbf{k}, \rho_1, \lambda_1; \omega)\\
&\quad + \hat{\alpha}_{\rho_1\lambda_1}(\mathbf{k}, \mathbf{q})\Gamma_{\mathbf{q}r}(\mathbf{k}-\mathbf{q}, \rho_1, \lambda_1; \omega)]G^{(0)}_{\rho\lambda}(\mathbf{k}; -\omega),
\end{aligned} \tag{28}$$

$$\begin{aligned}
&D^{(0)-1}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; -\omega)\hat{\Gamma}_{\mathbf{q}r}(\mathbf{q}-\mathbf{k}, \rho', \lambda'; \omega)\\
&= (1/2\pi) + \sum_{\substack{\mathbf{q},r,\rho,\lambda\\ \rho_1,\lambda_1}} \alpha^*_{\rho\lambda}(\mathbf{k}, \mathbf{q})[\alpha_{\rho_1\lambda_1}(\mathbf{k}, \mathbf{q})\hat{\Gamma}_{\mathbf{q}r}(\mathbf{q}-\mathbf{k}, \rho_1, \lambda_1; \omega)\\
&\quad + \hat{\alpha}_{\rho_1\lambda_1}(\mathbf{k}, \mathbf{q})\Gamma_{\mathbf{q}r}(\mathbf{k}-\mathbf{q}, \rho_1, \lambda_1; \omega)]G^{(0)}_{\rho\lambda}(\mathbf{k}; -\omega)\\
&\quad + \sum_{\substack{\mathbf{q},r,\rho,\lambda\\ \rho_1,\lambda_1}} \hat{\alpha}^*_{\rho\lambda}(\mathbf{k}, \mathbf{q})[\alpha_{\rho_1\lambda_1}(\mathbf{k}, \mathbf{q})\Gamma_{\mathbf{q}r}(\mathbf{k}-\mathbf{q}, \rho_1, \lambda_1; \omega)\\
&\quad + \hat{\alpha}_{\rho_1\lambda_1}(\mathbf{k}, \mathbf{q})\hat{\Gamma}_{\mathbf{q}r}(\mathbf{q}-\mathbf{k}, \rho_1, \lambda_1; \omega)]G^{(0)}_{\rho\lambda}(\mathbf{k}; \omega),
\end{aligned} \tag{29}$$

where

$$G^{(0)}_{\rho\lambda}(\mathbf{k}; \omega) = [\omega - \omega_{\rho\lambda}(\mathbf{k})]^{-1}$$

is the unperturbed polariton Green's function. Similarly, we find that the equation of motion for the Green's function

$$\langle\langle\tilde{\zeta}_{\rho'\lambda'}(\mathbf{k}-\mathbf{q})\tilde{\beta}_r(\mathbf{q}); \tilde{\zeta}^+_{\rho\lambda}(\mathbf{k})\rangle\rangle$$

is

$$\begin{aligned}
&\langle\langle\tilde{\zeta}_{\rho'\lambda'}(\mathbf{k}-\mathbf{q})\tilde{\beta}_r(\mathbf{q}); \tilde{\beta}_{\rho\lambda}(\mathbf{k})\rangle\rangle\\
&= \sum_{\rho_1,\lambda_1} \alpha^*_{\rho_1\lambda_1}(\mathbf{k}, \mathbf{q})[D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; \omega)G(\mathbf{k}, \rho_1, \lambda_1; \omega)\\
&\quad + D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; -\omega)\hat{G}(-\mathbf{k}, \rho_1, \lambda_1; \omega)]\\
&\quad + \sum_{\rho_1\lambda_1} \hat{\alpha}^*_{\rho_1,\lambda_1}(\mathbf{k}, \mathbf{q})[D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; -\omega)\\
&\quad \times G(\mathbf{k}, \rho_1, \lambda_1; \omega)\\
&\quad + D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; \omega)\hat{G}(-\mathbf{k}, \rho_1, \lambda_1; \omega)].
\end{aligned} \tag{30}$$

To solve the system of equations (22) and (23) as a first approximation, we shall neglect mixing of different polariton bands. This means that we make the assumption that the bands $\rho\lambda$ and $\rho_1\lambda_1 \neq \rho\lambda$ are separated by large energy gap so that the contributions from the terms with $\rho_1\lambda_1 \neq \rho\lambda$ are negligibly small. Then, using the fact that in the complex ω plane the relation $D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; -\omega) = D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; \omega)$ holds, we derive from (22) and (23) the following expressions:

$$G(\mathbf{k}, \rho, \lambda; \omega) = (1/2\pi)\frac{\omega + \Omega^{(2)}_{\mathbf{k}\rho}(\omega)}{[\omega^2 - \Omega^{(1)}_{\mathbf{k}\rho}(\omega)\Omega^{(2)}_{\mathbf{k}\rho}(\omega)]}, \tag{31}$$

$$\hat{G}(-\mathbf{k}, \rho, \lambda; \omega) = (1/2\pi)\frac{-\omega + \Omega^{(2)}_{\mathbf{k}\rho}(\omega)}{[\omega^2 - \Omega^{(1)}_{\mathbf{k}\rho}(\omega)\Omega^{(2)}_{\mathbf{k}\rho}(\omega)]}, \tag{32}$$

and hence

$$G(\mathbf{k}, \rho, \lambda; \omega) + \hat{G}(-\mathbf{k}, \rho, \lambda; \omega) = (1/\pi)\Omega^{(2)}_{\mathbf{k}\rho}(\omega)/[\omega^2 - \Omega^{(1)}_{\mathbf{k}\rho}(\omega)\Omega^{(2)}_{\mathbf{k}\rho}(\omega)], \tag{33}$$

where $\Omega^{(1)}_{\mathbf{k}\rho}(\omega)$ and $\Omega^{(2)}_{\mathbf{k}\rho}(\omega)$ are equal to

$$\Omega^{(1)}_{\mathbf{k}\rho}(\omega) = \omega_{\rho\lambda}(\mathbf{k}) + \sum_{\mathbf{q},r,\rho',\lambda'} |\alpha_{\rho\lambda}(\mathbf{k}, \mathbf{q}) + \hat{\alpha}_{\rho\lambda}(\mathbf{k}, \mathbf{q})|^2 \times D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; \omega) \tag{34}$$

and

$$\Omega^{(2)}_{\mathbf{k}\rho}(\omega) = \omega_{\rho\lambda}(\mathbf{k}) + \sum_{\mathbf{q},r,\rho',\lambda'} |\alpha_{\rho\lambda}(\mathbf{k}, \mathbf{q}) - \hat{\alpha}_{\rho\lambda}(\mathbf{k}, \mathbf{q})|^2 \times D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; \omega). \tag{35}$$

Thus the Green's functions $G(\mathbf{k}, \rho, \lambda; \omega)$ and $\hat{G}(-\mathbf{k}, \rho, \lambda; \omega)$ given by expressions (31) and (32) have the same poles that give the energies of excitation for the polariton–phonon spectrum. Considering the definition of $\alpha_{\rho\lambda}(\mathbf{k}, \mathbf{q})$ and $\hat{\alpha}_{\rho\lambda}(\mathbf{k}, \mathbf{q})$, expressions (34) and (35) then show that the coupling constant of the perturbed part of $\Omega^{(2)}_{\mathbf{k}}(\omega)$ depends entirely on $W_{\mathbf{q}r}$ and $\hat{W}_{\mathbf{q}r}$, i.e., on the interaction between both the electronic excitation and the intermolecular interactions with the

acoustic phonons, while in the corresponding term of $\Omega_{\mathbf{k}}^{(1)}(\omega)$ the dressed electromagnetic–phonon interaction is also fully included. The perturbed Green's function for the polariton spectrum (31) is reduced to that derived by Agranovich and Konobeev[9] if we consider the particular case studied by these authors, i.e., take the coupling constants as $\alpha_{\rho\lambda}(\mathbf{k}, \mathbf{q}) = W_{\mathbf{qr}}(\mathbf{k}, \rho, \lambda; \mathbf{k}-\mathbf{q}, \rho', \lambda')$, $\hat{\alpha}_{\rho\lambda}(\mathbf{k}, \mathbf{q}) = 0$, $E_{\mu\mu'}(\mathbf{k}, \mathbf{k}-\mathbf{q}) = 0$, and $F_{\mu\mu'}^{(2)}(\mathbf{k}, \mathbf{k}-\mathbf{q}) = 0$ instead of the expressions (25) and (26) that have been used here. In the same approximation, the solution of the system of equations (28) and (29) leads to

$$\Gamma_{\mathbf{qr}}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; \omega) = (2\pi)^{-1} D_{\mathbf{qr}}^{(0)}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; \omega) \times \left\{\omega^2 - \omega_{\rho\lambda}(\mathbf{k})\Omega_{\mathbf{k}\rho}^{(2)}(\omega) + \omega \sum_{\mathbf{q},r,\rho',\lambda'} [|\alpha_{\rho\lambda}(\mathbf{k}, \mathbf{q})|^2 - |\hat{\alpha}_{\rho\lambda}(\mathbf{k}, \mathbf{q})|^2] D_{\mathbf{qr}}^{(0)}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; \omega)\right\} \times [\omega^2 - \Omega_{\mathbf{k}\rho}^{(1)}(\omega)\Omega_{\mathbf{k}\rho}^{(2)}(\omega)]^{-1}, \tag{36}$$

$$\hat{\Gamma}_{\mathbf{q}}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; \omega) = (2\pi)^{-1} D_{\mathbf{qr}}^{(0)}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; \omega) \times \left\{\omega^2 - \omega_{\rho\lambda}(\mathbf{k})\Omega_{\mathbf{k}}^{(2)}(\omega) - \omega \sum_{\mathbf{q},r,\rho',\lambda'} [|\alpha_{\rho\lambda}(\mathbf{k}, \mathbf{q})|^2 - |\hat{\alpha}_{\rho\lambda}(\mathbf{k}, \mathbf{q})|^2] D_{\mathbf{qr}}^{(0)}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; \omega)\right\} \times [\omega^2 - \Omega_{\mathbf{k}\rho}^{(1)}(\omega)\Omega_{\mathbf{k}\rho}^{(2)}(\omega)]^{-1}, \tag{37}$$

and

$$\Gamma_{\mathbf{qr}}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; \omega) + \hat{\Gamma}_{\mathbf{qr}}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; \omega) = (\pi)^{-1} D_{\mathbf{qr}}^{(0)}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; \omega) \frac{[\omega^2 - \omega_{\rho\lambda}(\mathbf{k})\Omega_{\mathbf{k}\rho}^{(2)}(\omega)]}{[\omega^2 - \Omega_{\mathbf{k}\rho}^{(1)}(\omega)\Omega_{\mathbf{k}\rho}^{(2)}(\omega)]} = D_{\mathbf{qr}}^{(0)}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; \omega)\{(\pi)^{-1} + [\Omega_{\mathbf{k}\rho}^{(1)}(\omega) - \omega_{\rho\lambda}(\mathbf{k})][G(\mathbf{k}, \rho, \lambda; \omega) + \hat{G}(-\mathbf{k}, \rho, \lambda; \omega)]\}. \tag{38}$$

Also for the expression (30) we have

$$\langle\langle \tilde{\zeta}_{\rho'\lambda'}(\mathbf{k}-\mathbf{q})\tilde{\beta}_r(\mathbf{q}); \tilde{\zeta}_{\rho\lambda}^{+}(\mathbf{k})\rangle\rangle = D_{\mathbf{qr}}^{(0)}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; \omega)[\alpha_{\rho\lambda}^*(\mathbf{k}, \mathbf{q}) + \hat{\alpha}_{\rho\lambda}^*(\mathbf{k}, \mathbf{q})] \times [G(\mathbf{k}, \rho, \lambda; \omega) + \hat{G}(-\mathbf{k}, \rho, \lambda; \omega)]. \tag{39}$$

Substituting (33), (38), and (39), and its complex conjugate into (11), we obtain the following expression for the dielectric permeability due to the polariton–phonon interaction:

$$\epsilon_\lambda(\mathbf{k}, \omega) = \epsilon_\infty - \frac{\pi\omega_p^2}{\omega^2}(n_0 - n_\mu) \sum_\rho |A_{\rho\lambda}(\mathbf{k}, \omega)|^2 [G(\mathbf{k}, \rho, \lambda; \omega) + \hat{G}(-\mathbf{k}, \rho, \lambda; \omega)] - \frac{\omega_p^2}{\omega^2 N}(n_0 - n_{\mu'}) \times \sum_{\mathbf{q},r,\rho'} \hat{f}_{0\mu'}(\mathbf{k}-\mathbf{q}, \lambda')\omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q}) |Y_r(\mathbf{q})|^2 D_{\mathbf{qr}}^{(0)}(\mathbf{k}-\mathbf{q}, \lambda', \rho'; \omega), \tag{40}$$

where

$$A_{\rho\lambda}(\mathbf{k}, \omega) \equiv [\hat{f}_{0\mu}(\mathbf{k}, \lambda)\omega_{\rho\lambda}(\mathbf{k})]^{\frac{1}{2}} + N^{-\frac{1}{2}} \sum_{\mathbf{q},r,\rho'} [\hat{f}_{0\mu'}(\mathbf{k}-\mathbf{q}, \lambda')\omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q})]^{\frac{1}{2}} Y_r(\mathbf{q}) \times [\alpha_{\rho\lambda}^*(\mathbf{k}, \mathbf{q}) + \hat{\alpha}_{\rho\lambda}^*(\mathbf{k}, \mathbf{q})] D_{\mathbf{qr}}^{(0)}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; \omega). \tag{41}$$

The first term on the r.h.s. of (41) arises from the free-polariton field, while the second describes the coupling function corresponding to the coupled polariton–phonon system. The expression for $A_{\rho\lambda}(\mathbf{k}, \omega)$ is a complex function for values of frequencies ω corresponding to the zeros of the arguments of the Green's function $D_{\mathbf{qr}}^{(0)}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; \omega)$, but the square of its amplitude $|A_{\rho\lambda}(\mathbf{k}, \omega)|^2$ is real for the whole range of frequencies ω. The last term on the r.h.s. of (40) refers to the unperturbed polariton–phonon spectrum through the Green's function $D_{\mathbf{qr}}^{(0)}(\mathbf{k}-\mathbf{q}, \rho', \lambda'; \omega)$. This is due to the fact that we have only taken into account in the Hamiltonian (15) terms linear in the phonon operators and also restricted ourselves through the decoupling approximation only to processes involving the emission or absorption of one phonon.

In the range of wave vectors where retardation effects can be neglected, the expression for the dielectric permeability corresponding to interband excitations in the crystal can be derived from (40) by

taking the limit when $\omega_{\rho\lambda}(\mathbf{k}) = E_\mu(\mathbf{k})$ with the result

$$\epsilon_{ii}(\mathbf{k},\omega) = \epsilon_\infty - \frac{\pi\omega_p^2}{\omega^2}(n_{\mathbf{k}0} - n_{\mathbf{k}\mu})\sum_\mu |A_{\mathbf{k}\mu}(\omega)|^2 [G(\mathbf{k},\mu;\omega) + \hat{G}(-\mathbf{k},\mu;\omega)] - \frac{\omega_p^2}{\omega^2 N}(n_{\mathbf{k}-\mathbf{q}0} - n_{\mathbf{k}-\mathbf{q}\mu'})$$
$$\times \sum_{\mathbf{q},r,\mu'} f_{0\mu'}(\mathbf{k}-\mathbf{q},i)E_{\mu'}(\mathbf{k}-\mathbf{q})\,|Y_r(\mathbf{q})|^2 D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\mu';\omega), \tag{42}$$

where

$$A_{\mathbf{k}\mu}(\omega) = [f_{0\mu}(\mathbf{k},i)E_\mu(\mathbf{k})]^{\frac{1}{2}} + \frac{1}{N^{\frac{1}{2}}}\sum_{\mathbf{q},r,\mu'}[f_{0\mu'}(\mathbf{k}-\mathbf{q},i)E_{\mu'}(\mathbf{k}-\mathbf{q})]^{\frac{1}{2}} Y_r(\mathbf{q})$$
$$\times [\alpha^*_\mu(\mathbf{k},\mathbf{q}) + \hat{\alpha}^*_\mu(\mathbf{k},\mathbf{q})]D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\mu';\omega), \tag{43}$$

and the retarded Green's functions $G(\mathbf{k},\mu;\omega)$, $\hat{G}(-\mathbf{k},\mu;\omega)$, and $D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\mu';\omega)$ are equal to

$$G(\mathbf{k},\mu;\omega) + \hat{G}(-\mathbf{k},\mu;\omega) = \frac{1}{\pi}\frac{\Omega^{(2)}_{\mathbf{k}\mu}(\omega)}{\omega^2 - \Omega^{(1)}_{\mathbf{k}\mu}(\omega)\Omega^{(2)}_{\mathbf{k}\mu}(\omega)}, \tag{44}$$

$$D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\mu';\omega) = \frac{1+N_r(\mathbf{q})}{\omega - E_{\mu'}(\mathbf{k}-\mathbf{q}) - \omega_r(\mathbf{q})} + \frac{N_r(\mathbf{q})}{\omega - E_{\mu'}(\mathbf{k}-\mathbf{q}) + \omega_r(\mathbf{q})}. \tag{45}$$

The expressions for the functions $\Omega^{(1)}_{\mathbf{k}\mu}(\omega)$ and $\Omega^{(2)}_{\mathbf{k}\mu}(\omega)$ are now given by

$$\Omega^{(1)}_{\mathbf{k}\mu}(\omega) = E_\mu(\mathbf{k}) + \sum_{\mathbf{q},r,\mu'}|\alpha_\mu(\mathbf{k},\mathbf{q}) + \hat{\alpha}_\mu(\mathbf{k},\mathbf{q})|^2 D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\mu';\omega), \tag{46a}$$

$$\Omega^{(2)}_{\mathbf{k}\mu}(\omega) = E_\mu(\mathbf{k}) + \sum_{\mathbf{q},r,\mu'}|\alpha_\mu(\mathbf{k},\mathbf{q}) - \hat{\alpha}_\mu(\mathbf{k},\mathbf{q})|^2 D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\mu';\omega), \tag{46b}$$

with

$$\alpha_\mu(\mathbf{k},\mathbf{q}) \equiv W_{\mathbf{q}r}(\mathbf{k},\mu;\mathbf{k}-\mathbf{q},\mu')$$
$$= \frac{1}{N^{\frac{1}{2}}}[E_{\mu\mu'}(\mathbf{k},\mathbf{k}-\mathbf{q}) + F^{(1)}_{\mu\mu'}(\mathbf{k},\mathbf{k}-\mathbf{q})]Y_r(q)(n_0 - n_\mu)^{\frac{1}{2}}(n_0 - n_{\mu'})^{\frac{1}{2}}, \tag{46c}$$

and

$$\hat{\alpha}_\mu(\mathbf{k},\mathbf{q}) \equiv \frac{1}{N^{\frac{1}{2}}}F^{(2)}_{\mu\mu'}(\mathbf{k},\mathbf{k}-\mathbf{q})Y_r(\mathbf{q})(n_0 - n_\mu)^{\frac{1}{2}}(n_0 - n_{\mu'})^{\frac{1}{2}}. \tag{46d}$$

Since $\epsilon_{ii}(\mathbf{k},\omega)$ corresponds to the normal waves, the expression (42) for $i = \mu_\perp = 1, 2$ gives the two transverse components, while for $i = \mu_\parallel = 3$ it gives the longitudinal component for the dielectric permeability. The expression for the Green's function $G_\mu(\mathbf{k},\mu;\omega)$ is reduced to that derived by Davydov and Myasnikov[12] if we take $\hat{\alpha}_\mu(\mathbf{k},\mathbf{q}) = 0$ and $E_{\mu\mu'}(\mathbf{k},\mathbf{k}-\mathbf{q}) = 0$; physically, it corresponds to the extreme case, which occurs only when the resonance interaction between neighboring molecules dominates the interactions involved in the expressions for $E_{\mu\mu'}(\mathbf{k},\mathbf{k}+\mathbf{q})$ and $\hat{\alpha}_\mu(\mathbf{k},\mathbf{q})$. We note here that the coupling constant $\hat{\alpha}_\mu(\mathbf{k},\mathbf{q})$ given by (46d) depends entirely on the coupling between the matrix elements responsible for the simultaneous creation or annihilation of two bare excitons and the phonon field; therefore $\hat{\alpha}_\mu(\mathbf{k},\mathbf{q}) < \alpha_\mu(\mathbf{k},\mathbf{q})$. However, since $F^{(2)}_{\mu\mu'}(\mathbf{k},\mathbf{k}-\mathbf{q})$ is of the same order of magnitude as $F^{(1)}_{\mu\mu'}(\mathbf{k},\mathbf{k}-\mathbf{q})$, both should be taken into account, especially for crystals having large transition dipole moments. The Green's function $D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\mu';\omega)$ in the limit of zero temperature, $N_r(\mathbf{q}) \approx 0$, is identical to that derived by Rashba.[14]

B. Phonon Field

The equation of motion for the phonon Green's function, $\langle\langle\tilde{\beta}_r(\mathbf{q});\tilde{\beta}^+_r(\mathbf{q})\rangle\rangle$, is easily evaluated by means of the Hamiltonian (13) with the result

$$[\omega^2 - \omega_r^2(\mathbf{q}) - \omega_r(\mathbf{q})P_{\mathbf{q}r}(\omega^2)]$$
$$\times \langle\langle\tilde{\beta}_r(\mathbf{q});\tilde{\beta}^+_r(\mathbf{q})\rangle\rangle = \omega_r(\mathbf{q})/\pi, \tag{47}$$

where the polarization operator is equal to

$$P_{\mathbf{q}r}(\omega^2) = \sum_{\substack{\mathbf{k},\rho,\rho'\\ \lambda,\lambda'}}|\hat{\alpha}_{\rho\lambda}(\mathbf{k},q)|^2$$
$$\times [(\omega - \omega_{\rho\lambda}(\mathbf{k}) - \omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q}))^{-1}$$
$$+ (-\omega - \omega_{\rho\lambda}(\mathbf{k}) - \omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q}))^{-1}]. \tag{48}$$

In deriving (47) and (48) we have made use of the decoupling approximations (27) and set equal to zero terms proportional to the polariton occupation numbers. The expression (48) describes the physical process where two polaritons $(\mathbf{k},\rho)$ and $(\mathbf{k}-\mathbf{q},\rho')$ are created or annihilated simultaneously through the

[14] E. I. Rashba, Zh. Eksp. Teor. Fiz. **50**, 1064 (1966) [Sov. Phys.—JETP **23**, 708 (1966)].

exchange of a phonon $(\mathbf{q}, r)$. The spectral representation for the phonon spectrum turns out to be

$$J_{\mathbf{q}r}(\omega) = 2\,\mathrm{Im}\,\langle\langle\tilde{\beta}_r(\mathbf{q});\tilde{\beta}_r^+(\mathbf{q})\rangle\rangle(1-e^{\beta\omega})^{-1} = \frac{2\omega_r(\mathbf{q})}{\pi}\times\frac{\omega_r(\mathbf{q})\gamma_{\mathbf{q}r}(\omega)(e^{\beta\omega}-1)^{-1}}{[\omega^2-\omega_r^2(\mathbf{q})-\omega_r(\mathbf{q})\,\mathrm{Re}\,P_{\mathbf{q}r}(\omega^2)]^2+[\omega_r(\mathbf{q})\gamma_{\mathbf{q}r}(\omega^2)]^2}, \tag{49}$$

where

$$\mathrm{Re}\,P_{\mathbf{q}r}(\omega^2) = \mathfrak{P}\sum_{\substack{\mathbf{k},\rho,\rho'\\ \lambda,\lambda'}}|\hat{\alpha}_{\rho\lambda}(\mathbf{k},\mathbf{q})|^2[(\omega-\omega_{\rho\lambda}(\mathbf{k})-\omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q}))^{-1}+(-\omega-\omega_{\rho\lambda}(\mathbf{k})-\omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q}))^{-1}] \tag{50}$$

and

$$\gamma_{\mathbf{q}r}(\omega^2) = \pi\sum_{\substack{\mathbf{k},\rho,\rho'\\ \lambda,\lambda'}}|\hat{\alpha}_{\rho\lambda}(\mathbf{k},\mathbf{q})|^2[\delta(\omega-\omega_{\rho\lambda}(\mathbf{k})-\omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q}))+\delta(-\omega-\omega_{\rho\lambda}(\mathbf{k})-\omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q}))], \tag{51}$$

with $\beta = (K_B T)^{-1}$, K_B is Boltzmann's constant, T the absolute temperature, and $\mathfrak{P}$ indicates that the principal value must be taken. The function (49) shows a resonance shape centered around the real frequencies that are determined by the dispersion relation

$$\omega^2-\omega_r^2(\mathbf{q})-\omega_r(\mathbf{q})\,\mathrm{Re}\,P_{\mathbf{q}r}(\omega^2)=0. \tag{52}$$

If the damping $\gamma_{\mathbf{q}r}(\omega^2)$ is very small, the frequencies of the dressed phonons $\tilde{\omega}_r(\mathbf{q})$ are given by the roots of the equation

$$\tilde{\omega}_r^2(\mathbf{q})-\omega_r^2(\mathbf{q})-\omega_r(\mathbf{q})\,\mathrm{Re}\,P_{\mathbf{q}r}(\tilde{\omega}_r^2(\mathbf{q}))=0. \tag{53}$$

Then Eq. (49) shows a Lorentzian shape around the frequency $\tilde{\omega}_r(q)$, provided that $\gamma_{\mathbf{q}r}(\tilde{\omega}_r^2(\mathbf{q})) \ll \omega_r(\mathbf{q})$ and $\partial\gamma_{\mathbf{q}r}(\omega^2)/\partial\omega^2 \ll 1$, and has a width of the order

$$\frac{\gamma_{\mathbf{q}r}(\tilde{\omega}_r^2(\mathbf{q}))}{2\omega_r(\mathbf{q})}\left|1-\omega_r(\mathbf{q})\frac{\partial\,\mathrm{Re}\,P_{\mathbf{q}r}(\omega^2)}{\partial\omega^2}\right|^{-1}_{\omega=\tilde{\omega}r(\mathbf{q})}. \tag{54}$$

From the expression (51), it is easily seen that the function $\gamma_{\mathbf{q}r}(\tilde{\omega}_r^2(\mathbf{q}))$ is different from zero unless

$$\tilde{\omega}_r(\mathbf{q}) = \pm[\omega_{\rho\lambda}(\mathbf{k})+\omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q})]. \tag{55}$$

In the optical region, the condition (55) is not likely to be satisfied, for it means that the dressed phonon energy $\tilde{\omega}_r(\mathbf{q})$ must be equal to the sum of the energies of two polaritons with wave vectors $\mathbf{k}$ and $\mathbf{k}-\mathbf{q}$, respectively. Physically, the condition (55) indicates the simultaneous creation or annihilation of two polaritons through the dressed phonon $(\mathbf{q}, r)$. This kind of process could occur only in the infrared region of frequencies where the incoming radiation is of the order of magnitude of the phonon energies. Then, of course, allowance for anharmonicity will lead to further broadening of these peaks with the possibility of the appearance of new peaks (overtones) at multiple or combination frequencies. Therefore, in the optical region, $\gamma_{\mathbf{q}r}(\omega^2)$ goes to zero and then Eq. (49) has a delta-function distribution, i.e.,

$$J_{\mathbf{q}r}(\omega) \approx 2\omega_r(\mathbf{q})(e^{\beta\omega}-1)^{-1}\times\delta(\omega^2-\omega_r^2(\mathbf{q})-\omega_r(\mathbf{q})P_{\mathbf{q}r}(\omega^2)), \quad \text{for}\quad \gamma_{\mathbf{q}r}(\omega^2)\to 0. \tag{56}$$

Equation (56) indicates that the phonon spectrum will be strongly peaked at the frequencies $\tilde{\omega}_r^2(\mathbf{q})$ given now by the roots of the equation

$$\omega^2-\omega_r^2(\mathbf{q})-\omega_r(\mathbf{q})P_{\mathbf{q}r}(\omega^2)=0. \tag{57}$$

Using the relation

$$\langle\tilde{\beta}_r^+(\mathbf{q})\tilde{\beta}_r(\mathbf{q})\rangle = \int_{-\infty}^{+\infty}d\omega J_{\mathbf{q}r}(\omega)$$

and (56), we find that integration over ω leads to

$$\langle\tilde{\beta}_r^+(\mathbf{q})\tilde{\beta}_r(\mathbf{q})\rangle = [\omega_r(\mathbf{q})/\tilde{\omega}_r(\mathbf{q})]\coth\tfrac{1}{2}\beta\tilde{\omega}_r(\mathbf{q}), \tag{58}$$

and the phonon distribution function $N_r(\mathbf{q})$ is equal to

$$N_r(\mathbf{q}) = \frac{1}{2}\left[\frac{\omega_r(\mathbf{q})}{\tilde{\omega}_r(\mathbf{q})}\coth\tfrac{1}{2}\beta\tilde{\omega}_r(\mathbf{q})-1\right], \tag{59}$$

with

$$(\omega_r(\mathbf{q})/\tilde{\omega}_r(\mathbf{q})) = [1+P_{\mathbf{q}r}(\tilde{\omega}_r^2(\mathbf{q}))/\omega_r(\mathbf{q})]^{-\frac{1}{2}}\times\left|1-\omega_r(\mathbf{q})\frac{\partial}{\partial\omega^2}P_{\mathbf{q}r}(\omega^2)\right|^{-1}_{\omega=\tilde{\omega}_r(\mathbf{q})}. \tag{60}$$

From (59) we see that, even at zero temperature ($\beta\to\infty$), the phonon distribution function is different from zero, i.e.,

$$N_r^{(0)}(\mathbf{q}) = \frac{1}{2}\left[\frac{\omega_r(q)}{\tilde{\omega}_r(q)}-1\right], \tag{61}$$

and could be interpreted as the zero-point fluctuations in the phonon field caused by the polariton–phonon interaction. Using (59), we derive the average energy due to the phonon field

$$\langle\mathcal{H}_{\text{phonon}}\rangle = \sum_{\mathbf{q},r}\omega_r(\mathbf{q})[N_r(\mathbf{q})+\tfrac{1}{2}] = \frac{1}{2}\sum_{\mathbf{q},r}\frac{\omega_r(q)}{\tilde{\omega}_r(\mathbf{q})}\coth\tfrac{1}{2}\beta\tilde{\omega}_r(\mathbf{q}) \tag{62}$$

or

$$\langle\mathcal{H}_{\text{phonon}}\rangle = \tfrac{1}{2}\sum_{\mathbf{q},r}\omega_r(\mathbf{q})[1+P_{\mathbf{q}r}(\tilde{\omega}_r^2(\mathbf{q}))/\omega_r(\mathbf{q})]^{-\frac{1}{2}}\times\coth\tfrac{1}{2}\beta\tilde{\omega}_r(\mathbf{q})\times\left|1-\omega_r(\mathbf{q})\frac{\partial}{\partial\omega^2}P_{\mathbf{q}r}(\omega^2)\right|^{-1}_{\omega=\tilde{\omega}_r(\mathbf{q})}. \tag{63a}$$

In the absence of interaction $P_{\mathbf{q}r}(\tilde{\omega}_r(\mathbf{q})) = 0$, Eq. (63a) is reduced to the usual expression for the average energy in the harmonic approximation

$$\tfrac{1}{2}\sum_{\mathbf{q},r}\omega_r(\mathbf{q})\coth\tfrac{1}{2}\beta\omega_r(\mathbf{q}).$$

The constant factor

$$[1 + P_{\mathbf{q}r}(\tilde{\omega}_r^2(\mathbf{q}))/\omega_r(\mathbf{q})]^{-\frac{1}{2}}\left|1 - \omega_r(\mathbf{q})\frac{\partial}{\partial\omega^2}P_{\mathbf{q}r}(\omega^2)\right|^{-1}_{\omega=\tilde{\omega}_r(\mathbf{q})}$$

that appears in the expression (63a) could be interpreted as due to the polarization of the medium resulting from the polariton–phonon interaction. The temperature-dependent part, $\coth\frac{1}{2}\beta\tilde{\omega}_r(\mathbf{q})$, depends now on the dressed phonon energy $\tilde{\omega}_r(\mathbf{q})$. The modification to the average phonon energy that appears in the expression (63a) depends mainly on the value of the ratio $P_{\mathbf{q}r}(\tilde{\omega}_r^2(\mathbf{q}))/\omega_r(\mathbf{q})$ and could have some importance at very low temperatures for crystals having small anharmonicities. In the limiting case where retardation is not important, $\omega_{\rho\lambda}(\mathbf{k}) = E_\mu(\mathbf{k})$, the polarization operator becomes

$$P_{\mathbf{k}\mu}(\omega^2) = \sum_{\mathbf{k},\mu,\mu'}|\hat{\alpha}_\mu(\mathbf{k},\mathbf{q})|^2 \times[(\omega - E_\mu(\mathbf{k}) - E_{\mu'}(\mathbf{k}-\mathbf{q}))^{-1} + (-\omega - E_\mu(\mathbf{k}) - E_{\mu'}(\mathbf{k}-\mathbf{q}))^{-1}], \quad (63b)$$

with $\hat{\alpha}_\mu(\mathbf{k},\mathbf{q})$ given by (46d). The polarization operator now describes the physical process where two bare excitons are created or annihilated through the exchange of a phonon $(\mathbf{q}, r)$.

V. ABSORPTION COEFFICIENT

The absorption coefficient $\tilde{\alpha}(\mathbf{k},\omega)$, the index of refraction of the medium $\eta(\mathbf{k},\omega)$, and the real part of the electrical conductivity $\mathrm{Re}\,\sigma(\mathbf{k},\omega)$ are related by the well-known formula

$$\tilde{\alpha}(\mathbf{k},\omega) = \frac{4\pi}{c\eta(\mathbf{k},\omega)}\mathrm{Re}\,\sigma(\mathbf{k},\omega) \quad (64)$$

derived from Maxwell's equations for an isotropic cubic crystal. Using the relation

$$\mathrm{Re}\,\sigma(\mathbf{k},\omega) = \frac{\omega}{4\pi}\mathrm{Im}\,\epsilon(\mathbf{k},\omega),$$

we see that the expression (64) becomes

$$\tilde{\alpha}(\mathbf{k},\omega) = \frac{\omega}{c\eta(\mathbf{k},\omega)}\mathrm{Im}\,\epsilon(\mathbf{k},\omega). \quad (65)$$

In the optical region where the relations

$$\mathrm{Re}\,\epsilon(\mathbf{k},\omega) > 0 \quad \text{and} \quad \omega \gg \frac{4\pi}{\mathrm{Re}\,\epsilon(\mathbf{k},\omega)}\mathrm{Re}\,\sigma(\mathbf{k},\omega) \quad (66)$$

hold,[15] we may replace in (65) $\eta(\mathbf{k},\omega)$ by $\mathrm{Re}\,\epsilon^{\frac{1}{2}}(\mathbf{k},\omega)$; that is, in the optical region of frequencies, $\eta(\mathbf{k},\omega)$ is a slowly varying function of ω and may be regarded as constant.

A. Polariton–Phonon Spectrum

Substitution of the imaginary part of (40) into (65) leads to the following expression for the absorption coefficient:

$$\tilde{\alpha}_\lambda(\mathbf{k},\omega) = (n_0 - n_\mu)\frac{\omega_p^2}{c\omega\eta(\mathbf{k},\omega)}\sum_\rho|A_{\rho,\lambda}(\mathbf{k},\omega)|^2 S_\rho(\mathbf{k},\omega) + \frac{(n_0 - n_{\mu'})\omega_p^2}{Nc\omega\eta(\mathbf{k},\omega)}\sum_{\mathbf{q},r,\rho',\lambda'}\hat{f}_{0\mu'}(\mathbf{k}-\mathbf{q},\lambda')\omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q})\,|Y_r(\mathbf{q})|^2\,\mathrm{Im}\,D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\rho',\lambda';\omega), \quad (67)$$

where the shape function $S_\rho(\mathbf{k},\omega)$ is equal to

$$S_\rho(\mathbf{k},\omega) = \frac{2\omega^2_{\rho\lambda}(\mathbf{k})\tilde{\gamma}_{\mathbf{k}\rho}(\omega) + [\omega^2 - \omega^2_{\rho\lambda}(\mathbf{k})]\hat{\gamma}_{\mathbf{k}\rho}(\omega)}{[\omega^2 - \omega^2_{\rho\lambda}(\mathbf{k}) - 2\omega^2_{\rho\lambda}(\mathbf{k})\tilde{Q}_{\mathbf{k}\rho}(\omega)]^2 + [2\omega_{\rho\lambda}(\mathbf{k})\tilde{\gamma}_{\mathbf{k}\rho}(\omega)]^2} \quad (68)$$

and

$$\mathrm{Im}\,D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\rho',\lambda';\omega) = \pi\{[1 + N_r(\mathbf{q})]\delta(\omega - \omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q}) - \omega_r(\mathbf{q})) + N_r(\mathbf{q})\delta(\omega - \omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q}) + \omega_r(\mathbf{q}))\}. \quad (69)$$

In (68) we have made use of the following notation:

$$Q_{\mathbf{k}\rho}(\omega) = \sum_{\mathbf{q},r,\rho',\lambda'}|\alpha_{\rho\lambda}(\mathbf{k},\mathbf{q}) + \hat{\alpha}_{\rho\lambda}(\mathbf{k},\mathbf{q})|^2\,\mathrm{Re}\,D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\rho',\lambda';\omega), \quad (70a)$$

$$\hat{Q}_{\mathbf{k}\rho}(\omega) = \sum_{\mathbf{q},r,\rho',\lambda'}|\alpha_{\rho\lambda}(\mathbf{k},\mathbf{q}) - \hat{\alpha}_{\rho\lambda}(\mathbf{k},\mathbf{q})|^2\,\mathrm{Re}\,D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\rho',\lambda';\omega), \quad (70b)$$

[15] V. L. Bonch-Bruyevich, *The Electronic Theory of Heavily Doped Semiconductors* (American Elsevier Publ. Co., Inc., New York, 1966), p. 98.

$$\gamma_{\mathbf{k}\rho}(\omega) \equiv \operatorname{Im} \Omega^{(1)}_{\mathbf{k}\rho}(\omega) = \sum_{\mathbf{q},r,\rho',\lambda'} |\alpha_{\rho\lambda}(\mathbf{k},\mathbf{q}) + \hat{\alpha}_{\rho\lambda}(\mathbf{k},\mathbf{q})|^2 \operatorname{Im} D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\rho',\lambda';\omega), \tag{70c}$$

$$\hat{\gamma}_{\mathbf{k}\rho}(\omega) \equiv \operatorname{Im} \Omega^{(2)}_{\mathbf{k}\rho}(\omega) = \sum_{\mathbf{q},r,\rho',\lambda'} |\alpha_{\rho\lambda}(\mathbf{k},\mathbf{q}) - \hat{\alpha}_{\rho\lambda}(\mathbf{k},\mathbf{q})|^2 \operatorname{Im} D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\rho',\lambda';\omega), \tag{70d}$$

$$\tilde{\gamma}_{\mathbf{k}\rho}(\omega) \equiv \tfrac{1}{2}[\gamma_{\mathbf{k}\rho}(\omega) + \hat{\gamma}_{\mathbf{k}\rho}(\omega)] = \sum_{\mathbf{q},r,\rho',\lambda'} [|\alpha_{\rho\lambda}(\mathbf{k},\mathbf{q})|^2 + |\hat{\alpha}_{\rho\lambda}(\mathbf{k},\mathbf{q})|^2] \operatorname{Im} D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\rho',\lambda';\omega), \tag{70e}$$

$$\operatorname{Re} D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\rho',\lambda';\omega) \equiv P\left[\frac{1+N_r(\mathbf{q})}{\omega-\omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q})-\omega_r(\mathbf{q})} + \frac{N_r(\mathbf{q})}{\omega-\omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q})+\omega_r(\mathbf{q})}\right], \tag{70f}$$

$$\tilde{Q}_{\mathbf{k}\rho}(\omega) \equiv [Q_{\mathbf{k}\rho}(\omega) + \hat{Q}_{\mathbf{k}\rho}(\omega)]/2\omega_{\rho\lambda}(\mathbf{k}) + [Q_{\mathbf{k}\rho}(\omega)\hat{Q}_{\mathbf{k}\rho}(\omega) - \gamma_{\mathbf{k}\rho}(\omega)\hat{\gamma}_{\mathbf{k}\rho}(\omega)]/2\omega^2_{\rho\lambda}(\mathbf{k}). \tag{70g}$$

From (70a)–(70e) we note that the ω-dependent part of the functions $Q_{\mathbf{k}\rho}(\omega)$, $\hat{Q}_{\mathbf{k}\rho}(\omega)$, and $\gamma_{\mathbf{k}\rho}(\omega)$, $\hat{\gamma}_{\mathbf{k}\rho}(\omega)$ is entirely determined by the real and imaginary parts of the unperturbed polariton–phonon Green's function $D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\rho',\lambda';\omega)$, respectively. Inspection of the shape function (68) shows that each absorption peak is of asymmetric Lorentzian shape even if we neglect the ω dependence of the functions $\gamma_{\mathbf{k}\rho}(\omega)$, $\hat{\gamma}_{\mathbf{k}\rho}(\omega)$, and $\tilde{Q}_{\mathbf{k}\rho}(\omega)$, respectively. The asymmetry arises from the function $\hat{\gamma}_{\mathbf{k}\rho}(\omega)$ given by expression (70d); considering expressions (25) and (26) for $\alpha_{\rho\lambda}(\mathbf{k},\mathbf{q})$ and $\hat{\alpha}_{\rho\lambda}(\mathbf{k},\mathbf{q})$, respectively, we see that the coupling constant of the function $\hat{\gamma}_{\mathbf{k}\rho}(\omega)$ depends entirely on the couplings between both the dressed electronic subsystem in the excited state and the intermolecular interactions with the acoustic branches of the phonon field. The second term on the r.h.s. of (67) has a delta-function distribution given by the $\operatorname{Im} D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\rho',\lambda';\omega)$, and is peaked at the zeros of the arguments of the two delta functions that appear in expression (69) corresponding to the process of emission or absorption of a phonon, respectively, with frequency $\omega_r(\mathbf{q})$ for a given value of $\mathbf{k}$.

To investigate the frequency dependence of the function $\omega\tilde{\alpha}_\lambda(\mathbf{k},\omega)$, we shall assume that the real and imaginary parts of the function $D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\rho',\lambda';\omega)$ vary slowly with ω so that the conditions

$$\frac{\partial}{\partial\omega}\gamma_{\mathbf{k}\rho}(\omega) \ll 1, \quad \frac{\partial}{\partial\omega}\hat{\gamma}_{\mathbf{k}\rho}(\omega) \ll 1,$$

$$\frac{\partial}{\partial\omega}Q(\omega) \ll 1, \quad \frac{\partial}{\partial\omega}\hat{Q}_{\mathbf{k}\rho}(\omega) \ll 1,$$

and

$$\frac{\partial}{\partial\omega}\operatorname{Im} D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\rho',\lambda';\omega) \ll 1$$

are satisfied within the region of frequencies

$$|\omega^2 - \omega^2_{\rho\lambda}(\mathbf{k})| \gtrsim 2\omega^2_{\rho\lambda}(\mathbf{k})\tilde{\gamma}_{\mathbf{k}\rho}(\omega)/\hat{\gamma}_{\mathbf{k}\rho}(\omega).$$

Having these conditions in mind, it is sufficient to examine the solutions of the equation

$$\frac{\partial}{\partial\omega}S_\rho(\mathbf{k},\omega) = 0 \tag{71}$$

that result in the following expression for the frequencies $\tilde{\omega}^2_{\rho\lambda}(\mathbf{k})$:

$$|\tilde{\omega}^2_{\rho\lambda}(\mathbf{k}) - \omega^2_{\rho\lambda}(\mathbf{k})| = 2\omega^2_{\rho\lambda}(\mathbf{k})\tilde{\gamma}_{\mathbf{k}\rho}/\hat{\gamma}_{\mathbf{k}\rho}\{-1 \pm [(1+\tilde{Q}_{\mathbf{k}\rho}\hat{\gamma}_{\mathbf{k}\rho}/\tilde{\gamma}_{\mathbf{k}\rho})^2 + (\hat{\gamma}_{\mathbf{k}\rho}/\omega_{\rho\lambda}(\mathbf{k}))^2]^{\frac{1}{2}}\}, \tag{72}$$

where the notation $\tilde{\gamma}_{\mathbf{k}\rho} \equiv \tilde{\gamma}_{\mathbf{k}\rho}(\tilde{\omega}_{\rho\lambda}(\mathbf{k}))$,

$$\hat{\gamma}_{\mathbf{k}\rho} \equiv \hat{\gamma}_{\mathbf{k}\rho}(\tilde{\omega}_{\rho\lambda}(\mathbf{k})),$$

and $\tilde{Q}_{\mathbf{k}\rho} \equiv \tilde{Q}_{\mathbf{k}\rho}(\tilde{\omega}_{\rho\lambda}(\mathbf{k}))$ has been introduced. Thus the function $S_\rho(\mathbf{k},\omega)$ has steep maxima at some frequencies $\omega = \tilde{\omega}_{\rho\lambda}(\mathbf{k})$ that are determined from expression (72) provided that $\tilde{\omega}_{\rho\lambda}(\mathbf{k})$ also satisfies the equation

$$\tilde{\omega}_{\rho\lambda}(\mathbf{k}) - \omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q}) \pm \omega_r(\mathbf{q}) = 0, \tag{73}$$

a condition that is required for $\gamma_{\mathbf{k}\rho} \neq 0$ and $\hat{\gamma}_{\mathbf{k}\rho} \neq 0$. Then, if the conditions $\gamma_{\mathbf{k}\rho} \ll \omega_{\rho\lambda}(\mathbf{k})$, $\hat{\gamma}_{\mathbf{k}\rho} \ll \omega_{\rho\lambda}(\mathbf{k})$, and $\tilde{Q}_{\mathbf{k}\rho} \ll 1$ hold, the solutions of Eq. (72), $\tilde{\omega}^2_{\rho\lambda}(\mathbf{k})$, correspond to the resonance frequencies of the polariton–phonon spectrum and the r.h.s. of (72) gives the energy shift with respect to the unperturbed energy $\omega^2_{\rho\lambda}(\mathbf{k})$ and is temperature dependent. At the frequencies $\omega = \tilde{\omega}_{\rho\lambda}(\mathbf{k})$, the function $|S_\rho(\mathbf{k},\tilde{\omega}_{\rho\lambda}(\mathbf{k}))|$ is equal to

$$|S_\rho(\mathbf{k},\tilde{\omega}_{\rho\lambda}(\mathbf{k}))| = (4\tilde{\gamma}_{\mathbf{k}\rho})^{-1}\{(1+\tilde{Q}_{\mathbf{k}\rho}\hat{\gamma}_{\mathbf{k}\rho}/\tilde{\gamma}_{\mathbf{k}\rho}) \pm [(1+\tilde{Q}_{\mathbf{k}\rho}\hat{\gamma}_{\mathbf{k}\rho}/\tilde{\gamma}_{\mathbf{k}\rho})^2 + \hat{\gamma}^2_{\mathbf{k}\rho}/\omega^2_{\rho\lambda}(\mathbf{k})]^{\frac{1}{2}}\}, \tag{74}$$

which gives the height of the resonance band while the corresponding width $|\Gamma_{\mathbf{k}\rho}(\tilde{\omega}_{\rho\lambda}(\mathbf{k}))|$ in energy units is given by

$$|\Gamma_{\mathbf{k}\rho}(\tilde{\omega}_{\rho\lambda}(\mathbf{k}))| = 4\tilde{\gamma}_{\mathbf{k}\rho}\{(1+\tilde{Q}_{\mathbf{k}\rho}\hat{\gamma}/\tilde{\gamma}_{\mathbf{k}\rho}) \pm [(1+\tilde{Q}_{\mathbf{k}\rho}\hat{\gamma}_{\mathbf{k}\rho}/\tilde{\gamma}_{\mathbf{k}\rho})^2 + \hat{\gamma}^2_{\mathbf{k}\rho}/\omega^2_{\mathbf{k}\rho}(\mathbf{k})]^{\frac{1}{2}}\}^{-1}. \tag{75}$$

At high temperatures $\beta\omega_r(\mathbf{q}) < 1$, the quantities $\tilde{\gamma}_{\mathbf{k}\rho}$, $\hat{\gamma}_{\mathbf{k}\rho}$, and $\tilde{Q}_{\mathbf{k}\rho}$ vary linearly with temperature, because we have restricted ourselves to a one-phonon process, and it is easily shown that the expressions for the height and the width are approximately proportional to to T^{-1} and T, respectively. This is due to the fact that the terms in (74) and (75) causing the asymmetry in the absorption line appear as the ratio of quantities such as $\hat{\gamma}_{\mathbf{k}\rho}/\tilde{\gamma}_{\mathbf{k}\rho}$ and $\tilde{Q}_{\mathbf{k}\rho}/\tilde{\gamma}_{\mathbf{k}\rho}$, and they are practically independent of temperature. For $\hat{\gamma}_{\mathbf{k}\rho} = 0$, both expressions (74) and (75) for the height and the width of

the absorption band are reduced to that of $\frac{1}{2}\gamma_{\mathbf{k}\rho}^{-1}(\tilde{\omega}_{\rho\lambda}(\mathbf{k}))$ and $2\gamma_{\mathbf{k}\rho}(\tilde{\omega}_{\rho\lambda}(\mathbf{k}))$, respectively, corresponding to a symmetric Lorentzian line with $\tilde{\omega}_{\rho\lambda}(\mathbf{k})$ now determined from the roots of the equation

$$\tilde{\omega}_{\rho\lambda}^2(\mathbf{k}) - \omega_{\rho\lambda}^2(\mathbf{k}) - 2\omega_{\rho\lambda}^2(\mathbf{k})\tilde{Q}_{\mathbf{k}\rho}(\tilde{\omega}_{\rho\lambda}(\mathbf{k})) = 0, \quad (76)$$

provided that $(\partial/\partial\omega)\gamma_{\mathbf{k}\rho}(\omega) \ll 1$, and $(\partial/\partial\omega)Q_{\mathbf{k}\rho}(\omega) \ll 1$ in the neighborhood of frequencies $\omega \approx \tilde{\omega}_{\rho\lambda}(\mathbf{k})$ that satisfy Eq. (73).

Thus the expressions (74) and (75) arise from contributions resulting from the superposition of symmetric and asymmetric Lorentzian lines; and since $\hat{\gamma}_{\mathbf{k}\rho}/\tilde{\gamma}_{\mathbf{k}\rho} < 1$, the deviation from the symmetric Lorentzian shape depends mainly on the values of the quantities $\hat{\gamma}_{\mathbf{k}\rho}^2/\omega_{\rho\lambda}^2(\mathbf{k})$ and $\tilde{Q}_{\mathbf{k}\rho}$. In the case of exact resonance, one may replace $Q_{\mathbf{k}\rho} \approx 0$ in the expressions (68), (72), (74), and (75); then the asymmetry in the line shape will depend entirely on the value of the quantity $\hat{\gamma}_{\mathbf{k}\rho}^2/\omega_{\rho\lambda}^2(\mathbf{k})$. Because of the condition (73), the second term on the rhs of (67) will be peaked at $\omega = \tilde{\omega}_{\rho\lambda}(\mathbf{k})$ and will thus give a constant contribution to the expression for the absorption coefficient. Therefore, in the case of exact resonance at $\omega = \tilde{\omega}_{\rho\lambda}(\mathbf{k})$, we have

$$\begin{aligned}
&[\omega\tilde{\alpha}_{\rho\lambda}(\mathbf{k},\omega)]_{\omega=\tilde{\omega}_{\rho\lambda}(\mathbf{k})} \\
&= \frac{(n_0 - n_\mu)\omega_p^2}{c\eta(\mathbf{k},\omega_{\rho\lambda}(\mathbf{k}))} \sum_\rho \hat{f}_{0\mu}(\mathbf{k},\lambda)\omega_{\rho\lambda}(\mathbf{k})|S_\rho(\mathbf{k},\tilde{\omega}_{\rho\lambda}(\mathbf{k}))| \\
&\quad + \frac{\omega_p^2(n_0 - n_{\mu'})}{Nc\eta(\mathbf{k},\omega_{\rho\lambda}(\mathbf{k}))} \sum_{\mathbf{q},r,\rho',\lambda'} \hat{f}_{0\mu'}(\mathbf{k}-\mathbf{q},\lambda')\omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q}) \\
&\quad \times |Y_r(\mathbf{q})|^2 \operatorname{Im} D_{\mathbf{q}r}^{(0)}(\mathbf{k}-\mathbf{q},\rho',\lambda';\tilde{\omega}_{\rho\lambda}(\mathbf{k})) \\
&\quad \times \left[1 + \sum_\rho |S_\rho(\mathbf{k},\tilde{\omega}_{\rho\lambda}(\mathbf{k}))|\,\gamma_{\mathbf{k}\rho}\right], \quad (77)
\end{aligned}$$

and $|S_\rho(\mathbf{k},\tilde{\omega}_{\rho\lambda}(\mathbf{k}))|$ is given by (74) with $\tilde{Q}_{\mathbf{k}\rho} \approx 0$. At high temperatures $\beta\omega_r(\mathbf{q}) < 1$, the first term on the rhs of (77) varies like T^{-1}, while the remaining two terms are proportional to the temperature T. Therefore, at high temperatures the last two terms on the rhs of (77) will dominate the first term. Of course, at very high temperatures, where anharmonicities become substantial, then consideration of multiphonon effects might result in more complicated temperature dependence of the absorption coefficient. At low temperatures $\beta\omega_r(\mathbf{q}) > 1$, there will be a competition among the terms that appear on the rhs of (77) but for very small values of the damping constant $\tilde{\gamma}_{\mathbf{k}\rho}$, $\tilde{\gamma}_{\mathbf{k}\rho} \ll \omega_{\rho\lambda}(\mathbf{k})$, it seems that the first term will be predominant. Our discussion so far has been based on the assumption that the function $D_{\mathbf{q}r}^{(0)}(\mathbf{k}-\mathbf{q},\rho',\lambda';\omega)$ varies slowly with ω in the neighborhood of $\omega \approx \tilde{\omega}_{\rho\lambda}(\mathbf{k})$ which appears to be the case for most physical problems where the nature of the real final states is such that $(\partial/\partial\omega)\gamma_{\mathbf{k}\rho}(\omega) \ll 1$ and $(\partial/\partial\omega)\hat{\gamma}_{\mathbf{k}\rho}(\omega) \ll 1$. If the variation of $\operatorname{Im} D_{\mathbf{q}r}^{(0)}(\mathbf{k}-\mathbf{q},\rho',\lambda';\omega)$ with respect to ω is taken into consideration, further asymmetries will appear in the line shape with the possibility of the appearance of new peaks at frequencies corresponding to the so-called combination bands determined by the relation

$$(\partial/\partial\omega)\operatorname{Im} D_{\mathbf{q}r}^{(0)}(\mathbf{k}-\mathbf{q},\rho',\lambda';\omega) = 0.$$

In the vicinity of the edges of the polariton absorption band where $\omega^2 - \omega_{\rho\lambda}^2(\mathbf{k}) \gg 2\omega_{\rho\lambda}^2(\mathbf{k})\tilde{Q}_{\mathbf{k}\rho}(\omega)$, the absorption coefficient is described by the expression

$$\begin{aligned}
\tilde{\alpha}_\lambda(\mathbf{k},\omega) &\approx (n_0 - n_\mu)\frac{\omega_p^2}{c\omega\eta(\mathbf{k},\omega)} \sum_\mu \hat{f}_{0\mu}(\mathbf{k},\lambda)\omega_{\rho\lambda}(\mathbf{k})\left\{\frac{2\omega_{\rho\lambda}^2(\mathbf{k})\tilde{\gamma}_{\mathbf{k}\rho}(\omega)}{[\omega^2 - \omega_{\rho\lambda}^2(\mathbf{k})]^2} + \frac{\hat{\gamma}_{\mathbf{k}\rho}(\omega)}{[\omega^2 - \omega_{\rho\lambda}^2(\mathbf{k})]}\right\} \\
&\quad + \frac{(n_0 - n_{\mu'})\omega_p^2}{Nc\omega\eta(\mathbf{k},\omega)} \sum_{\mathbf{q},r,\rho',\lambda'} \hat{f}_{0\mu'}(\mathbf{k}-\mathbf{q},\lambda')\omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q})\,|Y_r(\mathbf{q})|^2 \operatorname{Im} D_{\mathbf{q}r}^{(0)}(\mathbf{k}-\mathbf{q},\rho',\lambda';\omega) \\
&\quad \times \left\{1 + \left[\frac{2\omega_{\rho\lambda}^2(\mathbf{k})\tilde{\gamma}_{\mathbf{k}\rho}(\omega)}{(\omega^2 - \omega_{\rho\lambda}^2(\mathbf{k}))^2} + \frac{\hat{\gamma}_{\mathbf{k}\rho}(\omega)}{\omega^2 - \omega_{\rho\lambda}^2(\mathbf{k})}\right]\gamma_{\mathbf{k}\rho}(\omega)\right\}. \quad (78)
\end{aligned}$$

At high temperatures $\beta\omega_r(\mathbf{q}) < 1$, the first three terms on the rhs of (78) vary linearly with T while the last term is proportional to T^3. We also notice that in the expression (78) the damping constants $\tilde{\gamma}_{\mathbf{k}\rho}(\omega)$ and $\hat{\gamma}_{\mathbf{k}\rho}(\omega)$ are multiplied by $[\omega^2 - \omega_{\rho\lambda}^2(\mathbf{k})]^{-2}$ and $[\omega^2 - \omega_{\rho\lambda}^2(\mathbf{k})]^{-1}$, respectively, which means that they behave differently with respect to $[\omega^2 - \omega_{\rho\lambda}^2(\mathbf{k})]$. The first term on the rhs of (78) can be reduced to that derived by Agranovich and Konobeev,[4] if one considers the particular case studied by these authors.

In the limiting case where the condition (73) is not satisfied, that is, when the arguments of the delta functions that appear in (69) are different than zero for $\omega = \tilde{\omega}_{\rho\lambda}(\mathbf{k})$, then the damping constants $\gamma_{\mathbf{k}\rho}(\tilde{\omega}_{\rho\lambda}(\mathbf{k}))$

and $\hat{\gamma}_{\mathbf{k}\rho}(\tilde{\omega}_{\rho\lambda}(\mathbf{k}))$ go to zero and the expression for the absorption coefficient is equal to

$$\tilde{\alpha}_{\lambda}(\mathbf{k},\omega) \approx \pi(n_0 - n_{\mu})\frac{\omega_p^2}{c\omega\eta(\mathbf{k},\omega)}\sum_{\rho}|A_{\rho\lambda}(\mathbf{k},\omega)|^2\,\omega_{\rho\lambda}(\mathbf{k})\delta(\omega^2 - \Omega^{(1)}_{\mathbf{k}\rho}(\omega)\Omega^{(2)}_{\mathbf{k}\rho}(\omega))$$
$$+\frac{(n_0 - n_{\mu'})\omega_p^2}{Nc\omega\eta(\mathbf{k},\omega)}\sum_{\mathbf{q},r,\rho',\lambda'}\hat{f}_{0\mu'}(\mathbf{k}-\mathbf{q},\lambda')\omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q})\,|Y_r(\mathbf{q})|^2\,\mathrm{Im}\,D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\rho',\lambda';\omega), \quad (79)$$

for

$$\gamma_{\mathbf{k}\rho}(\tilde{\omega}_{\rho\lambda}(\mathbf{k}))\to 0,\quad \hat{\gamma}_{\mathbf{k}\rho}(\tilde{\omega}_{\rho\lambda}(\mathbf{k}))\to 0 \quad\text{and}\quad A_{\rho\lambda}(\mathbf{k},\omega)$$

is given by the expression (41), which is now real. The expression (79) corresponds to the transparent region of frequencies of the crystal and has a delta-function distribution. The first term on the rhs of (79) is different than zero for values of ω given by the real roots of the equation

$$\omega^2 - \Omega^{(1)}_{\mathbf{k}\rho}(\omega)\Omega^{(2)}_{\mathbf{k}\rho}(\omega) = 0, \quad (80)$$

with $\Omega^{(1)}_{\mathbf{k}\rho}(\omega)$ and $\Omega^{(2)}_{\mathbf{k}\rho}(\omega)$ given now by the real expressions (34) and (35), respectively, while the second term obeys the dispersion relation

$$\omega - \omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q}) \pm \omega_r(\mathbf{q}) = 0. \quad (81)$$

It is easily seen that for a given value of $\mathbf{k}$ the two terms on the rhs of (79) are peaked at different frequencies and their energy spectrum is determined by the dispersion relations (80) and (81), respectively. The solution of Eq. (80) that corresponds to the zero-phonon–polariton spectrum is obtained by replacing $\Omega^{(1)}_{\mathbf{k}\rho}(\omega)$ and $\Omega^{(2)}_{\mathbf{k}\rho}(\omega)$ by $\Omega^{(1)}_{\mathbf{k}\rho}(\omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q}))$ and

$$\Omega^{(2)}_{\mathbf{k}\rho}(\omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q})),$$

respectively, with the result

$$\Omega^{(0)2}_{\rho\lambda}(\mathbf{k}) = \Omega^{(1)}_{\mathbf{k}\rho}(\omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q}))\Omega^{(2)}_{\mathbf{k}\rho}(\omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q})), \quad (82)$$

where

$$\Omega^{(1)}_{\mathbf{k}\rho}(\omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q})) \equiv \omega_{\rho\lambda}(\mathbf{k}) - \sum_{\mathbf{q},r,\rho',\lambda'}|\alpha_{\rho\lambda}(\mathbf{k},\mathbf{q}) + \hat{\alpha}_{\rho\lambda}(\mathbf{k},\mathbf{q})|^2/\omega_r(\mathbf{q}) \quad (83a)$$

and

$$\Omega^{(2)}_{\mathbf{k}\rho}(\omega_{\rho'\lambda'}(\mathbf{k}-\mathbf{q})) \equiv \omega_{\rho\lambda}(\mathbf{k}) - \sum_{\mathbf{k},r,\rho',\lambda'}|\alpha_{\rho\lambda}(\mathbf{k},\mathbf{q}) - \hat{\alpha}_{\rho\lambda}(\mathbf{k},\mathbf{q})|^2/\omega_r(\mathbf{q}). \quad (83b)$$

The expression (82) for $\Omega^{(0)}_{\rho\lambda}(\mathbf{k})$ is the renormalized energy of excitation for the zero-phonon–polariton spectrum, and the last terms on the rhs of (83a) and (83b) describe the scattering amplitudes in energy units for the corresponding process at zero temperature that cause a reduction of the unperturbed energy $\omega_{\rho\lambda}(\mathbf{k})$. The renormalized energies of elementary excitations $\omega = \pm\Omega_{\rho\lambda}(\mathbf{k})$, corresponding to the emission or absorption of a photon with energy $\omega_r(\mathbf{q})$, are obtained by solving Eq. (80) provided that the function $D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\rho',\lambda';\omega)$ varies slowly near the maximum frequency,

$$\Omega_{\rho\lambda}(\mathbf{k}) = [\Omega^{(1)}_{\mathbf{k}\rho}(\Omega_{\rho\lambda}(\mathbf{k}))\Omega^{(2)}_{\mathbf{k}\rho}(\Omega_{\rho\lambda}(\mathbf{k}))]^{\frac{1}{2}}. \quad (84)$$

The spectrum determined by the solutions of (84) corresponds to that of the coupled polariton–phonon system and the expressions for $\Omega^{(1)}_{\mathbf{k}\rho}(Q_{\rho\lambda}(\mathbf{k})) - \omega_{\rho\lambda}(\mathbf{k})$ and $\Omega^{(2)}_{\mathbf{k}\rho}(\Omega_{\rho\lambda}(\mathbf{k})) - \omega_{\rho\lambda}(\mathbf{k})$, i.e.,

$$\Omega^{(1)}_{\mathbf{k}\rho}(\Omega_{\rho\lambda}(\mathbf{k})) - \omega_{\rho\lambda}(k) = \sum_{\mathbf{q},r,\rho',\lambda'}|\alpha_{\rho\lambda}(\mathbf{k},\mathbf{q}) + \hat{\alpha}_{\rho\lambda}(\mathbf{k},\mathbf{q})|^2 \times D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\rho',\lambda';\Omega_{\rho\lambda}(\mathbf{k})), \quad (85)$$

$$\Omega^{(2)}_{\mathbf{k}\rho}(\Omega_{\rho\lambda}(\mathbf{k})) - \omega_{\rho\lambda}(\mathbf{k}) = \sum_{\mathbf{q},r,\rho',\lambda'}|\alpha_{\rho\lambda}(\mathbf{k},\mathbf{q}) - \hat{\alpha}_{\rho\lambda}(\mathbf{k},\mathbf{q})|^2 \times D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\rho',\lambda';\Omega_{\rho\lambda}(\mathbf{k})) \quad (86)$$

describe the scattering amplitudes in energy units involving the emission or absorption of a phonon by the dressed particle $\Omega_{\rho\lambda}(\mathbf{k})$. Since $\alpha_{\rho\lambda}(\mathbf{k},\mathbf{q}) > \hat{\alpha}_{\rho\lambda}(\mathbf{k},\mathbf{q})$, the rhs of (85) is much larger than that of (86) which depends entirely on the coupling of the dressed electron subsystem with the phonon field. The scattering amplitude (85) or (86) is reduced to that derived by Ovander,[2,5] apart from small corrections in the expression $\alpha_{\rho\lambda}(\mathbf{k},\mathbf{q})$, if we set $\hat{\alpha}_{\rho\lambda}(\mathbf{k},\mathbf{q}) = 0$, replace the sum over q by an integral, and then integrate over $\mathbf{q}$ by taking the poles of the propagator $D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\rho',\lambda';\Omega_{\rho\lambda}(\mathbf{k}))$ with $\Omega_{\rho\lambda}(\mathbf{k}) \sim \omega_{\rho\lambda}(\mathbf{k})$ in the limit of zero temperature. The energy spectrum determined by Eq. (81) corresponds to frequencies ω where the polariton and phonon fields are not coupled. Therefore, it is possible that for a given value of $\mathbf{k}$ the two terms on the rhs of (79) will be peaked at different frequency regions that depends on whether or not the polariton and the phonon field are coupled, respectively.

B. Bare Exciton–Phonon Spectrum

The expression for the absorption coefficient corresponding to the bare exciton–phonon spectrum is obtained from that of the polariton–phonon by taking the limit when $\omega_{\rho\lambda}(\mathbf{k}) = E_\mu(\mathbf{k})$. Thus substitution of the imaginary part of (42) into (65) yields

$$\tilde{\alpha}_{ii}(\mathbf{k},\omega) = \frac{(n_0 - n_\mu)\omega_p^2}{c\omega\eta(\mathbf{k},\omega)} \sum_\mu |A_\mu(\mathbf{k},\omega)|^2 S_\mu(\mathbf{k},\omega)$$

$$+ \frac{(n_0 - n_{\mu'})\omega_p^2}{Nc\omega\eta(\mathbf{k},\omega)} \sum_{\mathbf{q},r,\mu'} f_{0\mu'}(\mathbf{k}-\mathbf{q}, i) E_{\mu'}(\mathbf{k}-\mathbf{q}) |Y_r(\mathbf{q})|^2 \operatorname{Im} D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\mu';\omega), \tag{87}$$

where

$$S_\mu(\mathbf{k},\omega) = \frac{2E_\mu^2(\mathbf{k})\tilde{\gamma}_{\mathbf{k}\mu}(\omega) + [\omega^2 - E_\mu^2(\mathbf{k})]\hat{\gamma}_{\mathbf{k}\mu}(\omega)}{[\omega^2 - E_\mu^2(\mathbf{k}) - 2E_\mu^2(\mathbf{k})\tilde{Q}_{\mathbf{k}\mu}(\omega)]^2 + [2E_\mu(\mathbf{k})\tilde{\gamma}_{\mathbf{k}\mu}(\omega)]^2} \tag{88}$$

and

$$\operatorname{Im} D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\mu';\omega) = \pi\{[1 + N_r(\mathbf{q})]\delta(\omega - E_{\mu'}(\mathbf{k}-\mathbf{q}) - \omega_r(\mathbf{q})) + N_r(\mathbf{q})\delta(\omega - E_{\mu'}(\mathbf{k}-\mathbf{q}) + \omega_r(\mathbf{q}))\}. \tag{89}$$

The quantities $\tilde{\gamma}_{\mathbf{k}\mu}(\omega)$, $\hat{\gamma}_{\mathbf{k}\mu}(\omega)$, and $\tilde{Q}_{\mathbf{k}\mu}(\omega)$ appearing in (88) are now defined as follows:

$$\tilde{\gamma}_{\mathbf{k}\mu}(\omega) \equiv \gamma_{\mathbf{k}\mu}(\omega) + \hat{\gamma}_{\mathbf{k}\mu}(\omega), \tag{90a}$$

$$\gamma_{\mathbf{k}\mu}(\omega) = \sum_{\mathbf{q},r,\mu'} |\alpha_\mu(\mathbf{k},\mathbf{q}) + \hat{\alpha}_\mu(\mathbf{k},\mathbf{q})|^2 \operatorname{Im} D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\mu';\omega), \tag{90b}$$

$$\hat{\gamma}_{\mathbf{k}\mu}(\omega) = \sum_{\mathbf{q},r,\mu'} |\alpha_\mu(\mathbf{k},\mathbf{q}) - \hat{\alpha}_\mu(\mathbf{k},\mathbf{q})|^2 \operatorname{Im} D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\mu';\omega), \tag{90c}$$

$$\tilde{Q}_{\mathbf{k}\mu}(\omega) \equiv [Q_{\mathbf{k}\mu}(\omega) + \hat{Q}_{\mathbf{k}\mu}(\omega)]/2E_\mu(\mathbf{k}) + [Q_{\mathbf{k}\mu}(\omega)\hat{Q}_{\mathbf{k}\mu}(\omega) - \gamma_{\mathbf{k}\mu}(\omega)\hat{\gamma}_{\mathbf{k}\mu}(\omega)]/2E_\mu^2(\mathbf{k}), \tag{90d}$$

$$Q_{\mathbf{k}\mu}(\omega) = \sum_{\mathbf{q},r,\mu'} |\alpha_\mu(\mathbf{k},\mathbf{q}) + \hat{\alpha}_\mu(\mathbf{k},\mathbf{q})|^2 \operatorname{Re} D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\mu';\omega), \tag{90e}$$

$$\tilde{Q}_{\mathbf{k}\mu}(\omega) = \sum_{\mathbf{q},r,\mu'} |\alpha_\mu(\mathbf{k},\mathbf{q}) - \hat{\alpha}_\mu(\mathbf{k},\mathbf{q})|^2 \operatorname{Re} D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\mu';\omega), \tag{90f}$$

and

$$\operatorname{Re} D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\mu';\omega) = P\left[\frac{1 + N_r(\mathbf{q})}{\omega - E_{\mu'}(\mathbf{k}-\mathbf{q}) - \omega_r(q)} + \frac{N_r(\mathbf{q})}{\omega - E_{\mu'}(\mathbf{k}-\mathbf{q}) + \omega_r(\mathbf{q})}\right], \tag{90g}$$

with $\alpha_\mu(\mathbf{k},\mathbf{q})$ and $\hat{\alpha}_\mu(\mathbf{k},\mathbf{q})$ given by the expressions (46c) and (46d), respectively. It is easily shown that the whole discussion and the expressions derived for the polariton–phonon spectrum are applicable to that for the bare exciton–phonon spectrum, provided that we make the replacement of the expressions for $D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q}, \rho', \lambda';\omega)$, $\gamma_{\mathbf{k}\rho}(\omega)$, $\hat{\gamma}_{\mathbf{k}\rho}(\omega)$, $Q_{\mathbf{k}\rho}(\omega)$, and $\hat{Q}_{\mathbf{k}\rho}(\omega)$ given by (69)–(70g) by the corresponding expressions

$$D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\mu';\omega),$$

$\gamma_{\mathbf{k}\mu}(\omega)$, $\hat{\gamma}_{\mathbf{k}\mu}(\omega)$, $Q_{\mathbf{k}\mu}(\omega)$ and $\hat{Q}_{\mathbf{k}\rho}(\omega)$, and (89)–(90g), respectively. Substituting (46b) and (46d) into (90c), we obtain the expression for the damping $\hat{\gamma}_{\mathbf{k}\mu}(\omega)$:

$$\hat{\gamma}_{\mathbf{k}\mu}(\omega) = N^{-1} \sum_{\mathbf{q},r,\mu'} |[E_{\mu\mu'}(\mathbf{k},\mathbf{k}-\mathbf{q}) + F^{(1)}_{\mu\mu'}(\mathbf{k},\mathbf{k}-\mathbf{q}) - F^{(2)}_{\mu\mu'}(\mathbf{k},\mathbf{k}-\mathbf{q})] Y_r(\mathbf{q})|^2 \times (n_0 - n_\mu)(n_0 - n_{\mu'}) \operatorname{Im} D^{(0)}_{\mathbf{q}r}(\mathbf{k}-\mathbf{q},\mu';\omega). \tag{91}$$

Since $F^{(1)}_{\mu\mu'}(\mathbf{k},\mathbf{k}-\mathbf{q}) \sim F^{(2)}_{\mu\mu'}(\mathbf{k},\mathbf{k}-\mathbf{q})$, the main contribution to the damping $\hat{\gamma}_{\mathbf{k}\mu}(\omega)$ that causes the asymmetry to the spectral line arises from the first term on the rhs of (91) corresponding to the coupling between the isolated bare exciton and the phonon field.

In our discussion for the expression of the absorption coefficient corresponding to the coupled polariton–phonon, as well as for the bare exciton–phonon spectrum, the shape function is found to be an asymmetric Lorentzian even if we neglect the ω dependence of the damping function. The asymmetry in each case is caused by the coupling of the electronic excitations and the intermolecular interactions with the acoustic branches of the phonon field. In our treatment we have ignored not only the mixing of different exciton bands, assuming that the exciton bands are well separated, but also the ω dependence of the energy shift and damping functions. If these two

effects are taken into consideration, additional asymmetries will appear in the lineshape. In the studies of Toyozawa[16] and Davydov and Myasnikov[12] for the bare exciton–phonon interaction, the deviations from the symmetric Lorentzian line appeared only when mixing of different bands and the ω dependence of the energy shift and spectral width were taken into account, respectively. We believe that both effects are second order in comparison with the expression causing the asymmetry of the spectral line derived in the present study. This is due to the fact that the mixing of different bands becomes substantial only in the case where the energy bands are very close together, while the variation of the energy shift and spectral width with respect to ω is important for certain regions of frequencies where the density of the final states is large. Of course, in such cases both effects should be taken into consideration leading to more complicated lineshapes with the possibility of the appearance of new peaks corresponding to combination bands when multiphonon effects are also included.

VI. AVERAGE ENERGY

The average energy of the system is obtained by averaging the Hamiltonian (13), i.e.,

$$\langle \mathcal{H} \rangle = \langle \mathcal{H}^{(0)} \rangle + \langle \mathcal{H}_{\text{phonon}} \rangle + \langle \mathcal{H}_{\text{int}} \rangle. \tag{92}$$

Using (12), (31), (32), and (62), we find the following expression for the average energy (92):

$$\langle \mathcal{H} \rangle = \langle \mathcal{H}_0^{(0)} \rangle + \frac{1}{2} \sum_{\mathbf{q},r} \frac{\omega_r^2(\mathbf{q})}{\tilde{\omega}_r(\mathbf{q})} \coth \tfrac{1}{2}\beta\tilde{\omega}_r(\mathbf{q}) + \frac{1}{\pi} \sum_{\mathbf{k},\rho,\lambda} \int_{-\infty}^{+\infty} d\omega \, \mathrm{Im} \frac{[\omega + \omega_{\rho\lambda}(\mathbf{k})]\omega(1 - e^{\beta\omega})^{-1}}{[\omega^2 - \Omega_{\mathbf{k}\rho}^{(1)}(\omega)\Omega_{\mathbf{k}\rho}^{(2)}(\omega)]} \tag{93}$$

To integrate over ω we have to assume that ω is far from any absorption frequency of the crystal, which means that the functions $\Omega_{\mathbf{k}\rho}^{(1)}(\omega)$ and $\Omega_{\mathbf{k}\rho}^{(2)}(\omega)$ are real; then integration over ω leads to

$$\langle \mathcal{H} \rangle = \langle \mathcal{H}_0^{(0)} \rangle + \frac{1}{2} \sum_{\mathbf{q},r} \frac{\omega_r^2(\mathbf{q})}{\tilde{\omega}_r(\mathbf{q})} \coth \tfrac{1}{2}\beta\tilde{\omega}_r(\mathbf{q}) + \frac{1}{2} \sum_{\mathbf{k},\rho,\lambda} [\Omega_{\rho\lambda}(\mathbf{k}) \coth \tfrac{1}{2}\beta\Omega_{\rho\lambda}(\mathbf{k}) - \omega_{\rho\lambda}(\mathbf{k})] \times \left[1 - \left| \frac{\partial \Omega_{\mathbf{k}\rho}^{(1)}(\omega)\Omega_{\mathbf{k}\rho}^{(2)}(\omega)}{\partial \omega^2} \right|_{\omega=\Omega_{\rho\lambda}(\mathbf{k})}\right]^{-1} \tag{94}$$

where $\Omega_{\rho\lambda}(\mathbf{k})$ is the positive solution of Eq. (84). The last term on the rhs of (94) represents the energy of interaction resulting from the polariton–phonon interaction at finite temperatures. It is given as the difference between the energy of excitation of the coupled polariton–phonon system and the unperturbed polariton energy. In the limit $\beta \to \infty$, the expression (94) gives the ground-state energy of the crystal. When retardation effects are not important, i.e., when $\omega_{\rho\lambda}(\mathbf{k}) = E_\mu(\mathbf{k})$, then expression (94) is reduced to that of the average energy due to the coupling between the bare exciton–phonon system.

VII. SUMMARY

The present study is concerned with the excitation spectrum arising from the interaction between polaritons and acoustic phonons in molecular and insulating crystals. In the limit of weak polariton–phonon coupling, the various Green's functions appearing in the expression for the dielectric permeability have been calculated in the one-phonon approximation by means of a model Hamiltonian for the crystal. The described physical process is due to the scattering of a polariton from $(\mathbf{k}, \rho, \lambda)$ to $(\mathbf{k} - \mathbf{q}, \rho', \lambda')$, with the emission or absorption of a phonon $(\mathbf{q}, r)$. The lineshape for the process in question at finite temperatures is considered and is found to be an asymmetric Lorentzian line, if the frequency dependence of the energy shift and the spectral width is neglected. The temperature dependence of the frequency- and wave-vector-dependent absorption coefficient is discussed. In the range of wave vectors where retardation may be neglected, the lineshape of the exciton absorption bands is studied in detail and compared with the results derived from previous studies.

The present work was motivated by the fact that most of the observed exciton absorption spectra show some asymmetry.[16] Toyozawa[16] attributed the asymmetry of the absorption lines as due to the interband as well as intraband scattering of excitons by phonons in insulating crystals. We refer to his papers,[16] where comparison between existing theories and experimental results is made. A different process causing asymmetry in the lineshape of the absorption bands is proposed in the present study. It is hoped that this work will stimulate interest in measuring lineshapes of exciton absorption spectra, particularly in molecular crystals where the existing data are not sufficient to warrant comparison with theory.

ACKNOWLEDGMENT

The author is indebted to Dr. M. W. C. Dharmawardana for reading the manuscript.

[16] Y. Toyozawa, J. Phys. Chem. Solids **25**, 59 (1964) and references therein.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

On Some Relations between the Solutions and the Parameters of Second-Order Linear Differential Equations*

NARENDRA S. GOEL
Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627
AND
FRANK P. BUFF
Department of Chemistry, University of Rochester, Rochester, New York 14627

(Received 19 May 1969)

The calculation of the Green's function for a point charge near an anisotropic planar diffuse layer leads to the differential equation $U'' - k^2D^2(u)U = -2k\delta(u - u')$. Solutions for small and large k are obtained readily by a procedure equivalent to the principle of invariant imbedding. It is shown that the additional terms introduced in the solution for large k, due to the lack of smoothness in $D(u)$ specifying the dielectric properties of the layer, may lead to singularities in the image potential. Similar considerations apply in the case of 1-dimensional wave propagation through a stratified medium.

In many physical problems when the properties of the medium vary as a function of position, as a simplification one assumes that the variation is limited to a strip or layer of finite width h. Some examples of such physical problems are the electrostatics of a point charge near a diffuse boundary between two dielectrics, wave propagation through a stratified medium, scattering from a potential field, etc. The equations which describe these phenomena depend on a parameter k in addition to those describing the medium. It is the purpose of this note to show that the dependence on small and large k may be readily calculated by a simple method. The lack of smoothness of the physical parameters introduces additional terms in the solution for large k. We illustrate this approach by the electrostatic problem mentioned above, where these additional terms introduce singularities in the image potential. We restrict ourselves to a brief discussion of a planar boundary with discontinuities in dielectric constant and its derivatives only at the boundaries of this layer and refer for further details to our forthcoming paper.[1]

Let us label the three different regions defining the diffuse boundary between two dielectrics of dielectric constants, D_1 and D_3, by 1, 2, and 3 as shown in Fig. 1.

Let there be a unit charge at the point $\mathbf{R}'$. Let us choose the z axis as the axis perpendicular to the interface (with $\mathbf{l}$ as the unit vector), the $+z$ axis being in the direction D_3 to D_1. Because of the cylindrical symmetry around this z axis, the dielectric tensor D is characterized by two dielectric parameters through the relation

$$\mathsf{D} = D_N(z)\mathbf{ll} + D_T(z)(\mathsf{I} - \mathbf{ll}), \tag{1}$$

where I is the unit tensor. D_N characterizes the dielectric properties of the medium perpendicular to the interface and D_T, parallel to the interface.

In general, we require the solution of the Poisson equation

$$\nabla \cdot \mathsf{D} \cdot \nabla\psi = -4\pi\delta(\mathbf{R} - \mathbf{R}'), \tag{2}$$

where ψ is the electrostatic potential at an arbitrary point $\mathbf{R}$.

Upon substitution of Eq. (1) into (2), making the transformation

$$D_N(z)\,du = dz, \tag{3}$$

and substituting for ψ the expression

$$\int_0^\infty J_0(kr)U(k, u, u')\,du,$$
$$r = [(x - x')^2 + (y - y')^2]^{\frac{1}{2}}, \tag{4}$$

we get for the three regions the equations[1]

$$\frac{d^2U_1}{du^2} - k^2D_1^2U_1 = -2k\delta(u - u'), \tag{5}$$

$$\frac{d^2U_2}{du^2} - k^2D^2U_2 = 0, \quad D^2 \equiv D_TD_N, \tag{6}$$

$$\frac{d^2U_3}{du^2} - k^2D_3^2U_3 = 0. \tag{7}$$

Using the standard method of variation of constants,[1,2] the general solution of Eq. (5) consistent

* Supported by a grant from the Office of Saline Water.
[1] F. P. Buff and N. S. Goel, J. Chem. Phys. **51**, 4983, 5363 (1969).
[2] F. Brauer and J. A. Nohel, *Ordinary Differential Equations* (W. A. Benjamin, Inc., New York, 1967).

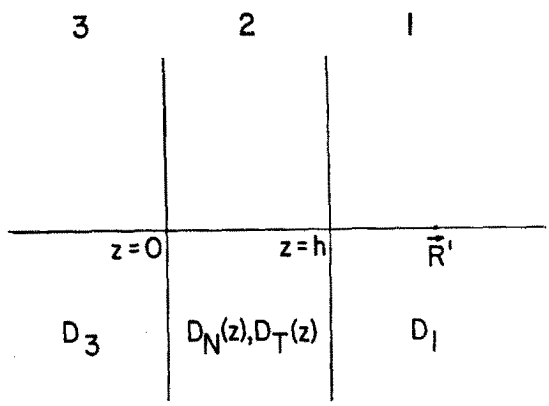


FIG. 1. The regions defining the diffuse boundary between two dielectrics.

with the boundary condition $U_1 \to 0$, $u \to \infty$ is

$$U_1 = (a/D_1)e^{-kD_1u} + U_p, \tag{8}$$

where a is an unknown constant and

$$U_p = (D_1)^{-1}e^{-kD_1|u-u'|}. \tag{9}$$

Further, the solution of Eq. (7) consistent with the boundary condition $U_3 \to 0$, $u \to -\infty$ is

$$U_3 = (d/D_3)e^{kz}, \tag{10}$$

where d is an unknown constant.

The solution of Eq. (6) can be taken as

$$U_2 = bU_{21} + cU_{22}, \tag{11}$$

where b and c are unknown constants. U_{21} and U_{22} are the two linearly independent solutions of Eq. (6), which may be chosen such that

$$U_{21}(0) = 1, \quad U'_{21}(0) = 1,$$
$$U_{22}(0) = 0, \quad U'_{22}(0) = 1. \tag{12}$$

By using the conditions that U and dU/du are continuous at $z = 0$ and $z = h$, it can easily be seen that[1]

$$U_1 = \frac{e^{-k|z-z'|}}{D_1} + \frac{e^{-k(z+z'-2h)}}{D_1} R(z = h), \tag{13}$$

where R is given by

$$R = (kD_1\mu - \mu')/(kD_1\mu + \mu'), \tag{14}$$

$$\mu' = d\mu/du, \tag{15a}$$

$$\mu = U_{21}/D_3 + U_{22}, \tag{15b}$$

$$R(z = 0) = (D_1 - D_3)/(D_1 + D_3). \tag{16}$$

$R(0)$ is the value of R for the simple case of an abrupt interface. To calculate R, we differentiate Eq. (14) with respect to u. Using the fact that μ satisfies Eq. (6) and utilizing Eq. (14), we get

$$\frac{dR}{du} = \frac{k}{2D_1}[D_1^2(1 - R)^2 - D^2(1 + R)^2]. \tag{17}$$

We observe that if one adapts the method of invariant imbedding[3,4] to this problem, one can recover Eqs. (16) and (17).

To obtain R, one has to solve the Ricatti equation (17) together with the boundary condition (16). In general, there is no analytical solution for this equation. However, as we will show, it is possible to find formal asymptotic expansions for R for small k (useful for calculating ψ_1 for large z) and for large k (useful for ψ_1 as $z = z' \to h$).

To obtain an expansion for R for small k, we substitute

$$R = \sum_0^\infty R_n k^n \tag{18}$$

into Eq. (17) and equate the coefficients of equal powers of k on the left- and right-hand sides of Eq. (17) to get

$$\frac{dR_0}{du} = 0, \quad \text{that is,} \quad R_0 = R(0), \tag{19a}$$

$$\frac{dR_1}{du} = (2D_1)^{-1}[D_1^2(1 - R_0)^2 - D^2(1 + R_0)^2], \quad \text{etc.} \tag{19b}$$

Integrating Eqs. (19) and restricting ourselves to first-order correction (higher-order corrections are examined in Ref. 1), we get

$$Re^{2kh}|_{z=h} = R_0(1 + 2kz_1) + O(k^2)$$
$$\simeq R_0e^{2kz_1} + O(k^2), \tag{20}$$

where

$$z_1 = h - \frac{D_1}{D_1^2 - D_3^2}\int_0^h \left(D_T - \frac{D_3^2}{D_N}\right) dz. \tag{21}$$

It may be noted that Eq. (21) can also be written as

$$\int_{-\infty}^{\infty} \left(D_T - \frac{D_3^2}{D_N}\right) dz = \int_{z_1}^{\infty} \left(D_1 - \frac{D_3^2}{D_1}\right) dz. \tag{22}$$

This equation is true for all h and, hence, for small k, the layer model is a convenient device for finding

[3] R. Bellman and R. Kalaba, J. Math. Mech. 8, 683 (1959).
[4] G. M. Wing, "The Method of Invariant Imbedding with Applications to Transport Theory and Other Areas of Mathematical Physics," in *Colloquium Lectures in Pure and Applied Science, No. 10* (Field Research Laboratory, Socony Mobil Oil Co., Inc., Dallas, Texas, 1965).

z_1 for any diffuse boundary. Substituting Eq. (20) into (13), we get

$$U_1 = (D_1^{-1})(e^{-k|z-z'|} + R_0 e^{-k(z+z'-2z_1)}). \quad (23)$$

It may be noted that the above procedure for calculating R_n is significantly more efficient than the procedure based on Green's functions. Equation (23) when substituted into Eq. (4) gives asymptotically the potential ψ in region 1, so that

$$\lim_{\substack{R\to R' \\ z'\gg h}} \psi_{\text{im}} \sim \frac{R_0}{2D_1\,|z - z_1|}, \quad (24a)$$

where

$$\psi_{\text{im}} = \psi - \frac{1}{D_1\,|\mathbf{R} - \mathbf{R}'|}. \quad (24b)$$

$\frac{1}{2}[\psi_{\text{im}}(\mathbf{R} \to \mathbf{R}')]$ is the image potential of the charge in region 1.

Therefore, for distances far from the interface, the leading term in the electrostatic potential is the same as one would get by assuming a sharp boundary between the two dielectrics, but shifted to the point $z = z_1 \neq 0$.

As $z = z' \to h$, it is necessary to examine Eq. (17) for large k. For this purpose, we make the transformation

$$Q = \frac{D_1}{D}\frac{1 - R}{1 + R}. \quad (25)$$

Substituting this equation into (16) and (17), we get

$$Q(0) = D_3/D(0), \quad (26)$$

$$\frac{dQ}{d\omega} = k(1 - Q^2) - Q\frac{\dot{D}}{D}, \quad (27)$$

where

$$d\omega \equiv \left(\frac{D_T}{D_N}\right)^{\frac{1}{2}} dz, \quad (28a)$$

$$\dot{D} = \frac{dD}{d\omega}. \quad (28b)$$

It may be noted that the Ricatti equation (27) enables us to calculate R (and hence ψ) exactly, if $\dot{D}/D$ is either zero or constant.

To get an asymptotic expansion for

$$k \gg 1/\omega(h), \quad k \gg \dot{D}/D,$$

we substitute

$$Q = \sum_0^\infty Q_n k^{-n} \quad (29)$$

into Eq. (27) and choose Q_0 so that

$$\frac{dQ_0}{d\omega} = k(1 - Q_0^2) \quad (30a)$$

and then equate the coefficients of equal powers of k on both sides of the equation to get

$$0 = 2Q_1 + \dot{D}/D, \quad (30b)$$

$$\dot{Q}_1 = -2Q_0Q_2 - Q_1(Q_1 + \dot{D}/D), \quad (30c)$$

and

$$\dot{Q}_n = -\sum_{\substack{l\neq 1 \\ m\neq 1}} Q_l Q_m \delta_{l+m,n+1}, \quad n \geq 2. \quad (30d)$$

Equations (30) can be solved to give

$$Q_0(\omega) = \frac{Q_0(0) + \tanh k\omega}{1 + Q_0(0)\tanh k\omega}, \quad (31a)$$

$$Q_1 = -r(\omega), \quad r(\omega) = \frac{d \ln D^{\frac{1}{2}}}{d\omega}, \quad (31b)$$

$$Q_2 = \frac{\dot{r}(\omega) + r^2(\omega)}{2Q_0}, \quad (31c)$$

$$Q_3 = -\frac{\dot{Q}_2}{2Q_0}, \quad (31d)$$

and

$$Q_n = -\frac{1}{2Q_0}\left(\dot{Q}_{n-1} + \sum_{\substack{l\geq 2 \\ n-l\geq 2}} Q_l Q_{n-l}\right), \quad n \geq 4, \quad (31e)$$

where $Q_0(0)$ is to be determined by the boundary condition (26). Thus

$$Q_0(0) = \frac{D_3}{D(0)} + \frac{r(0)}{k} - \frac{D(0)}{D_3}\frac{\dot{r}(0) + r^2(0)}{2k^2} + \cdots.$$

Substituting Eqs. (31) into Eq. (29), we get Q, which, when substituted into Eq. (25), gives R. The asymptotic value of R thus obtained is

$$R|_{z=h} \sim \frac{1}{(D_1 + qD_h)^2} \times \left[(D_1^2 - q^2D_h^2) + \frac{2D_1D_h r(h)}{k} - \frac{D_1D_h}{k^2} \times \left(\frac{\dot{r}(h) + r^2(h)}{q} - \frac{2D_h}{D_1 + qD_h}r^2(h)\right) + \cdots\right], \quad (32)$$

where

$$D_h = D(z = h),$$

$$q = Q_0(\omega = h) \xrightarrow[k\to\infty]{} 1.$$

We observe that when the first-order WKB[5] calculation associated with Eq. (6) is carried out, the result is identical to that in the above equation to order k^{-1}. Similarly, the second-order WKB calculation gives result identical to that in Eq. (32) to order k^{-2}, etc. Thus, the method which we have discussed is a simple method of obtaining the dominant terms in the potential at a point close to the boundary, which one gets from WKB calculations in various orders, the latter one involving a significant amount of calculations. We may add that the Ricatti Eq. (27) can also be used to get an expression for Q (and hence for R) for small k by making a substitution of type (18).

Using Eqs. (32), (13), and (4), it can be shown[1] that, for $z = z' \to h$, ψ_{im} diverges as $(z-h)^{-1}$, if at $z = h$ D is discontinuous, or as $-\log(z-h)$, if D is continuous, but $dD/d\omega$ is discontinuous and is finite when both D and $dD/d\omega$ are continuous. Thus, the lack of smoothness at $z = h$ may introduce singularities into the potential at the boundary.

It may be noted that knowledge of Q (or R) is sufficient to determine d as defined by Eq. (10) and, hence, the potential at an arbitrary point in region 3. We also observe[1] that $R(z)$ and a functionally similar quantity $\bar{R}(z)$ permit calculation of ψ everywhere with arbitrary location of charge.

Let us now consider the wave propagation through a plane slab of finite width h where the wavenumber varies as a function of position coordinate u. Let us suppose that there is a wave coming from the right-hand side through a medium 1 with a constant wavenumber k_1 ($\equiv kD_1$) and impinging on the slab 2 with a position-dependent wavenumber $k_2(u) \equiv (kD(u))$. This wave, in part, will be reflected and, in part, will be transmitted through the medium 3 with constant wavenumber k_3 ($\equiv kD_3$) (Fig. 2). The wave vectors ψ_1, ψ_2, and ψ_3 in the three regions will satisfy the wave equations

$$\frac{d^2\psi_1}{du^2} + k^2 D_1^2 \psi_1 = 0, \quad h \leq u < \infty, \tag{33}$$

$$\frac{d^2\psi_2}{du^2} + k^2 D^2(u)\psi_2 = 0, \quad 0 < u < h, \quad D(\infty) = D_1, \quad D(-\infty) = D_3, \tag{34}$$

$$\frac{d^2\psi_3}{du^2} + k^2 D_3^2 \psi_3 = 0, \quad -\infty < u \leq 0. \tag{35}$$

As in the case of electrostatics, by using the conditions that ψ and $d\psi/du$ are continuous at $u = 0$ and

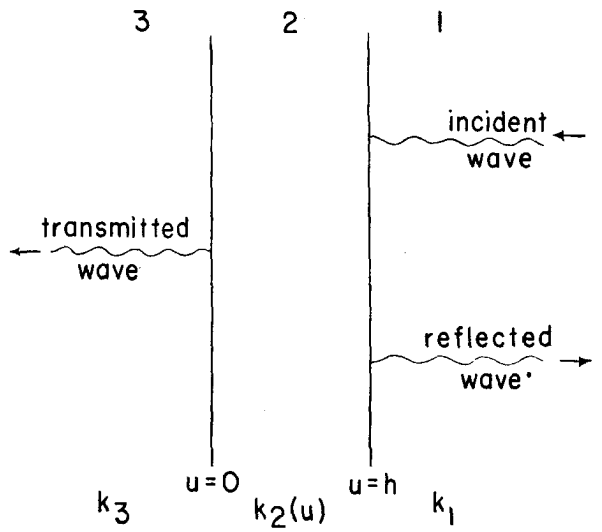


FIG. 2. Transmitted and reflected waves originating from single incident wave at the boundary between slab and region 2.

$u = h$, it can be easily seen that[3]

$$\psi_1 = e^{-ik_1(u-h)} + e^{ik_1(u-h)}R(h), \tag{36}$$

where R, the ratio of the amplitudes of the reflected and incident waves, satisfies the equations

$$R(0) = (D_1 - D_3)/(D_1 + D_3), \tag{37}$$

$$\frac{dR}{du} = \frac{ik}{2D_1}[D^2(1+R)^2 - D_1^2(1-R)^2]. \tag{38}$$

Equations (37) and (38) are very similar to Eqs. (16) and (17) and so can be treated likewise.

To obtain an asymptotic expansion for small k, we substitute

$$R = \sum R_n (ik)^n \tag{39}$$

into Eq. (38) and equate the coefficients of equal powers of k on the left- and right-hand sides of Eq. (38) to get

$$R = R(0)\left(1 + \frac{2ikD_1}{D_1^2 - D_3^2}\int_0^h (D^2 - D_3^2)\,du + \cdots\right). \tag{40}$$

This is the same result as obtained by MacLaurin[6] using the more laborious conventional Green's function method. Elsewhere,[7] we have applied this small k expansion for the calculation of ellipticity of light reflected from an optically anisotropic stratified medium.

[5] R. Bellman, *Perturbation Techniques in Mathematics, Physics, and Engineering* (Holt, Rinehard and Winston, Inc., New York, 1966).

[6] R. C. MacLaurin, Proc. Roy. Soc. (London) **A74**, 49 (1905).

[7] F. P. Buff and R. A. Lovett, in *Simple Dense Fluids*, H. L. Frisch and Z. W. Salsburg, Eds. (Academic Press Inc., New York, 1968).

To obtain a formal asymptotic expansion for R for large k, we make the transformation

$$Q = \frac{D_1}{D}\frac{R-1}{R+1} \tag{41}$$

to obtain the equations

$$Q(0) = -D_3/D(0), \tag{42}$$

$$\frac{dQ}{d\omega} = ik(1-Q^2) - Q\frac{\dot{D}}{D}, \tag{43}$$

where

$$d\omega = D\,du \tag{44a}$$

and

$$\dot{D} = dD/d\omega. \tag{44b}$$

It may be noted that Eq. (43) enables us to calculate the reflection coefficient exactly, if either $\dot{D}$ is equal to 0 (Airy's model[8]) or $\dot{D}/D$ is equal to a constant (Rayleigh's model[9]).

Equation (43) is exactly Eq. (27), if one replaces k in Eq. (27) by ik. Therefore, we can write

$$Q = \sum Q_n(ik)^{-n}, \tag{45}$$

where

$$Q_0 = \frac{Q_0(0) + i\tan k\omega}{1 + iQ_0(0)\tan k\omega} \tag{46}$$

and the Q_n, $n > 1$, are given by Eqs. (31b)–(31e). To determine $Q_0(0)$, as before, we have to use the boundary condition (42).

Contrary to the electrostatic case, we do not achieve the simplification $Q_0(h) \to 1$ for $k \to \infty$. However, if we limit ourselves to the special case in which the discontinuity in D and its derivative occurs only at $z = h$, the calculations are simplified. For this case,

$$Q_0 = -1. \tag{47}$$

Using Eqs. (31b)–(31e), (47), and (41), we get

$$R|_{u=h} \sim \frac{1}{(D_1 + D_h)^2} \times \left[(D_1^2 - D_h^2) - \frac{2D_hD_1r(h)}{ik} - \frac{D_hD_1}{k^2} \times \left(\dot{r}(h) + r^2(h) - \frac{2D_h}{D_1 + D_h}r^2(h)\right) + \cdots\right]. \tag{48}$$

For $D_1 = D_h$ up to order k^{-1}, this reduces to the expression given by Landau and Lifshitz[10] in connection with a similar problem of reflection of a particle above a barrier within the quasiclassical limit, when the potential has a discontinuity in slope.

We finally observe that the transmission coefficient can be expressed as a quadrature in terms of Q.

[8] G. B. Airy, Phil. Mag.,Ser. 3, **2**, 20 (1833).
[9] Lord Rayleigh, Proc. London Math. Soc. **11**, 51 (1880).
[10] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics, Non-relativistic Theory* (Pergamon Press Inc., New York, 1965), p. 184.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Construction of Weight Spaces for Irreducible Representations of A_n, B_n, C_n, D_n

ROBERT GILMORE*
Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 9 October 1968)

An algebraic technique is presented by means of which the weight space for the irreducible representations (λ) of A_n, B_n, C_n, D_n can be constructed from the weight spaces associated with the representations (λ') of the subalgebras A_{n-1}, B_{n-1}, C_{n-1}, D_{n-1}. Since each chain ends with A_1, all weight spaces of the classical simple Lie algebras associated with (λ) can be constructed, ultimately, from the well-known representations of A_1.

I. INTRODUCTION

Algebraic techniques are presented for determining the irreducible representations (λ') of a canonical subalgebra A', subduced from the representation λ of A under the restriction $A \downarrow A'$, for the following six cases: $A_n \downarrow A_{n-1}$, $B_n \downarrow D_n \downarrow B_{n-1} \downarrow D_{n-1}$, $C_n \downarrow C_{n-1}$. The converse problem, that of determining which irreducible representations (λ) of A contain (λ') of A' under the subalgebra restriction $A \downarrow A'$, is solved using similar techniques. These two processes are then used, back to back, to solve the weight-multiplicity problem. That is, the irreducible representations λ of A can be synthesized from the representations λ' of the subalgebra A' in the following cases: $A_n \downarrow A_{n-1}$, $D_n \downarrow D_{n-1}$, $B_n \downarrow B_{n-1}$, and $C_n \downarrow C_{n-1}$. Since each of these four chains of simple Lie algebras terminates with A_1, whose representations are well known and thoroughly understood, the representations (λ) of the classical simple Lie algebras can all be constructed by such bootstrap techniques as

$$A_1 \uparrow A_2 \uparrow A_3 \uparrow \cdots \uparrow A_{n-1} \uparrow A_n \cdots$$

or

$$A_1 \uparrow A_1 + A_1 \cong D_2 \uparrow D_3 \uparrow \cdots \uparrow D_{n-1} \uparrow D_n \cdots$$

from the representations of A_1.

II. BRANCHING RULES FOR CANONICAL SUBGROUPS

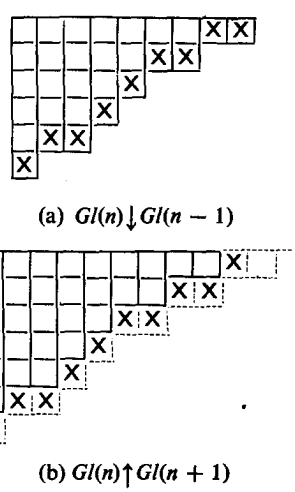
The finite set of irreducible representations for the general linear groups $Gl(n, c)$, $Gl(n, r)$ is in one-to-one correspondence with the set of all partitions of an integer λ into no more than n parts[1]:

$$\lambda = (\lambda_1, \cdots, \lambda_n), \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0. \quad (1)$$

Under restriction to the canonical subgroup $Gl(n-1)$, obtained by leaving the nth coordinate of the carrier space invariant, an arbitrary irreducible representation λ of $Gl(n)$ will subduce a number of irreducible representations of the subgroup $Gl(n-1)$:

$$\Gamma_\lambda \xrightarrow{Gl(n)\downarrow Gl(n-1)} \sum_{\lambda'} \Gamma^{\lambda'} \quad (2)$$

In a tableau representing a basis vector for the representation λ of $Gl(n)$, the only positions in which the index n can occur lie at the end of a column, as shown in Fig. 1(a). Representations of $Gl(n-1)$



(a) $Gl(n)\downarrow Gl(n-1)$

(b) $Gl(n)\uparrow Gl(n+1)$

FIG. 1. (a) In the partition λ, the only positions in which an n can occur lie at the bottom of each column. These are removed by the "box annihilation operator" P_n. (b) In the partition λ' the only positions to which a box with an $n+1$ can occur lie at the bottom of each column. These are placed there by the "box creation operator" $P^\dagger_{n+1}$.

subduced from λ of $Gl(n)$ do not contain any coordinate indices n in the tableaux which represent basis vectors. Therefore, those representations of $Gl(n-1)$ occur in this reduction corresponding to the removal of all combinations of the boxes labeled X in Fig. 1(a):

$$\Gamma^\lambda \xrightarrow{Gl(n)\downarrow Gl(n-1)} \sum_{\lambda'} \Gamma^{\lambda'}. \quad (3)$$

* This work is supported by Office of Naval Research and National Science Foundation.

[1] H. Weyl, *Classical Groups* (Princeton University Press, Princeton, N.J., 1946).

The sum in Eq. (3) is over all partitions λ' which obey

$$\lambda_1 \geq \lambda_1' \geq \lambda_2 \geq \lambda_2' \geq \cdots \geq \lambda_{n-1}' \geq \lambda_n \geq 0. \quad (4)$$

Each representation $\Gamma^{\lambda'}$, with λ' described by Eq. (4), occurs exactly once.[1]

Equations 3 and 4 can be summarized neatly by introducing the ordered annihilation operator

$$P_n = (1 - \epsilon_n)^{-1}(1 - \epsilon_{n-1})^{-1} \cdots (1 - \epsilon_2)^{-1}(1 - \epsilon_1)^{-1}. \quad (5)$$

The operator ϵ_i is to be interpreted as a box annihilation operator in the ith row:

$$\epsilon_i(\lambda_1 \cdots \lambda_{i-1}, \lambda_i, \lambda_{i+1} \cdots \lambda_n) = (\lambda_1 \cdots \lambda_{i-1}, \lambda_i - 1, \lambda_{i+1} \cdots \lambda_n) \quad (6)$$

and each term $(1 - \epsilon_i)^{-1}$ is a representation for the power series of annihilation operators

$$(1 - \epsilon_i)^{-1} = 1 + \epsilon_i + \epsilon_i^2 + \epsilon_i^3 + \cdots. \quad (7)$$

Then Eqs. (3) and (4) can be rewritten

$$(\lambda)_n \xrightarrow{Gl(n)\downarrow Gl(n-1)} \sum_{\lambda' \text{ as in (4)}} (\lambda') = P_n(\lambda)_{n-1}. \quad (8)$$

The operator in Eq. (8) will sometimes leave us with partitions containing columns of length n. Thus, if λ has n rows, the leading term in $P_n = \prod_{i=1}^{n} (\epsilon_i^{j_i}) = 1 + \cdots$ will yield a partition $\lambda' = \lambda$. Since completely antisymmetric nth-order tensors cannot be constructed in an $(n-1)$-dimensional vector space, representations of $Gl(n-1)$ corresponding to partitions with $\lambda_n' > 0$ vanish. The complete expression of Eqs. (3) and (4) is given by Eq. (8), coupled with the modification rules for $Gl(n) \downarrow Gl(n-1)$:

$$(\lambda_1' \cdots \lambda_{n-1}'\lambda_n')_{n-1} = (\lambda_1' \cdots \lambda_{n-1}')_{n-1}, \quad \lambda_n' = 0, \qquad = 0, \quad \lambda_n' > 0. \quad (9)$$

Representations of the special linear groups $Sl(n+1)$ are in one-to-one correspondence with partitions into no more than n parts. But, since the completely skew-symmetric tensor (1^{n+1}) has the transformation properties of the identity,

$$(\lambda_1 \cdots \lambda_n\lambda_{n+1}) \equiv (\lambda_1 - \lambda_{n+1}, \lambda_2 - \lambda_{n+1}, \cdots, \lambda_n - \lambda_{n+1}).$$

Using arguments similar to those for $Gl(n)$, we easily construct the spectrum of irreducible representations of $Sl(n)$ subduced from λ of $Sl(n+1)$:

$$(\lambda)_{n+1} \xrightarrow{Sl(n+1)\downarrow Sl(n)} \sum (\lambda')_n = P_n(\lambda)_n. \quad (10)$$

Modification rules for $Sl(n+1) \downarrow Sl(n)$ are

$$(\lambda_1'\lambda_2' \cdots \lambda_{n-1}'\lambda_n')_n \equiv (\lambda_1' - \lambda_n', \lambda_2' - \lambda_n' \cdots \lambda_{n-1}' - \lambda_n')_n. \quad (11)$$

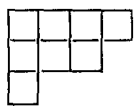
These arguments can be extended directly[1] to the groups $SU(n+1)$ and $SU(n)$ or their corresponding algebras A_n and A_{n-1}:

$$\{\lambda\}_{n+1} \xrightarrow{A_n \downarrow A_{n-1}} \sum \{\lambda'\}_n = P_n\{\lambda\}_n. \quad (12)$$

Modification rules for $A_n \downarrow A_{n-1}$ are

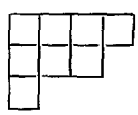
$$\{\lambda_1' \cdots \lambda_{n-1}'\lambda_n'\}_{n+1} \equiv \{\lambda_1' - \lambda_n', \lambda_2' - \lambda_n' \cdots \lambda_{n-1}' - \lambda_n'\}_n. \quad (13)$$

For example, let us determine the spectrum of representations subduced from the representation associated with

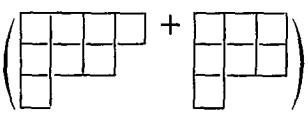


under the restrictions $Gl(n) \downarrow Gl(n-1)$, $Sl(n+1) \downarrow Sl(n)$, $SU(n+1) \downarrow SU(n)$. Applying P_3 to the partition (4, 3, 1), we obtain the solution

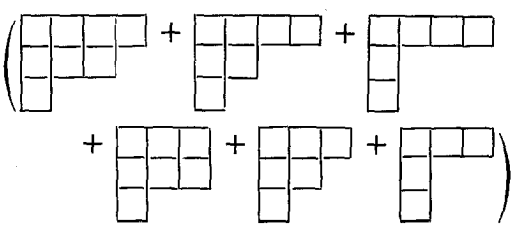
$P_3(4, 3, 1)$

$= (1 - \epsilon_3)^{-1}(1 - \epsilon_2)^{-1}(1 + \epsilon_1 + \cdots)$  (14)

$= (1 - \epsilon_3)^{-1}(1 + \epsilon_2 + \epsilon_2^2 + \cdots)$

$\times$ 

$= (1 + \epsilon_3 + \cdots)$

$\times$ 

$= (4, 3, 1) + (4, 3) + (4, 2, 1) + (4, 2)$
$+ (4, 1, 1) + (4, 1) + (3, 3, 1) + (3, 3)$
$+ (3, 2, 1) + (3, 2) + (3, 1, 1) + (3, 1).$ (15)

For $Gl(n) \downarrow Gl(n-1)$ and $n - 1 \geq 3$, Eq. (15) stands without modification. For $n - 1 = 2$, all partitions into more than two parts vanish:

$$(4, 3, 1) \xrightarrow{Gl(3)\downarrow (2} (4, 3)_2 + (4, 2)_2 + (4, 1)_2 + (3, 3)_2 + (3, 2)_2 + (3, 1)_2. \quad (16)$$

For $n - 1 = 1$, the partition (4, 3, 1) is not a representation of $Gl(2)$. For the restrictions

$$Sl(n+1, c) \downarrow Sl(n, c), \quad Sl(n+1, r) \downarrow Sl(n, r),$$

and

$$SU(n+1) \downarrow SU(n) \quad \text{with} \quad n > 3,$$

Eq. (15) remains unmodified. But, for $n = 3$,

$$\{\lambda_1' \lambda_2' 1\}_3 \equiv \{\lambda_1' - 1, \lambda_2' - 1\}_3:$$

$$\{4,3,1\}_4 \xrightarrow{SU(4)\downarrow \quad (3)} \{3,2\}_3 + \{4,3\}_3 + \{3,1\}_3 + \{4,2\}_3 + \{3,0\}_3 + \{4,1\}_3 + \{2,2\}_3 + \{3,3\}_3 + \{2,1\}_3 + \{3,2\}_3 + \{2,0\}_3 + \{3,1\}_3. \tag{17}$$

For $n = 3$, $\{4,3,1\}_3 \equiv \{3,2\}_3$, and the modification rules (11) and (13) give

$$\{4,3,1\}_2 \equiv \{3,2\}_3 \xrightarrow{SU(3)\downarrow SU(2)} \{1\}_2 + \{2\}_2 + \{3\}_2 + \{0\}_2 + \{1\}_2 + \{2\}_2. \tag{18}$$

Similar results hold for the unimodular orthogonal groups $SO(2n+1)$ and $SO(2n)$.[2] The representations of the algebra B_n, whose associated compact group is $SO(2n+1)$, are in one-to-one correspondence with partitions into no more than n parts. The character of the representation of $SO(2n+1)$ associated with the partition $(\boldsymbol{\lambda})$ is denoted[3] $[\boldsymbol{\lambda}]_{2n+1}$. Similarly, the representations of the algebra D_n are in one-to-one correspondence with partitions into n parts $(\boldsymbol{\lambda})$. The character of the associated connected compact group is denoted[3] $[\boldsymbol{\lambda}]_{2n}$. By arguments similar to those used for $Gl(n)$, $Sl(n+1)$, and $SU(n+1)$, we find[2,3]

$$[\boldsymbol{\lambda}]_{2n+1} \xrightarrow{B_n \downarrow D_n} P_n[\boldsymbol{\lambda}]_{2n}, \tag{19}$$

$$[\boldsymbol{\lambda}]_{2n} \xrightarrow{D_n \downarrow B_{n-1}} P_n[\boldsymbol{\lambda}]_{2n-1}. \tag{20}$$

There are no modification[3-5] rules for $B_n \downarrow D_n$.

Modification rules for $D_n \downarrow B_{n-1}$ are

$$[\lambda_1' \cdots \lambda_{n-1}' \lambda_n']_{2n-1} = [\lambda_1' \cdots \lambda_{n-1}']_{2n-1}, \quad \lambda_n' = 0, 1, \quad = 0, \quad \text{for} \quad \lambda_n' > 1. \tag{21}$$

In order to discuss directly the cases $B_n \downarrow B_{n-1}$ and $D_n \downarrow D_{n-1}$, we merely apply (19)–(21), back to back:

$$[\boldsymbol{\lambda}]_{2n+1} \xrightarrow{B_n \downarrow B_{n-1}} P_n P_n[\boldsymbol{\lambda}]_{2n-1}, \tag{22}$$

$$[\boldsymbol{\lambda}]_{2n} \xrightarrow{D \downarrow D_{n-1}} P_{n-1} P_n[\boldsymbol{\lambda}]_{2n-2}. \tag{23}$$

In Eqs. (22) and (23), the modification rules (21) are applied after the last annihilation operator P_n has been applied. It can be easily checked that the processes indicated in Eqs. (23) and (21) are equivalent to (24) and (25):

$$[\boldsymbol{\lambda}]_{2n} \xrightarrow{D_n \downarrow D_{n-1}} P_n P_n[\boldsymbol{\lambda}]_{2n-2}. \tag{24}$$

Modification rules for $D_n \downarrow D_{n-1}$ are

$$\begin{aligned} [\lambda_1' \cdots \lambda_{n-1}'; 2]_{2n-2} &= -[\lambda_1' \cdots \lambda_{n-1}']_{2n-2}, \\ [\lambda_1' \cdots \lambda_{n-2}', 1; 1]_{2n-2} &= +[\lambda_1' \cdots \lambda_{n-2}', 0]_{2n-2}, \\ \lambda_1' \cdots \lambda_{n-1}'; \lambda_n']_{2n-2} &= 0, \quad \text{otherwise.} \end{aligned} \tag{25}$$

The fourth and final series of simple Lie algebras C_n is associated with the symplectic groups $Sp(2n, r)$. The representations of C_n are in one-to-one correspondence with partitions $(\boldsymbol{\lambda})$ into no more than n parts. The compact group associated with C_n is called $USp(2n)$. The character of the unitary representation of $USp(2n)$ associated with partition $(\boldsymbol{\lambda})$ is denoted[3] $\langle \boldsymbol{\lambda} \rangle_{2n}$. The spectrum of representations $\langle \boldsymbol{\lambda}' \rangle_{2n-2}$, subduced from $\langle \boldsymbol{\lambda} \rangle_{2n}$ under the restriction $C_n \downarrow C_{n-1}$, is given by Eqs. (26) and (27):

$$\langle \boldsymbol{\lambda} \rangle_{2n} \xrightarrow{C_n \downarrow \quad 1} P_n P_n \langle \boldsymbol{\lambda} \rangle_{2n-2}. \tag{26}$$

Modification[5,6] rules for $C_n \downarrow C_{n-1}$ are

$$\langle \lambda_1' \cdots \lambda_{n-1}'; \lambda_n' \rangle_{2n-2} = \langle \lambda_1' \cdots \lambda_{n-1}' \rangle_{2n-2}, \quad \lambda_n' = 0, \quad = 0, \quad \text{for} \quad \lambda_n' > 0. \tag{27}$$

Results derived from a character analysis on the unitary representations of the compact groups associated with algebras A_n, B_n, C_n, D_n hold for the algebras as well, both compact and noncompact. These results, therefore, apply even to the finite-dimensional nonunitary representations of the noncompact groups associated with these four series of simple Lie algebras.

Similar results can be applied directly to discussion of the representations of the symmetric group. The representation of S_n are in one-to-one correspondence with partitions of the integer n. Thus, the representations of S_{n-1}, subduced from $(\boldsymbol{\lambda})_n$ of S_n, are

$$(\boldsymbol{\lambda})_n \xrightarrow{S_n \downarrow S_{n-1}} \left(\sum_{i=1}^{n} \epsilon_i\right)(\boldsymbol{\lambda})_{n-1}. \tag{28}$$

The operator $(\sum_{i=1}^{n} \epsilon_i)$ acts to reduce the number of boxes in the partition $(\boldsymbol{\lambda})$ by one. Since this operator is homogeneous of first order, this expression can be brought into a form similar to the results for simple Lie algebras:

$$(\boldsymbol{\lambda}) \xrightarrow{S_n \downarrow S_{n-1}} P_n(\boldsymbol{\lambda})_{n-1}.$$

[2] H. Boerner, *Representations of Groups* (North-Holland Publ. Co., Amsterdam, 1963).
[3] D. E. Littlewood, *Theory of Group Characters* (The Clarendon Press, Oxford, 1950).
[4] F. Murnaghan, *The Theory of Group Representations* (Dover Press, New York, 1963).
[5] M. J. Newell, Proc. Roy. Irish Acad. **A54**, 153 (1951).
[6] G. C. Hegerfeldt, J. Math. Phys. **8**, 1195 (1967).

TABLE I. Spectrum of representations (λ') of the subalgebra A' subduced from the representation (λ) of A under the restriction $A \downarrow A'$.

Subalgebra restriction $A \downarrow A'$	Spectrum of subduced representations	Modification rules
$A_n \downarrow A_{n-1}$	$\{\lambda\}_{n+1} \to P_n\{\lambda\}_n$	$\{\lambda_1'\lambda_2'\cdots\lambda_{n-1}'\lambda_n'\}_n \equiv \{\lambda_1' - \lambda_n', \lambda_2' - \lambda_n', \cdots, \lambda_{n-1}' - \lambda_n'\}_n$
$B_n \downarrow D_n$	$[\lambda]_{2n+1} \to P_n[\lambda]_{2n}$	
$D_n \downarrow B_{n-1}$	$[\lambda]_{2n} \to P_n[\lambda]_{2n-1}$	$[\lambda_1'\cdots\lambda_{n-1}'\lambda_n']_{2n-1} \equiv [\lambda_1'\cdots\lambda_{n-1}']_{2n-1}, \quad \lambda_n' = 0, 1$ $= 0, \quad \lambda_n' > 1$
$D_n \downarrow D_{n-1}$	$[\lambda]_{2n} \to P_nP_n[\lambda]_{2n-2}$	$[\lambda_1'\cdots\lambda_{n-1}'\lambda_n']_{2n-2} \equiv -[\lambda_1'\cdots\lambda_{n-1}']_{2n-2}, \quad \lambda_n' = 2$ $[\lambda_1'\cdots\lambda_{n-2}'1, 1]_{2n-2} \equiv +[\lambda_1'\cdots\lambda_{n-2}'0]_{2n-2}$
$B_n \downarrow B_{n-1}$	$[\lambda]_{2n+1} \to P_nP_n[\lambda]_{2n-1}$	$[\lambda_1'\cdots\lambda_{n-1}'\lambda_n']_{2n-1} = [\lambda_1\cdots\lambda_{n-1}']_{2n-1}, \quad \lambda_n' = 0, 1$ $= 0, \quad \lambda_n' > 1$
$C_n \downarrow C_{n-1}$	$\langle\lambda\rangle_{2n} \to P_nP_n\langle\lambda\rangle_{2n-2}$	$\langle\lambda_1'\cdots\lambda_{n-1}'\lambda_n'\rangle_{2n-2} = 0, \quad \lambda_n' > 0$
$S_n \downarrow S_{n-1}$	$(\lambda)_n \to P_n(\lambda)_{n-1}$	Partitions with more or less than $n-1$ boxes vanish

Modification rules for $S_n \downarrow S_{n-1}$ are: All partitions with more or fewer than exactly $n-1$ parts vanish. The foregoing results have been collected in Table I.

III. EMBEDDINGS IN LARGER ALGEBRAS

The problem solved in the last section can be turned around and asked again. That is, we might ask what representations (λ) of, say, $Gl(n+1)$ contain the representation (λ') of $Gl(n)$ under the restriction $Gl(n+1) \downarrow Gl(n)$. This problem is essentially one of adding boxes containing an $(n+1)$ to the partition (λ') in all possible ways. Such additions can only occur at the end of a column. This is essentially the converse of the problem described by Eq. (4) and in Fig. 1(a). The solution is presented in Eq. (4′) and Fig. 1(b):

$$\lambda_1 \geq \lambda_1' \geq \lambda_2 \geq \lambda_2' \geq \cdots \geq \lambda_n' \geq \lambda_{n+1} \geq 0. \quad (4')$$

It is clear that all the representations λ of $Gl(n+1)$ with this property are contained in the spectrum of

$$P_{n+1}^{\dagger}(\lambda') \quad (8')$$

with the creation operator $P_n^{\dagger}$ given by

$$P_n^{\dagger} = \{(1-\epsilon_n)^{-1}\cdots(1-\epsilon_1)^{-1}\}^{\dagger} = (1-\epsilon_1^{\dagger})^{-1}(1-\epsilon_2^{\dagger})^{-1}\cdots(1-\epsilon_n^{\dagger})^{-1}. \quad (5')$$

Referring to Table I, we easily construct Table II, which gives the spectrum of representations (λ) of A which contain the representation (λ') of the subalgebra A' under the restriction $A \downarrow A'$.

For example, let us determine the spectrum of representations of $SU(3)$ which contain $\square\square$ of $SU(2)$, under $SU(3) \downarrow SU(2)$. The solution is as follows:

$$P_2^{\dagger}\,\square\square = (1-\epsilon_1^{\dagger})^{-1}(1+\epsilon_2^{\dagger}+(\epsilon_2^{\dagger})^2+\cdots)\,\square\square = \sum_{a=0}^{\infty}(\epsilon_1^{\dagger})^a\left\{\square\square,\ \begin{array}{|c|c|}\hline\ \\\hline\end{array}\!\!\begin{array}{|c|}\hline\ \\\hline\end{array},\ \begin{array}{|c|c|}\hline\ & \\\hline\end{array}\right\} = \{2+a, 0\}_3, \{2+a, 1\}_3, \{2+a, 2\}_3. \quad (29)$$

We observe that all representations in this spectrum with $a > 0$ contain larger representations than $\square\square$ of $SU(2)$. In fact, the largest representation of $SU(2)$ contained in $\{2+a, 0 \text{ or } 1 \text{ or } 2\}_3$ of $SU(3)$ is $\{2+a\}_2$. By "larger" we are not referring to the dimensionality of the representation, but rather to its dominant weight or partition to which it corresponds. We say $(\mu) > (\nu)$ if the first nonvanishing term $\mu_i - \nu_i$ is positive.

TABLE II. Spectrum of representations (λ) of A which contain (λ') of A' under the subalgebra $A \downarrow A'$.

Algebra extension $A' \uparrow A$	Spectrum of representations (λ) of A containing	
	(λ') of A'	(λ') of A' as the largest subduced representation
$A_{n-1} \uparrow A_n$	$P_n^{\dagger}\{\lambda'\}_{n+1}$	$(1-\epsilon_n^{\dagger})^{-1}\{\lambda'\}_{n+1}$
$D_n \uparrow B_n$	$P_n^{\dagger}[\lambda']_{2n+1}$	$[\lambda']_{2n+1}$
$B_{n-1} \uparrow D_n$	$P_n^{\dagger}[\lambda']_{2n}$	$(1-\epsilon_n^{\dagger})^{-1}[\lambda']_{2n}$
$D_{n-1} \uparrow D_n$	$P_n^{\dagger}P_n^{\dagger}[\lambda']_{2n}$	$(1-\epsilon_n^{\dagger})^{-1}[\lambda']_{2n}$
$B_{n-1} \uparrow B_n$	$P_n^{\dagger}P_n^{\dagger}[\lambda']_{2n+1}$	$(1-\epsilon_n^{\dagger})^{-1}[\lambda']_{2n+1}$
$C_{n-1} \uparrow C_n$	$P_n^{\dagger}P_n^{\dagger}\langle\lambda'\rangle_{2n}$	$(1-\epsilon_n^{\dagger})^{-1}\langle\lambda'\rangle_{2n}$
$S_{n-1} \uparrow S_n$	$P_n^{\dagger}(\lambda)_{n-1}$	

Let us also consider the following problem: What representations of A_2 contain $\square\square$ of A_1 as the largest of the subduced representations? What other representations of A_1 are subduced from each of these representations of A_2? Referring to the last column in Table II, we obtain the spectrum

$$(1-\epsilon_2^+)^{-1}\ \square\square = \square\square,\ \begin{array}{|c|}\hline\ \\\hline\ \\\hline\end{array}\!\!\begin{array}{|c|}\hline\ \\\hline\end{array},\ \begin{array}{|c|c|}\hline\ & \\\hline\ & \\\hline\end{array}. \quad (30)$$

To determine the spectrum of representations of A_1 subduced from each of these representations of A_2, we merely apply the annihilation operator P_2:

$$P_2\ \square\square = \square\square + \square + \bullet \quad (31a)$$

$$P_2\ \begin{array}{|c|}\hline\ \\\hline\end{array}\!\!\begin{array}{|c|}\hline\ \\\hline\end{array} = \begin{array}{|c|}\hline\ \\\hline\end{array}\!\!\begin{array}{|c|}\hline\ \\\hline\end{array} + \square + \square\square + \square$$

$$\equiv \square + \bullet + \square\square + \square \quad (31b)$$

$$P_2\ \begin{array}{|c|c|}\hline\ & \\\hline\ & \\\hline\end{array} = \begin{array}{|c|c|}\hline\ & \\\hline\ & \\\hline\end{array} + \begin{array}{|c|c|}\hline\ & \\\hline\end{array} + \square\square$$

$$\equiv \bullet + \square + \square\square. \quad (31c)$$

Equations (31) tell us that, under the canonical restriction $SU(3) \downarrow SU(2)$, the 6 and $\bar{6}$ representations break up into a triplet, a doublet, and a singlet, while the octet breaks up into a triplet, two doublets, and a singlet.

IV. CONSTRUCTION OF WEIGHT SPACES

The examples of the last section show us that Tables I and II do not completely determine the structure of the representations (λ) of A in terms of the representations (λ') of the canonical subalgebra A'. For example, under the restriction $SU(3) \downarrow SU(2)$, the octet breaks up into two irreducible doublets and a four-dimensional representation which is further reducible into a singlet and a triplet (Fig. 2).

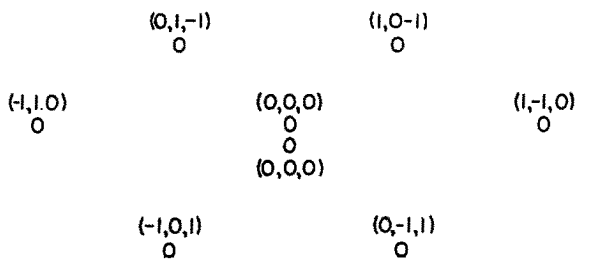


FIG. 2. Reduction of the representation $\begin{array}{|c|c|}\hline\ & \\\hline\ & \\\hline\end{array}$ of A_2 under the restriction to the algebra A_1 consisting of H_1, H_2, $E_{\pm(e_1-e_2)}$. Basis vectors with different m_3 values necessarily belong to different irreducible representations of A_1. The shift operators $E_{\pm(e_1-e_2)}$ can only connect basis vectors in the same horizontal row (fixed value of m_3).

Is it possible to extend Tables I and II in such a way that representations (λ) of A are completely determined in terms of the representations (λ') of A'?

To answer this important question, we look in more detail at the structure of the simple algebras A_n, B_n, C_n, D_n. The algebras A_n consist of $n+1$ mutually commuting generators

$$\mathbf{H} = (H_1, H_2, \cdots, H_n, H_{n+1}), \quad (32)$$

subject to the constraint

$$0 = H_1 + H_2 + \cdots + H_{n+1}, \quad (33)$$

together with the shift operators E_α:

$$\alpha = \mathbf{e}_i - \mathbf{e}_j, \quad 1 \leq i, \quad j \leq n+1. \quad (34)$$

The canonical subalgebra A_{n-1} of A_n contains n mutually commuting generators

$$h_i = H_i, \quad i = 1, 2, \cdots, n, \quad (32')$$

together with the shift operator f_α:

$$\alpha = \mathbf{e}_i - \mathbf{e}_j, \quad 1 \leq i, j \leq n. \quad (34')$$

Basis vectors for an arbitrary representation of A_n are described by $n+1$ numbers $|m_1 m_2 \cdots m_n m_{n+1}\rangle$. These states are eigenvalues of the diagonal operators $\mathbf{H}$:

$$H_i\,|m_1 m_2 \cdots m_i, \cdots, m_{n+1}\rangle = m_i\,|\mathbf{m}\rangle. \quad (35)$$

The m_i are subject to

$$m_1 + m_2 + \cdots + m_{n+1} = 0. \quad (33')$$

Basis vectors within the same irreducible subspace are connected by the shift operators E_α:

$$E_{e_i-e_j}\,|m_1, \cdots, m_i, \cdots, m_j, \cdots, m_{n+1}\rangle \cong |m_1, \cdots, m_i + 1, \cdots, m_j - 1, \cdots, m_{n+1}\rangle. \quad (36)$$

Since shift operators of the form $E_{e_i-e_{n+1}}$, $E_{e_{n+1}-e_i}$ are not contained in the canonical subgroup A_{n-1} of A_n, there are no shift operators in the subalgebra A_{n-1} which are capable of changing the value of m_{n+1}. We conclude, then, that bases $|\mathbf{m}\rangle$, $|\mathbf{m}'\rangle$ with $m_{n+1} \neq m'_{n+1}$ belong to different irreducible representations of A_{n-1}. Basis vectors with $m_{n+1} = m'_{n+1}$ may or may not belong to the same irreducible representations of A_{n-1}. Analogous statements hold for

$$D_n \downarrow D_{n-1}, \quad B_n \downarrow B_{n-1}, \quad C_n \downarrow C_{n-1}.$$

Representations of A_n can be described either by their dominant weights $\mathbf{M}$ or by the partition (λ) to which they correspond. There is a simple relation

between λ and $\mathbf{M}$:

$$M_i = \lambda_i - (n+1)^{-1}\sum_{j=1}^{n} j(\lambda_j - \lambda_{j+1}),$$
$$i = 1, 2, \cdots, n+1. \quad (37)$$

Therefore, the ϵ operators, which act on partitions, should be closely related to the shift operators E_α, which shift dominant weights. In expressions of the form (8), the annihilation operators act on a representation of A_n to yield a spectrum of representations of A_{n-1}. They do this by severing the sublattices of constant m_{n+1} value from each other. Therefore,

$$\epsilon_i \cong E_{\pm(e_i - e_{n+1})}. \quad (38)$$

Since the annihilation operators create smaller representations, they must act to decrease the dominant weights of (λ'). Thus, we make the association

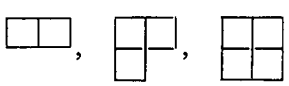
$$\epsilon_i \cong E_{-e_i + e_{n+1}}. \quad (39)$$

Equation (39) now allows us to solve the problem suggested at the beginning of this section: Whenever the annihilation operator ϵ_i acts on a partition, it shifts the M_{n+1} value up by one unit. All representations of A_{n-1} with the same value of M_{n+1}, under the restriction

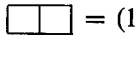
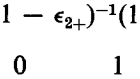
$$\{\lambda\}_{n+1} \xrightarrow{A_n \downarrow A_{n-1}} \{\lambda'\}_n = P_{n+}(\lambda)_n, \quad (12')$$

occur in the same representation sublattice of A_n. These representation sublattices are ordered according to M_{n+1} value.

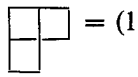
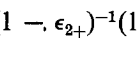
For example, let us determine the detailed structure of the representations



of $SU(3)$. The solution is as follows:

P_{2+}  $= (1-\epsilon_{2+})^{-1}(1-\epsilon_{1+})^{-1}$ 

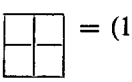
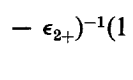
$$= \overset{0}{\square\square} + \overset{1}{\square} + \overset{2}{\bullet}, \quad (40a)$$

P_{2+}  $= (1-\epsilon_{2+})^{-1}(1-\epsilon_{1+})^{-1}$ 

$$= (1-\epsilon_{2+})^{-1}\left\{\overset{0}{\begin{smallmatrix}\square\square\\\square\end{smallmatrix}} + \overset{1}{\begin{smallmatrix}\square\\\square\end{smallmatrix}}\right\}$$

$$= \overset{0}{\begin{smallmatrix}\square\square\\\square\end{smallmatrix}} + \overset{1}{\square\square} + \overset{1}{\begin{smallmatrix}\square\\\square\end{smallmatrix}} + \overset{2}{\square} \quad (40b)$$

$$= \overset{0}{\square} + \{\overset{1}{\square\square} + \overset{2}{\bullet}\} + \overset{2}{\square},$$

P_{2+}  $= (1-\epsilon_{2+})^{-1}(1-\epsilon_{1+})^{-1}$ 

$$= \overset{0}{\begin{smallmatrix}\square\square\\\square\square\end{smallmatrix}} + \overset{1}{\begin{smallmatrix}\square\square\\\square\end{smallmatrix}} + \overset{2}{\square\square}$$

$$= \overset{0}{\bullet} + \overset{1}{\square} + \overset{2}{\square\square}, \quad (40c)$$

where the small integer over each partition describes the relative $M_{n+1} = M_3$ value. It is equal to the number of annihilation operators ϵ_i which have acted on the partition. In Fig. 3 we show how these representations of $SU(2)$ are "piled onto" each other to synthesize irreducible representations of $SU(3)$.

Let us also determine the general structure of the representations $\{\lambda_1, \lambda_2\}$ of $SU(3)$. The solution is as

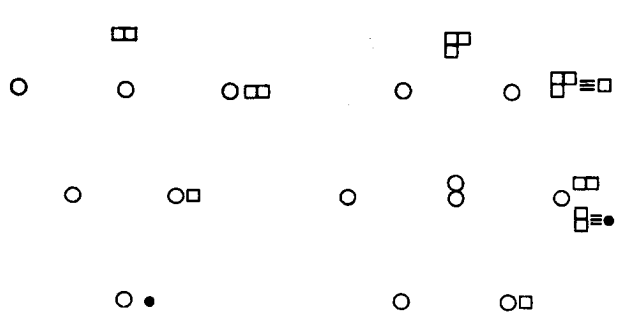
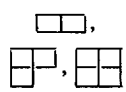


FIG. 3. Explicit construction of the weight spaces for the representations  of $A_2 \cong SU(3)$.

follows:

$$P_2\{\lambda_1,\lambda_2\} \xrightarrow{A_2\downarrow A_1} \begin{matrix} 0 & 1 & \lambda_1-\lambda_2 \\ \{\lambda_1,\lambda_2\}+\{\lambda_1-1,\lambda_2\}+\cdots+\{\lambda_2,\lambda_2\} \\ 1 & & \lambda_1-\lambda_2+1 \\ +\{\lambda_1,\lambda_2-1\}+\cdots+\{\lambda_2,\lambda_2-1\} \\ \vdots & & \vdots \\ \lambda_2 & & \lambda_1 \\ +\{\lambda_1,0\}+\cdots+\{\lambda_2,0\} \end{matrix} \tag{41}$$

$$\begin{matrix} \{\lambda_1-\lambda_2\} & & 0 \\ =\{\lambda_1-\lambda_2+1\}+\{\lambda_1-\lambda_2-1\} & & 1 \\ \vdots & & \vdots \\ \{\lambda_1-1\}+\{\lambda_1-3\}+\cdots+\{\lambda_1-2\lambda_2+3\}+\{\lambda_1-2\lambda_2+1\} & & \lambda_2-1 \\ \{\lambda_1\}+\{\lambda_1-2\}+\cdots+\{\lambda_1-2\lambda_2+2\}+\{\lambda_1-2\lambda_2\} & & \lambda_2 \\ \{\lambda_1-1\}+\{\lambda_1-3\}+\cdots+\{\lambda_1-2\lambda_2+3\}+\{\lambda-2\lambda_2+1\} & & \lambda_2+1 \\ \vdots & & \vdots \\ \{\lambda_2+1\}+\{\lambda_2-1\} & & \lambda_1-1 \\ \{\lambda_2\} & & \lambda_1 \end{matrix} \tag{42}$$

The general structure of $\{\lambda_1, \lambda_2\}$ is shown in Fig. 4.

The discussion of $A_n \downarrow A_{n-1}$ is, in some ways, simpler, in other ways, more difficult, than the discussion of the subalgebra restrictions $D_n \downarrow D_{n-1}$, $B_n \downarrow B_{n-1}$, $C_n \downarrow C_{n-1}$. It is simpler because we have to apply the annihilation operator P_n only once and also because we can make the unique association (39). It is more difficult because we are always operating in an n-dimensional hyperplane orthogonal to the vector

$$R = e_1 + e_2 + \cdots + e_{n+1}.$$

The relation between the partitions $(\boldsymbol{\lambda})$ and dominant weights $\mathbf{M}$, describing the irreducible representations of D_n, B_n, C_n, is simply

$$\boldsymbol{\lambda} = \mathbf{M}. \tag{43}$$

By arguments similar to those for A_n, we conclude

$$\epsilon_{i\pm} \cong E_{-e_1 \pm e_n}. \tag{44}$$

In terms of these annihilation operators, the representations $(\boldsymbol{\lambda}')$ subduced from $(\boldsymbol{\lambda})$, together with the

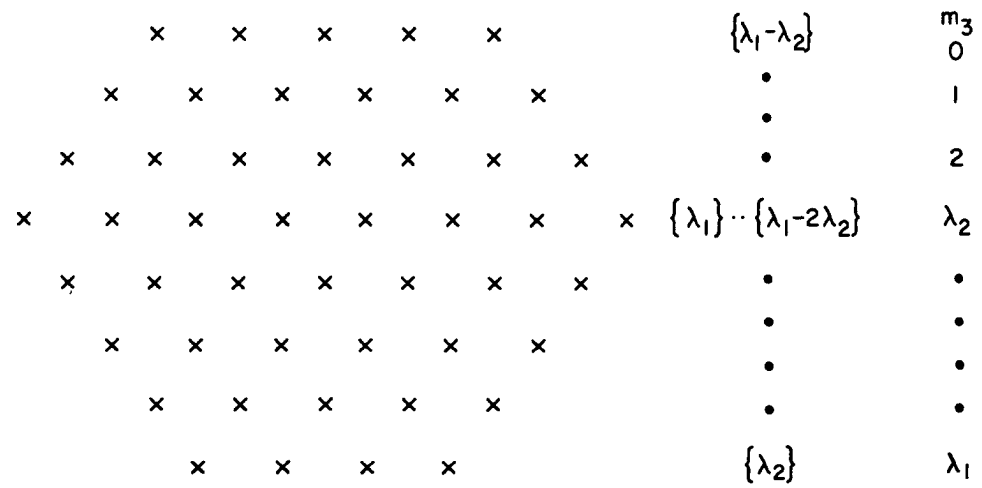


FIG. 4. General structure of the representation $\{\lambda_1, \lambda_2\}_3$ of A_2. From Eq. (42) we see the weight degeneracy increases by 1 as we proceed by hexagons toward the center. Once we reach a triangular pattern, and within it, the weight degeneracy remains constant.

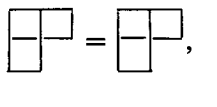
appropriate M_n ordering, are given by

$$(\boldsymbol{\lambda}) \to \sum_{\lambda'} (\lambda') = P_{n-} P_{n+} (\lambda) \tag{45}$$

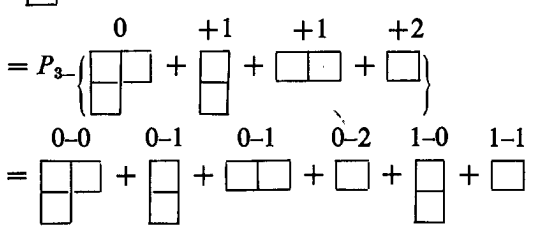
for each of the three cases $D_n \downarrow D_{n-1}$, $B_n \downarrow B_{n-1}$, $C_n \downarrow C_{n-1}$. For example, let us determine and construct the representations of D_3 which contain

 of D_2

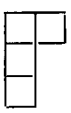
as their largest subduced representation. The solution is as follows:

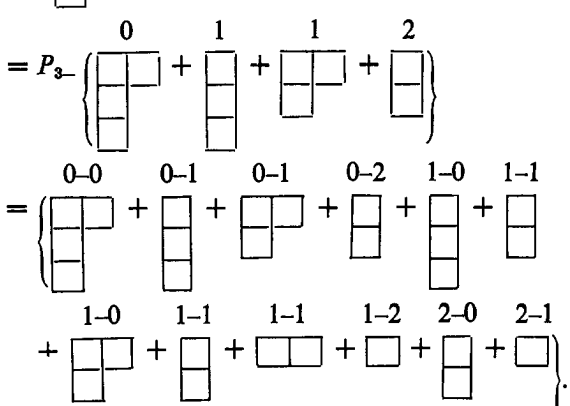
$(1 - \epsilon_3^+)^{-1}$ 

$P_{3-}P_{3+}$ 



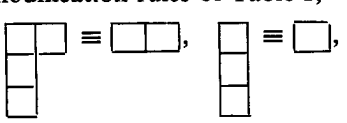
(46a)

$P_{3-}P_{3+}$ 

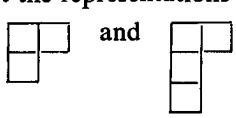


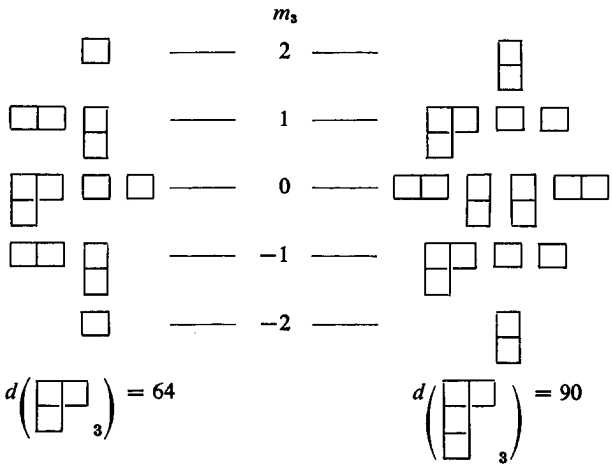
(46b)

Using the modification rules of Table I,



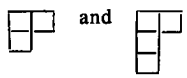
we can construct the representations

 and



(a) (b)

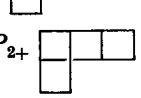
FIG. 5. These representations of D_2 are ordered as shown to construct the representations

 and

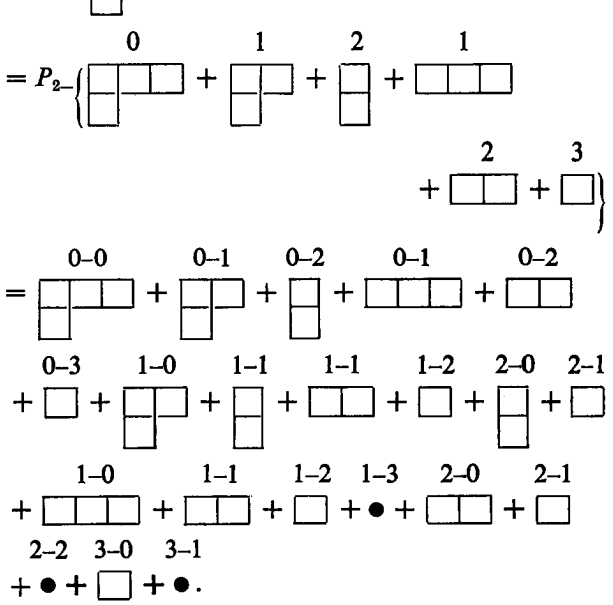
of D_3. The dimensionalities are

$d(\square_2) = 4$, $d\left(\begin{smallmatrix}\square\\\square\end{smallmatrix}_2\right) = 6$, $d(\square\square_2) = 9$, $d\left(\begin{smallmatrix}\square\square\\\square\end{smallmatrix}_2\right) = 16$.

of D_3, shown in Figs. 5 and 6. The representations of D_n with $\lambda_n = M_n \neq 0$ are not simple but compound, consisting of the two irreducible representations with dominant weights $(\lambda_1\lambda_2 \cdots \lambda_{n-1}, +\lambda_n)$ and $(\lambda_1\lambda_2 \cdots \lambda_{n-1}, -\lambda_n)$.

Finally, let us construct the irreducible representation  of C_2:

$P_{2-}P_{2+}$



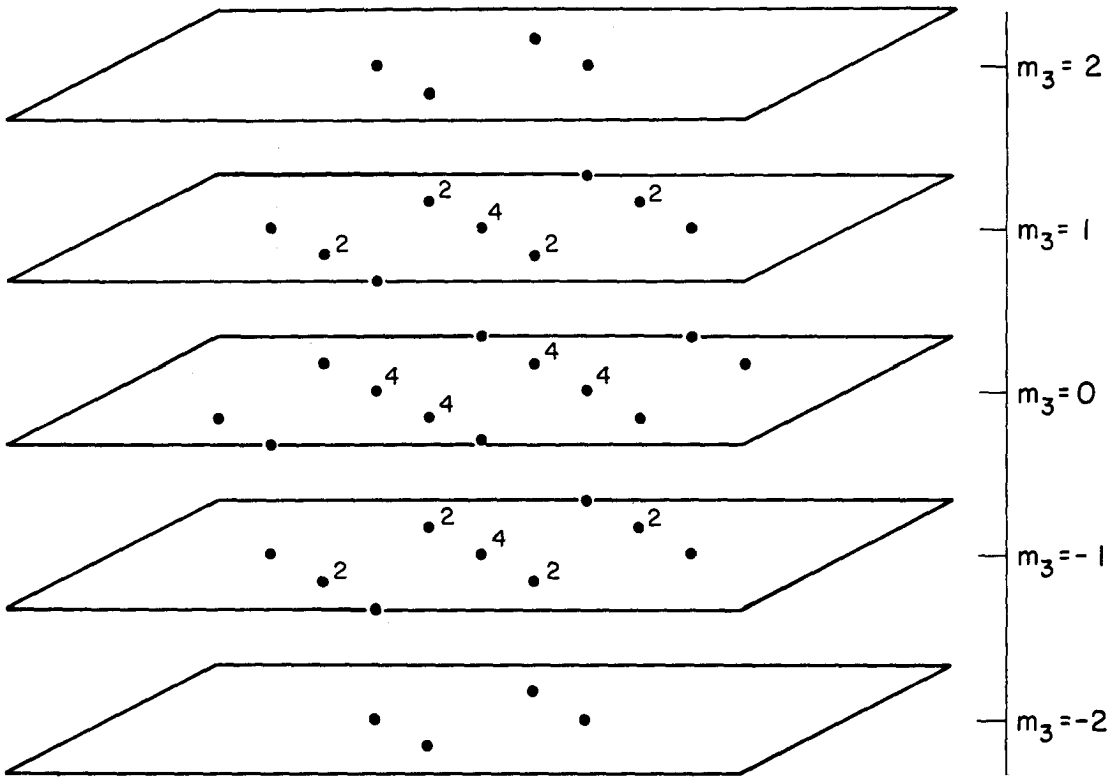


FIG. 6(a). Weight diagram for  of D_3.

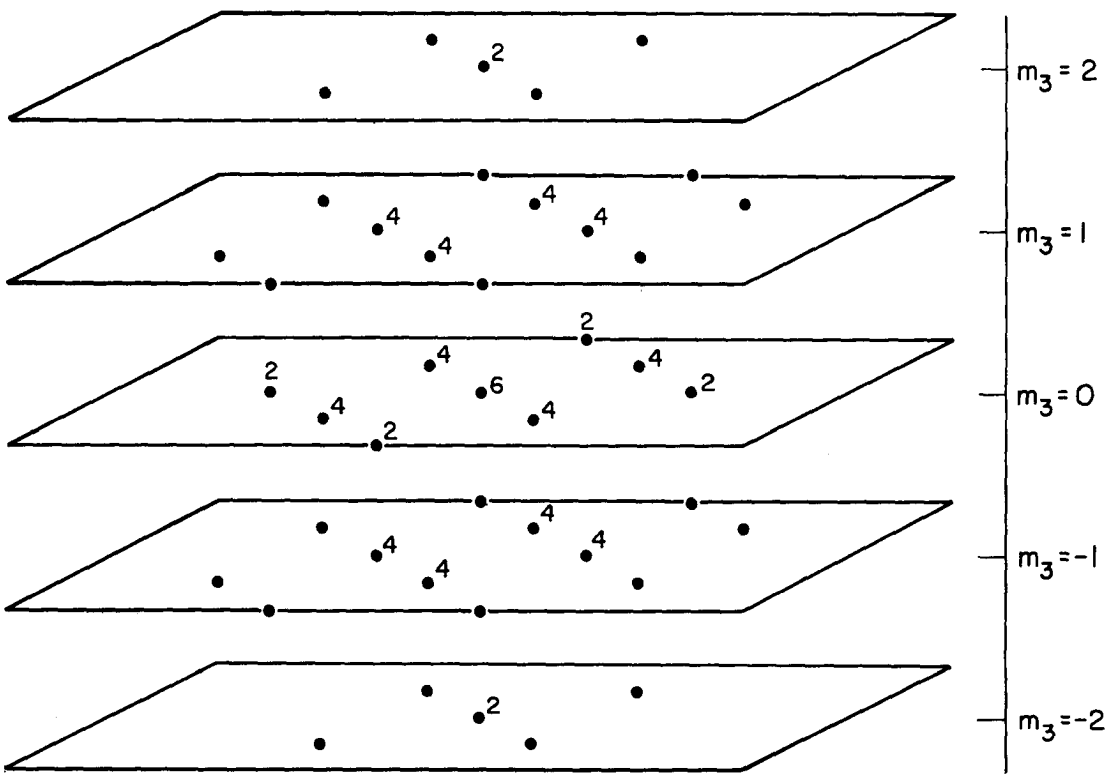


FIG. 6(b). Weight diagram for  of D_3.

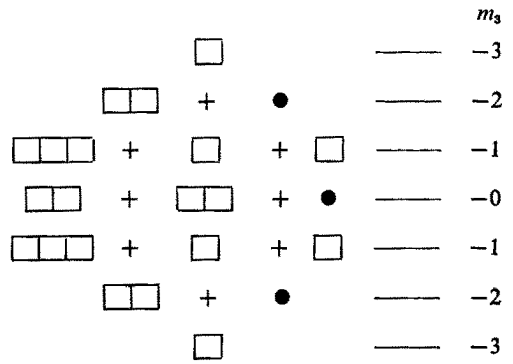


FIG. 7. Representations of C_1, with ordering, synthesized into the irreducible representation of C_2. A dimension check, using $d(\square) = 2$, $d(\square\square) = 3$, $d(\square\square\square) = 4$, gives

$$d\left(\begin{array}{|c|c|c|}\hline \\ \hline \\ \cline{1-1}\end{array}_{2}\right) = 35.$$

Using the modification rules of Table I, we reproduce this spectrum of representations of C_1 in Fig. 7 and use it to construct the weight space for

of C_2 in Fig. 8. Each of the basis vectors in any representation of C_2 is uniquely described by its eigenvalues under the subgroup $C_2 \downarrow C_1^{(1)} + C_1^{(2)}$, with the subgroups $C_1^{(1)}$ and $C_1^{(2)}$ consisting of the generators H_1, $E_{\pm 2e_1}$ and H_2, $E_{\pm 2e_2}$, respectively.

V. CONCLUSIONS

Algebraic techniques for determining the branching rules for the representations (λ) of the algebras A_n, B_n, C_n, D_n, under the restrictions $A_n \downarrow A_{n-1}$, $B_n \downarrow B_{n-1}$, $C_n \downarrow C_{n-1}$, $D_n \downarrow D_{n-1}$, have been described and used to construct the weight spaces for arbitrary representations (λ) of these algebras in terms of the weight spaces for (λ') of their subalgebras. Since the weight spaces for the representations of $A_1 = B_1 = C_1$ are well known, this allows us, in principle as well as

FIG. 8. Complete labeling for the basis states of of C_2.

in practice, to construct the weight-space diagrams for arbitrary representations $\boldsymbol{\lambda}$ of these algebras using bootstrap techniques. Equations (12′) and (45), together with those of Table II, may be regarded as the analog of the Frobenius reciprocity theorem[7] for Lie algebras.

[7] L. Jansen and M. Boon, *Theory of Finite Groups* (John Wiley & Sons, New York, 1967).

ACKNOWLEDGMENT

It is a pleasure to thank Dr. E. E. Lafon for bringing the Frobenius reciprocity theorem to my attention. It is this very beautiful theorem which convinced the author that an analog should exist for the classical simple Lie algebras.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Born Approximations in Statistical Mechanics

JEAN-PAUL MARCHAND
Department of Mathematics and Physics, University of Denver, Denver, Colorado 80210

(Received 15 April 1968)

There are various possibilities in defining Born approximations for the evolution of density matrices. They differ, however, from the ordinary Born approximation for pure states by the fact that they do not, in general, conserve the fundamental properties of physical states. We shall describe these possibilities geometrically in the so-called Liouville space of Hilbert–Schmidt operators and investigate their significance in nonequilibrium dynamics. The results admit of an analogous interpretation in classical statistical mechanics.

I. INTRODUCTION

The mathematical concept of Born approximation is as follows: In a Hilbert space two continuous groups $\{U_t^{(0)}\}$ and $\{U_t\}$ of unitary operators are given with generators $H^{(0)}$, H and resolvents $R^{(0)}(z)$, $R(z)$. The Laplace transform of the second resolvent equation

$$R(z) = R^{(0)}(z) - R^{(0)}(z)(H - H^{(0)})R(z)$$

then yields the convolution integral equation

$$U_t = U_t^{(0)} - i\int_0^t d\tau U_{t-\tau}^{(0)}(H - H^{(0)})U_\tau,$$

whose iterative solution

$$U_t^{(\nu)} = U_t^{(0)} - i\int_0^t d\tau U_{t-\tau}^{(0)}(H - H^{(0)})U_\tau^{(\nu-1)} \quad (1)$$

defines the Born series $\{U_t^{(\nu)}\}$.

In physics the Hilbert space is interpreted as the space of the physical states, the two families of unitary operators as their Schrödinger time propagators, and the generators as the Hamiltonians for the free and the "total" evolution, respectively. This can be done in two different ways: In *ordinary quantum mechanics* the elements ψ of the complex Hilbert space $\mathcal{H}$ correspond to *pure states*, while in *statistical mechanics* we may introduce the so-called Liouville space $\mathfrak{L}$ of Hilbert–Schmidt operators in which *all states* ρ, pure and mixed, are represented on an equal basis as a convex subset.

In both interpretations the Schrödinger evolution of the states ψ (respectively, ρ) is induced by groups of unitary operators U_t in $\mathcal{H}$ (respectively, $\mathfrak{U}_t$ in $\mathfrak{L}$) and in both cases a Born series may be set up formally by (1). We shall call the Born series so obtained in $\mathfrak{L}$ "statistical" in order to distinguish it from the "ordinary" series in $\mathcal{H}$. This statistical series has, for instance, been used by Kubo (see Appendix A).

Besides these two fundamental interpretations of (1), one applying to the space $\mathcal{H}$ of pure states and the other to the Liouville space $\mathfrak{L}$ of the density matrices, one might also consider a true extension of the ordinary Born approximation from $\mathcal{H}$ to $\mathfrak{L}$. This intermediate possibility is also briefly discussed in Sec. III.

Although these different Born series have some basic mathematical convergence properties in common, their physical significance is entirely different. While the ordinary Born propagators map pure states into pure states, the statistical propagators not only change the entropy of the states, but may even destroy their positivity for sufficiently large times.

The natural question then arises whether the statistical Born approximations give a correct information about the approach to equilibrium in statistical

in practice, to construct the weight-space diagrams for arbitrary representations $\boldsymbol{\lambda}$ of these algebras using bootstrap techniques. Equations (12′) and (45), together with those of Table II, may be regarded as the analog of the Frobenius reciprocity theorem[7] for Lie algebras.

[7] L. Jansen and M. Boon, *Theory of Finite Groups* (John Wiley & Sons, New York, 1967).

ACKNOWLEDGMENT

It is a pleasure to thank Dr. E. E. Lafon for bringing the Frobenius reciprocity theorem to my attention. It is this very beautiful theorem which convinced the author that an analog should exist for the classical simple Lie algebras.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Born Approximations in Statistical Mechanics

JEAN-PAUL MARCHAND
Department of Mathematics and Physics, University of Denver, Denver, Colorado 80210

(Received 15 April 1968)

There are various possibilities in defining Born approximations for the evolution of density matrices. They differ, however, from the ordinary Born approximation for pure states by the fact that they do not, in general, conserve the fundamental properties of physical states. We shall describe these possibilities geometrically in the so-called Liouville space of Hilbert–Schmidt operators and investigate their significance in nonequilibrium dynamics. The results admit of an analogous interpretation in classical statistical mechanics.

I. INTRODUCTION

The mathematical concept of Born approximation is as follows: In a Hilbert space two continuous groups $\{U_t^{(0)}\}$ and $\{U_t\}$ of unitary operators are given with generators $H^{(0)}$, H and resolvents $R^{(0)}(z)$, $R(z)$. The Laplace transform of the second resolvent equation

$$R(z) = R^{(0)}(z) - R^{(0)}(z)(H - H^{(0)})R(z)$$

then yields the convolution integral equation

$$U_t = U_t^{(0)} - i\int_0^t d\tau U_{t-\tau}^{(0)}(H - H^{(0)})U_\tau,$$

whose iterative solution

$$U_t^{(\nu)} = U_t^{(0)} - i\int_0^t d\tau U_{t-\tau}^{(0)}(H - H^{(0)})U_\tau^{(\nu-1)} \quad (1)$$

defines the Born series $\{U_t^{(\nu)}\}$.

In physics the Hilbert space is interpreted as the space of the physical states, the two families of unitary operators as their Schrödinger time propagators, and the generators as the Hamiltonians for the free and the "total" evolution, respectively. This can be done in two different ways: In *ordinary quantum mechanics* the elements ψ of the complex Hilbert space $\mathcal{H}$ correspond to *pure states*, while in *statistical mechanics* we may introduce the so-called Liouville space $\mathfrak{L}$ of Hilbert–Schmidt operators in which *all states* ρ, pure and mixed, are represented on an equal basis as a convex subset.

In both interpretations the Schrödinger evolution of the states ψ (respectively, ρ) is induced by groups of unitary operators U_t in $\mathcal{H}$ (respectively, $\mathfrak{U}_t$ in $\mathfrak{L}$) and in both cases a Born series may be set up formally by (1). We shall call the Born series so obtained in $\mathfrak{L}$ "statistical" in order to distinguish it from the "ordinary" series in $\mathcal{H}$. This statistical series has, for instance, been used by Kubo (see Appendix A).

Besides these two fundamental interpretations of (1), one applying to the space $\mathcal{H}$ of pure states and the other to the Liouville space $\mathfrak{L}$ of the density matrices, one might also consider a true extension of the ordinary Born approximation from $\mathcal{H}$ to $\mathfrak{L}$. This intermediate possibility is also briefly discussed in Sec. III.

Although these different Born series have some basic mathematical convergence properties in common, their physical significance is entirely different. While the ordinary Born propagators map pure states into pure states, the statistical propagators not only change the entropy of the states, but may even destroy their positivity for sufficiently large times.

The natural question then arises whether the statistical Born approximations give a correct information about the approach to equilibrium in statistical

mechanics. More specifically, it may be asked whether certain typical oscillations found in the expectation values of "observables of interest" could be due to the approximation procedure rather than to physics. One of our aims is to answer this question.

Section II gives a brief exposition of the geometry of states in Liouville space; in Sec. III, we shall compare their motions induced by the various kinds of Born approximations; in Sec. IV, the first statistical Born approximation is calculated explicitly in the simple example of a spin $\frac{1}{2}$ and the result compared to the exact solution; and in Sec. V, we consider this system as part of a statistical ensemble in order to obtain some conclusions about the suitability of this approximation in nonequilibrium dynamics.

Our treatment in this paper will be quantum mechanical. But we should note that the results could be interpreted in classical statistical mechanics as well (Appendix A). Using the translation code

density matrix $\leftrightarrow$ density in phase space,
trace operation $\leftrightarrow$ integration over phase space,
commutators $\leftrightarrow$ Poisson brackets,

we conclude, for example, that the analog to the statistical Born approximation in classical mechanics yields a stochastic process in which the distributions conserve their normalization throughout time, but not the positivity of their densities.

II. DESCRIPTION OF STATES IN LIOUVILLE SPACE

A. Definition of States as Elements of a Liouville Space[1]

General states in ordinary quantum mechanics are defined as positive σ-additive normalized functionals f on the atomic lattice of the projectors in a separable Hilbert space $\mathcal{H}$. According to Gleason, these functionals are in a one-to-one correspondence to the so-called density operators ρ in $\mathcal{H}$ (dimension $\mathcal{H} > 2$):

$$\rho\colon \operatorname{Tr}\rho = 1; \quad \rho^* = \rho; \quad \rho > 0;$$
$$f(A) = \operatorname{Tr}\rho A \ \forall \text{ self-adj. } A \text{ in } \mathcal{H}. \quad (2)$$

Since the density operators are Hilbert–Schmidt, they form a subset $\mathfrak{L}'$ of the real Hilbert space $\mathfrak{L}$ of all self-adjoint operators A for which $\operatorname{Tr}A^2 < \infty$, and in which the scalar product is defined by $(\vec{A}, \vec{B}) = \operatorname{Tr}AB$. (We shall endow the operators A with an arrow if they are to be considered as elements of $\mathfrak{L}$.)

[1] For a more complete exposition see G. G. Emch, Helv. Phys. Acta 37, 532 (1964).

B. Geometrical Characterization of States in $\mathfrak{L}$

We define the following objects in $\mathfrak{L}$:

the hyperplane[2]

$$\Upsilon = \{\vec{\rho} \mid (\vec{I}, \vec{\rho}) = \operatorname{Tr}\rho = 1\},$$

the sphere

$$\mathfrak{S} = \{\vec{\rho} \mid \|\vec{\rho}\|^2 = \operatorname{Tr}\rho^2 \leq 1\},$$

their intersection

$$\Sigma = \Upsilon \cap \mathfrak{S} = \{\vec{\rho} \mid \operatorname{Tr}\rho^2 \leq \operatorname{Tr}\rho = 1\},$$

its boundary

$$\bar{\Sigma} = \{\vec{\rho} \mid \operatorname{Tr}\rho^2 = \operatorname{Tr}\rho = 1\}.$$

Furthermore, we note that the geometrical locus of convex combinations $\vec{\rho} = \lambda\vec{\rho_1} + (1-\lambda)\vec{\rho_2}$ $(0 \leq \lambda \leq 1)$ of vectors $\vec{\rho_1}$, $\vec{\rho_2}$ is the segment of the straight line connecting $\vec{\rho_1}$ and $\vec{\rho_2}$. From (2) we now conclude:

(1) *The states are in* Σ.

Proof: $\operatorname{Tr}\rho = 1$, $\rho > 0$ implies that ρ has a discrete spectrum $0 \leq \lambda_i \leq 1$. Hence

$$\operatorname{Tr}\rho^2 = \sum_i \lambda_i^2 \leq \sum_i \lambda_i = \operatorname{Tr}\rho = 1.$$

(2) *The states form a convex set.*

Proof: Let $\rho_i > 0$, $\operatorname{Tr}\rho_i = 1$. Then $(\varphi, \rho_i\varphi) > 0\ \forall$ $\varphi \in \mathcal{H}$. For a convex combination we have $0 \leq \lambda_i \leq 1$, $\sum_i \lambda_i = 1$. So

$$\left(\varphi, \sum_i \lambda_i\rho_i\varphi\right) = \sum_i \lambda_i(\varphi, \rho_i\varphi) > 0$$

and

$$\operatorname{Tr}\sum \lambda_i\rho_i = \sum_i \lambda_i \operatorname{Tr}\rho_i = \sum_i \lambda_i = 1.$$

(3) *All states are a convex combination of states on* $\bar{\Sigma}$.

Proof: Let ψ_i denote a complete orthonormal set of eigenstates of ρ with eigenvalues λ_i and P_i the corresponding one-dimensional projectors. Then $\vec{\rho} = \sum_i \lambda_i\vec{P_i}$ is the desired convex combination, since $\vec{P_i}$ are states and lie on $\bar{\Sigma}$.

(4) *All states on* $\bar{\Sigma}$ *are extremal* (= *pure*) *states.*

Proof: Otherwise they would lie inside a segment, one of whose end points is outside Σ and cannot represent a state according to (1).

(5) *All extremal states are on* $\bar{\Sigma}$.

Proof: Immediate consequence of (3) and (4).

(6) *Not all elements of* $\bar{\Sigma}$ *are states unless* $\dim\mathcal{H} = 2$.

[2] Here we made a trivial abuse in notation since the unit operator I in $\mathcal{H}$ is not Hilbert–Schmidt unless $\mathcal{H}$ is finite dimensional.

Proof: $\mathrm{Tr}\,\rho^2 = \mathrm{Tr}\,\rho = 1$ implies positivity if and only if $\dim \mathcal{H} = 2$.

If:

$$\sum_i \lambda_i^2 = \sum_i \lambda_i = 1$$

$$\Rightarrow 1 = \lambda_1^2 + \lambda_2^2 = (\lambda_1 + \lambda_2)^2 - 2\lambda_1\lambda_2 = 1 - 2\lambda_1\lambda_2$$

$$\Rightarrow \lambda_1\lambda_2 = 0$$

$$\Rightarrow \lambda_1 = 0 \text{ or } 1, \quad \lambda_2 = 1 \quad \text{or} \quad 0 \Rightarrow \rho > 0.$$

Only if: the matrix

$$\rho = \begin{pmatrix} \frac{1}{2} & & \\ & \frac{1}{4}(1+\sqrt{5}) & \\ & & \frac{1}{4}(1-\sqrt{5}) \end{pmatrix}$$

satisfies $\mathrm{Tr}\,\rho^2 = \mathrm{Tr}\,\rho = 1$, but ρ is not positive.

C. Schrödinger Evolution of States in $\mathfrak{L}$

Let H be the Hamiltonian of the system. Then the Schrödinger equation for density matrices reads in integral form

$$\rho_t = U_t\rho_0 U_t^{-1}; \quad U_t = \exp(-iHt), \tag{3a}$$

in differential form

$$\frac{d}{dt}\rho_t = -i[H, \rho], \tag{3b}$$

where $\{U_t\}$ is a group of unitary operators in $\mathcal{H}$. In Liouville space we introduce the operators $\mathfrak{H}\rho^{\rightarrow} = [H, \rho]^{\rightarrow}$ and $\mathfrak{U}_t\rho^{\rightarrow} = (U_t\rho U_t^{-1})^{\rightarrow}$.[3] So (3) becomes

$$\rho_t^{\rightarrow} = \mathfrak{U}_t\rho_0^{\rightarrow}; \quad \mathfrak{U}_t = \exp(-i\mathfrak{H}t); \quad \frac{d}{dt}\rho_t^{\rightarrow} = -i\mathfrak{H}\rho_t^{\rightarrow}. \tag{3'}$$

For the characterization of the evolution of the system in $\mathfrak{L}$ we furthermore introduce the *entropy of state* ρ *by*

$$S = -k \ln \mathrm{Tr}\,\rho^2 = -k \ln \|\rho^{\rightarrow}\|^2.$$

Then we have:

(7) $\mathfrak{U}_t$ *transforms states into states and leaves their spectrum invariant.*

Proof: $U_t\rho U_t^{-1}$ is unitarily equivalent to ρ and thus leaves the spectrum and all properties characterizing a state invariant.

(8) $\{\mathfrak{U}_t\}$ *are rotations in* $\mathfrak{L}$ [*leaving* Υ *invariant according to* (7)].

Proof: The scalar products (lengths and angles) of the vectors are conserved:

$$(\mathfrak{U}_t\rho_1^{\rightarrow}, \mathfrak{U}_t\rho_2^{\rightarrow}) = \mathrm{Tr}\; U_t\rho_1 U_t^{-1} U_t\rho_2 U_t^{-1} = \mathrm{Tr}\;\rho_1\rho_2 = (\rho_1^{\rightarrow}, \rho_2^{\rightarrow}).$$

[3] If $\mathfrak{L}$ is infinite dimensional, $\mathfrak{H}$ may not be defined everywhere.

(9) *All rotations in* $\mathfrak{L}$ *conserve the entropy.*

Proof: Immediate consequence of the definition and (8).

(10) *Not all rotations in* $\mathfrak{L}$ *are Schrödingerian unless* $\dim \mathcal{H} = 2$ *or the states are pure.*

Proof: For any two points on a sphere there exists a rotation transforming one into the other. But

$$\|\rho^{\rightarrow}\|^2 = \mathrm{Tr}\,\rho^2 = \sum_i \lambda_i^2 = a \leq 1, \quad \mathrm{Tr}\,\rho = \sum_i \lambda_i = 1$$

does not uniquely define $\lambda_i > 0$ unless $\dim \mathcal{H} = 2$. So a rotation may change the spectrum of ρ in contradiction to (7). On the other hand, any pure state is on $\bar{\Sigma}$, any state on $\bar{\Sigma}$ is a projector, and all one-dimensional projectors are unitarily equivalent.

(11) *Not all rotations transform states into states, unless* $\dim \mathcal{H} = 2$.

Proof: According to (6), the spheres do not contain only states unless $\dim \mathcal{H} = 2$.[4]

III. BORN APPROXIMATIONS

Let us split the total Hamiltonian $\mathfrak{H}$ in $\mathfrak{L}$ into a free part $\mathfrak{H}^{(0)}$ and an interacting part $\mathfrak{V}$. In Liouville space both the free and the physical motion lead to ordinary rotations with rotation matrices $\mathfrak{U}_t^{(0)} = \exp(-i\mathfrak{H}^{(0)}t)$ and $\mathfrak{U}_t = \exp(-i\mathfrak{H}t)$, respectively. Both rotations leave a hyperplane $\mathfrak{E}^{(0)}$ and $\mathfrak{E}$ in Υ invariant (in the 3-dimensional case these are simply the rotation axes $n^{\rightarrow(0)}$ and $n^{\rightarrow}$); these fixed axes differ from each other except in the trivial case when $\mathfrak{H}$ commutes with $\mathfrak{H}^{(0)}$.

We shall now give a global description in $\mathfrak{L}$ of the Born approximations mentioned in the Introduction.

A. Ordinary Born Approximation for Pure States

The ordinary Born series is originally defined in the Hilbert space $\mathcal{H}$ of pure states by the iterative formula (1). The $\{U_t^{(\nu)}\}$ are not unitary in $\mathcal{H}$, and they do not form a semigroup in t except, of course, for $\nu = 0$ (free motion) and $\nu = \infty$ (physical motion), if the series converges. We may now define the corresponding Liouville operators

$${}^{O}\mathfrak{U}_t^{(\nu)}P_{\psi}^{\rightarrow} = P_{U_t^{(\nu)}\psi}^{\rightarrow}, \tag{4}$$

where "O" denotes "ordinary," P_{ψ} are the 1-dimensional projectors on ψ, and $U_t^{(\nu)}$ is defined by (1).

[4] It is not difficult to locate on the spheres in $\mathfrak{L}$ the subsets of elements with identical spectrum. This problem is, in fact, equivalent to characterizing among all rotations in $\mathfrak{L}$ those which are implementable by unitary operators in $\mathcal{H}$. The particularity of the 2-dimensional space in this respect is due to the fact that the dimension of the sphere $\Sigma = \Upsilon \cap \mathfrak{S}$ (three) equals the number of generators of rotations in Σ.

The operators ${}^{0}\mathfrak{U}_t^{(\nu)}$ map the states of $\bar{\Sigma}$ (= pure states) into themselves, thus conserving their norm and entropy. But they are, in general, not rotations, since they do not conserve their angles. In fact, suppose $(P_\varphi^{\rightarrow}, P_\psi^{\rightarrow}) = 0$. It is easy to show that this is true if and only if $(\varphi, \psi) = 0$. But, since $U_t^{(\nu)}$ is not unitary in $\mathcal{H}$, this does not imply that $(U_t^{(\nu)}\varphi, U_t^{(\nu)}\psi)$, and thus $(P^{\rightarrow}_{U_t^{(\nu)}\varphi}, P^{\rightarrow}_{U_t^{(\nu)}\psi})$, vanishes.

In a slightly different fashion we could also start from the remark that ${}^{0}\mathfrak{U}_t^{(\nu)}$ is not implementable by unitary operators in $\mathcal{H}$, although all 1-dimensional projectors are unitarily equivalent. In fact, we have $P_{U_t^{(\nu)}\varphi} = U_\varphi P_\varphi U_\varphi^{-1}$ and $P_{U_t^{(\nu)}\psi} = U_\psi P_\psi U_\psi^{-1}$, where both U_φ and U_ψ are unitary, but in general different. From this we immediately conclude that $\mathfrak{U}_t^{(\nu)}$ conserves the norm of $P_\psi^{\rightarrow}$ but not its scalar product with $P_\varphi^{\rightarrow}$. So we have:

The ordinary Born approximation for pure states corresponds in $\mathfrak{L}$ *to a family of operators* $\{\mathfrak{U}_t^{(\nu)}\}$ *defined by Eqs.* (4) *and* (1), *which maps the states on* $\bar{\Sigma}$ (= *pure states*) *into themselves, conserving their norm* (*entropy*) *but not their angles* (*transition probabilities*).

B. Ordinary Born Approximations for Arbitrary States

There is a straightforward extension of the ordinary Born approximation from pure states on $\bar{\Sigma}$ to arbitrary states in Σ by

$$ {}^{0}\mathfrak{U}_t^{(\nu)}\rho^{\rightarrow} = U_t^{(\nu)}\rho^{\rightarrow}U_t^{(\nu)\dagger}/\mathrm{Tr}\,(U_t^{(\nu)}\rho U_t^{(\nu)\dagger}) \qquad (5) $$

since, for pure states

$$ \begin{aligned} (U_t^{(\nu)})_{ik}(P_\psi)_{kl}(U_t^{(\nu)\dagger})_{lm} &= (U_t^{(\nu)})_{ik}\psi_k\psi_l^*(U_t^{(\nu)})_{ml}^* \\ &= (U_t^{(\nu)}\psi)_m(U_t^{(\nu)}\psi)_i \\ &= (P_{U^{(\nu)}\psi})_{im}\cdot \|U_t^{(\nu)}\psi\|_{\mathcal{H}}^2 \end{aligned} $$

and

$$ \mathrm{Tr}\,(U_t^{(\nu)}P_\psi U_t^{(\nu)\dagger}) = (U_t^{(\nu)}\psi)_i^*(U_t^{(\nu)}\psi)_i = \|U_t^{(\nu)}\psi\|_{\mathcal{H}}^2. $$

The transformations ${}^{0}\mathfrak{U}_t^{(\nu)}$ do not, however, conserve the norm of mixed states since

$$ \mathrm{Tr}\,U_t^{(\nu)}\rho U_t^{(\nu)\dagger}U_t^{(\nu)}\rho U_t^{(\nu)} \neq \mathrm{Tr}\,\rho^2 $$

unless $U_t^{(\nu)}$ is unitary. In particular, the spectrum of these states (and possibly even its positivity) is not conserved. Hence:

The extended ordinary Born approximation $\{{}^{0}\mathfrak{U}_t^{(\nu)}\}$ *defined by Eqs.* (5) *and* (1) *does not, in general, conserve the norm* (*entropy*) *of states* (*and possibly not even their positivity*) *except for the states on the boundary* $\bar{\Sigma}$ (*pure states*) *which are mapped into themselves.*[5]

C. Statistical Born Approximation

We now consider a totally different type of a Born approximation for density matrices where the idea of extending the ordinary approximation for pure states is abandoned and where we are guided instead by the formal analogy between the Hilbert spaces $\mathcal{H}$ and $\mathfrak{L}$.

The statistical Born series is the formal analog of (1) in the space $\mathfrak{L}$ of the density matrices, i.e.,

$$ \mathfrak{U}_t^{(\nu)} = \mathfrak{U}_t^{(0)}\left(I - i\int_0^t d\tau\,\mathfrak{U}_{-\tau}^{(0)}(\mathfrak{H} - \mathfrak{H}^{(0)})\mathfrak{U}_t^{(\nu-1)}\right). \qquad (6) $$

This definition has been used by Kubo[6] (see Appendix A), and we shall now discuss it more thoroughly here and in the following sections.

The statement that the ordinary Born approximation $U_t^{(\nu)}$ is not unitary in $\mathcal{H}$ implies immediately that *the statistical Born approximation* $\{\mathfrak{U}_t^{(\nu)}\}$ *defined by* (6) *is not a rotation in* $\mathfrak{L}$ *and thus does not conserve the norm* (*entropy*) *of the states.*

In order to obtain a more detailed qualitative picture of this nonconservation of the norm, let us restrict ourselves to the first Born approximation

$$ \mathfrak{U}_t^{(1)} = \mathfrak{U}_t^{(0)}\left(I - i\int_0^t d\tau\,\mathfrak{U}_{-\tau}^{(0)}(\mathfrak{H} - \mathfrak{H}^{(0)})\mathfrak{U}_\tau^{(0)}\right). $$

In the trivial case where the interaction $\mathfrak{V} = \mathfrak{H} - \mathfrak{H}^{(0)}$ commutes with $\mathfrak{H}^{(0)}$, the integral reduces to $(t\mathfrak{V})$, i.e., to a term which increases linearly with time. The other extreme, in which the integral is periodic in t, arises if $\mathfrak{V}$ generates a rotation perpendicular to $\mathfrak{U}_t^{(0)}$. So we might expect in the general case a Born approximation of the form

$$ \mathfrak{U}_t^{(1)} = \mathfrak{U}_t^{(0)}(I - it\mathfrak{V}' + \text{periodic term in } t), $$

where $\mathfrak{V}'$ is the maximal part of $\mathfrak{V}$ commuting with $\mathfrak{H}^{(0)}$.

This guess is confirmed by the concrete example of a spin system treated in Sec. IV, in which it turns out, moreover, that $\mathfrak{V}'$ is a multiple of $\mathfrak{H}^{(0)}$, so that the first Born approximation reads

$$ \mathfrak{U}_t^{(1)} = \mathfrak{U}_t^{(0)}\left(I + it(1-\lambda)\mathfrak{H}^{(0)} - i\int_0^t d\tau\,\mathfrak{U}_{-\tau}^{(0)}(\mathfrak{H} - \lambda\mathfrak{H}^{(0)})\mathfrak{U}_\tau^{(0)}\right). \qquad (6') $$

[5] One might think to replace definition (5) by

$$ \mathfrak{U}_t^{(\nu)}\rho^{\rightarrow} = (U_t^{(\nu)}\rho U_t^{(\nu)-1})^{\rightarrow} $$

in order to preserve the norm. This is worse, however, since now the hermiticity of the states is destroyed and the propagators $\mathfrak{U}_t^{(\nu)}$ cease to be operators in the real Liouville space $\mathfrak{L}$.

[6] R. Kubo, in *Lectures in Theoretical Physics, Boulder, Colorado, 1958*, W. E. Brittin and L. G. Dunham, Eds. (Interscience Publ. Co., New York, 1959), Vol. I, p. 138.

Whenever the interaction $\mathfrak{V}$ can be decomposed in this way, the following more stringent statements hold:

The statistical first Born approximation $\mathfrak{U}_t^{(1)}$ defined by Eq. (6′) is not a rotation in $\mathfrak{L}$; the positivity of $\mathfrak{U}_t^{(1)}\rho_0^{\rightarrow}$ is not guaranteed for all times; in fact, for large t the norm $\|U_t^{(1)}\rho_0^{\rightarrow}\|^2$ becomes arbitrarily large, unless the initial state ρ_0 commutes with $H^{(0)}$, in which case the linear term in (6′) vanishes and the entropy becomes a periodic function of t.

We also remark that the propagator describing the evolution of the diagonal part of the density matrix in the basis in which $H^{(0)}$ is diagonal has no linear terms in t. In fact, if we introduce the "coarse-graining" operator $\mathfrak{D}$ in $\mathfrak{L}$ which projects the states on their diagonal part, it is easily verified that $\mathfrak{H}^{(0)}\mathfrak{D} = 0$.

Let us make this now more specific in the example of a spin-$\frac{1}{2}$ system.

IV. STATISTICAL BORN APPROXIMATION IN THE EXAMPLE OF SPIN $\frac{1}{2}$

The $\frac{1}{2}$-integer spin is described in a 2-dimensional vector space $\mathcal{H}$ in which the two eigenstates ψ^+ and ψ^- of the spin in z direction may serve as a basis. The corresponding Liouville space $\mathfrak{L}$ is 4-dimensional and we may introduce the orthonormal basis

$$V_0^{\rightarrow} = 2^{-\frac{1}{2}} I_{\mathcal{H}}; \quad V_i^{\rightarrow} = 2^{-\frac{1}{2}}\sigma_i, \quad i = 1, 2, 3,$$

where σ_i are the three Pauli matrices. The hyperplane Υ of trace-one operators is defined by

$$\Upsilon = \{\rho^{\rightarrow} \mid (V_0^{\rightarrow}, \rho^{\rightarrow}) = 2^{-\frac{1}{2}} \operatorname{Tr} \rho = 2^{-\frac{1}{2}}\},$$

i.e., by the elements of $\mathfrak{L}$ with identical zero-component $2^{-\frac{1}{2}}$. So we may disregard the zero-component altogether as long as we consider only motions in Υ, and we may work in the space spanned by the three traceless Pauli matrices.

From the 3-dimensional plane Υ the 4-dimensional sphere $\|\rho^{\rightarrow}\|^2 \leq 1$ cuts out the 3-dimensional hypersphere Σ with radius $2^{-\frac{1}{2}}$, and, according to Sec. II, we have the particularly simple situation that

(1) all elements of Σ are physical states[7];

(2) the boundary $\bar{\Sigma}$ contains exactly the pure states;

(3) any trace-conserving rotation in $\mathfrak{L}$ is implementable by unitary operators in $\mathcal{H}$ and thus leaves the spectrum of ρ invariant. (See Appendix B.)

A constant magnetic field may now serve "to define a z direction," i.e., to create a free-energy split in the levels of ψ^+ and ψ^-. The corresponding free Hamiltonian is $H^{(0)} = \frac{1}{2}\alpha\sigma_3$ with unitary group $U_t^{(0)} = \exp(-\frac{1}{2}i\alpha\sigma_3 t)$. The most general form for the total Hamiltonian is $H = \frac{1}{2}\alpha n_i\sigma_i$ and its unitary group $U_t = \exp(-\frac{1}{2}i\alpha n_i\sigma_i t)$. The corresponding Liouville operators in Υ are

$$\left.\begin{aligned} \mathfrak{H}_{ik}^{(0)} &= i\alpha\epsilon_{ik3}, \\ (\mathfrak{U}_t^{(0)})_{ik} &= \cos\alpha t \cdot \delta_{ik} \\ &\quad + (1 - \cos\alpha t)\delta_{i3}\delta_{k3} + \sin\alpha t \cdot \epsilon_{ik3}, \\ \mathfrak{H}_{ik} &= i\alpha\epsilon_{ikl}n_l, \\ (\mathfrak{U}_t)_{ik} &= \cos\alpha t \cdot \delta_{ik} \\ &\quad + (1 - \cos\alpha t)n_i n_k + \sin\alpha t \cdot \epsilon_{ikl}n_l, \end{aligned}\right\} \tag{7}$$

and we immediately verify that $\mathfrak{U}_t$ induces [in agreement with remark (3)] the most general rotation in Υ with arbitrary rotation axis $n^{\rightarrow}$ and constant speed α, while $\mathfrak{U}_t^{(0)}$ is a constant speed rotation with axis $n^{\rightarrow(0)} = (0, 0, 1)$.

In using (6) we first split $\mathfrak{H}^{(0)}$ in two parts such that $\mathfrak{H} - \lambda\mathfrak{H}^{(0)}$ induces a rotation with an axis perpendicular to $n^{\rightarrow(0)}$. This is readily achieved choosing $\lambda = (n^{\rightarrow(0)}, n^{\rightarrow}) = n_3$; in fact, $\mathfrak{H} - \lambda\mathfrak{H}^{(0)}$ then generates a rotation around $n^{\rightarrow} - (n^{\rightarrow(0)}, n^{\rightarrow})n^{\rightarrow(0)} \perp n^{\rightarrow(0)}$. The result of the straightforward calculation of (6) is

$$\mathfrak{U}_t^{(1)} = \begin{pmatrix} C & S & \\ -S & C & \\ & & 1 \end{pmatrix} + (1 - n_3)\alpha t \begin{pmatrix} S & -C & \\ C & S & \\ & & \end{pmatrix} + \begin{pmatrix} & & n_1(1-C) - n_2 S \\ & & n_2(1-C) + n_1 S \\ n_1(1-C) + n_2 S & n_2(1-C) - n_1 S & 0 \end{pmatrix}, \tag{8}$$

where $C = \cos\alpha t$, $S = \sin\alpha t$.

The assertions of Sec. III can easily be verified:

(1) For an arbitrary initial state $\rho_0^{\rightarrow} = (\rho_1, \rho_2, \rho_3)$ the norm $\|\mathfrak{U}_t^{(1)}\rho_0^{\rightarrow}\|^2$ oscillates in general with an amplitude increasing indefinitely in time, and $\mathfrak{U}_t^{(1)\rightarrow}{}_0\rho$ does not represent a state at all times.

(2) Only if ρ_0 commutes with $H^{(0)}$ does the linear term in (8) vanish and the norm (as well as the entropy) become periodic in time:

$$\|\mathfrak{U}_t^{(1)}\rho_0^{\rightarrow}\|^2 = \tfrac{1}{2} + \rho_3^2 + 2\rho_3^2(1 - n_3^2)(1 - \cos\alpha t),$$

with average $\frac{1}{2} + \rho_3^2$.

[7] According to Gleason's theorem, there exist, in fact, many more physical states which cannot be described by density matrices in this case; but there is no need to consider them for our purpose.

(3) Even if the initial state commutes with $H^{(0)}$, its Born trajectory may lead periodically outside the sphere Σ of physical states. This is so in particular if $\rho_0^{\rightarrow} = 2^{-\frac{1}{2}}(0, 0, 1)$ is pure; the two points $\lambda_\pm$ of its spectrum develop according to

$$\lambda_\pm(t) = \tfrac{1}{2}\{1 \pm [1 + (1 - n_3^2)(1 - \cos \alpha t)]^{\frac{1}{2}}\},$$

and although its ergodic mean

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T dt \mathfrak{U}_t^{(1)} \rho_0^{\rightarrow}$$

exists, it lies slightly outside Σ at $2^{-\frac{1}{2}}(n_1, n_2, 1)$.

We also verify explicitly in (8) the following general property of the first Born approximation:

(4) The derivative at $t = 0$ of the first statistical Born approximation $\mathfrak{U}_t^{(1)}$ coincides with that of $\mathfrak{U}_t$.

In order to give an idea of the evolution of expectation values in the first Born approximation compared to their true evolution, we label the expectation values of spins S_z and S_x for the two initial states with spin in positive z (resp. x) direction:

Initial state	Observable	True evolution	Evolution in first Born approximation
↑	$\langle S_z \rangle$	$\frac{1}{2}((1 - n_3^2) \cos \alpha t + n_3^2)$	$\frac{1}{2}$
	$\langle S_x \rangle$	$\frac{1}{2}(n_1 n_3(1 - \cos \alpha t) - n_2 \sin \alpha t)$	$\frac{1}{2}(n_1(1 - \cos \alpha t) - n_2 \sin \alpha t)$
→	$\langle S_z \rangle$	$\frac{1}{2}(n_1 n_3(1 - \cos \alpha t) + n_2 \sin \alpha t)$	$\frac{1}{2}(n_1(1 - \cos \alpha t) + n_2 \sin \alpha t)$
	$\langle S_x \rangle$	$\frac{1}{2}(\cos \alpha t + n_1^2(1 - \cos \alpha t))$	$\frac{1}{2}(\cos \alpha t + \alpha t(1 - n_3) \sin \alpha t)$.

(9)

We note that the general type of the oscillation coincides; but, while their amplitude differs in the first three cases by a constant factor only, the amplitude of the sine in the last case increases linearly in time.

V. STATISTICAL BORN APPROXIMATION AND APPROACH TO EQUILIBRIUM

If we want to visualize the typical effect of the statistical Born approximation (6) in calculations of the approach to equilibrium of certain variables in statistical mechanics, we should exhibit a model which is soluble both exactly and in first Born approximation. We consider first the most primitive situation, where a single spin is coupled to a thermostat at infinite temperature whose state develops according to a semigroup of contractions admitting the vector $\rho_\infty^{\rightarrow} = (0, 0, 0)$ of maximal entropy as its unique asymptotic state. Such a behavior could, for instance, be realized by assuming a sort of *Stosszahlansatz* for the spin-bath interaction. Mathematically, this would result in assigning an over-all decay factor $\exp(-\lambda t)$ to the expectation values labeled in (9).

We see that in this naive model the Born approximations would not modify the set of equilibrium states since the exponential convergence factor dominates all powers in t. However, if we are interested in a detailed description of the approach to equilibrium at finite times, the situation is quite different. Even if we restrict ourselves to initial states which commute with $H^{(0)}$, the Born approximation may create [or wipe out, as in the first example (9)] oscillations.

A slightly different situation which arises in statistical mechanics should also be considered briefly. Suppose we are only interested in the populations of spin up and down in the z direction. Mathematically, this means that we consider only the σ_3 component of the states $\rho^{\rightarrow}$ (or, in $\mathfrak{K}$ terminology, the diagonal elements of ρ). Let $\mathfrak{D}_{ik} = \delta_{i3k3}$ be the corresponding projection operator in $\mathfrak{L}$. The reduced evolution operator $\mathfrak{D}\mathfrak{U}_t$ then satisfies the generalized master equation

$$\frac{d}{dt}(\mathfrak{D}\mathfrak{U}_t)$$
$$= -i\mathfrak{D}\mathfrak{H}\mathfrak{D}\mathfrak{U}_t - i\mathfrak{D}\mathfrak{H} \exp[-i(I - \mathfrak{D})\mathfrak{H} \cdot t](I - \mathfrak{D})$$
$$- \int_0^t d\tau \mathfrak{D}\mathfrak{H} \exp[-i(I - \mathfrak{D})\mathfrak{H}(t - \tau)](I - \mathfrak{D})\mathfrak{H}\mathfrak{D}\mathfrak{U}_\tau.$$

If we restrict ourselves to initial states $\rho^{\rightarrow} = \mathfrak{D}\rho^{\rightarrow}$, the second term on the right-hand side vanishes and we can set up the iterative solution

$$i\frac{d}{dt}(\mathfrak{D}\mathfrak{U}_t)^{(\nu)}$$
$$= -i\mathfrak{D}\mathfrak{H}(\mathfrak{D}\mathfrak{U}_t)^{(\nu-1)} - \int_0^t d\tau \mathfrak{D}\mathfrak{H}$$
$$\times \exp[-i(I - \mathfrak{D})\mathfrak{H}(t - \tau)](I - \mathfrak{D})\mathfrak{H}(\mathfrak{D}\mathfrak{U}_\tau^{(\nu-1)}),$$

where $(\mathfrak{D}\mathfrak{U}_t)^{(0)}$ satisfies the equation

$$i\frac{d}{dt}(\mathfrak{D}\mathfrak{U}_t)^{(0)} = -i\mathfrak{D}\mathfrak{H}(\mathfrak{D}\mathfrak{U}_t)^{(0)}.$$

The series so obtained suffers, however, from the same defect as (1), namely, that the positivity of the states is not necessarily conserved during the evolution, as can be verified in the example of spin.

ACKNOWLEDGMENT

I thank Professor G. Mohan for many interesting discussions.

APPENDIX A

Here we establish the relation between our statistical Born approximation (6) and the formulation of Kubo .[6]

Applying (6) on a state $\rho^{\rightarrow}$ and introducing the quantities

$$\rho_t^{\rightarrow(0)} = \mathfrak{U}_t^{(0)}\rho^{\rightarrow} \quad \text{and} \quad (\Delta\rho_t)^{\rightarrow} = (\mathfrak{U}_t^{(1)} - \mathfrak{U}_t^{(0)})\rho^{\rightarrow},$$

we obtain

$$(\Delta\rho_t)^{\rightarrow} = -i\int_0^t d\tau \mathfrak{U}_{t-\tau}^{(0)}(\mathfrak{H} - \mathfrak{H}^{(0)})\rho_\tau^{\rightarrow(0)}$$

$$= -i\int_0^t d\tau \mathfrak{U}_{t-\tau}^{(0)}[V, \rho_t^{(0)}]^{\rightarrow},$$

and, inserting the time-dependent interaction $V = -AF(t)$, we have

$$(\Delta\rho_t)^{\rightarrow} = i\int_0^t d\tau \mathfrak{U}_{t-\tau}^{(0)}[A, \rho_\tau^{(0)}]^{\rightarrow}F(\tau).$$

This corresponds to Kubo's formula (2.14) if the disturbance function $F(t)$ has positive support.

As mentioned in the Introduction, this statistical Born approximation has its analog in classical mechanics. Replacing ρ_t by the phase-space density f_t, we obtain

$$\Delta f_t = -\int_0^t d\tau e^{i(t-\tau)L}(A, f_\tau^{(0)})F(\tau)$$

[Ref. 6, formula (2.9)], where

$$iLf \equiv (H, f) = \sum_i \left(\frac{\partial H}{\partial q_i}\frac{\partial f}{\partial p_i} - \frac{\partial H}{\partial p_i}\frac{\partial f}{\partial q_i}\right),$$

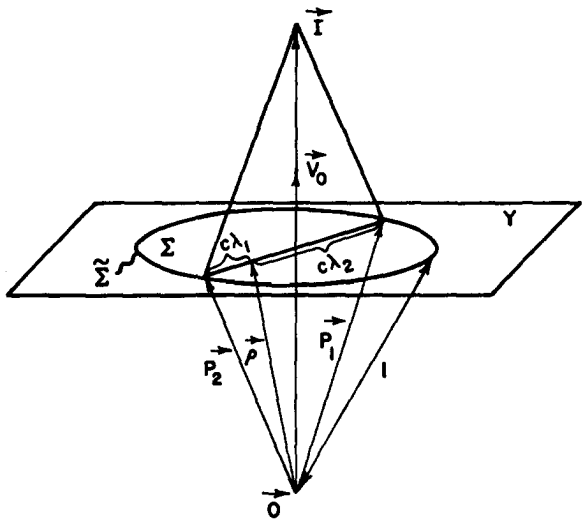


FIG. 1. Geometrical construction of the spectrum $\{\lambda_1, \lambda_2\}$ of ρ.

with q_i and p_i being the canonical variables in phase space.

APPENDIX B

In the 2-dimensional case the density matrices satisfy the relations

$$\rho^{\rightarrow} = \lambda_1 P_1^{\rightarrow} + \lambda_2 P_2^{\rightarrow}; \quad P_1^{\rightarrow} + P_2^{\rightarrow} = I;$$

$$P_1^{\rightarrow}, P_2^{\rightarrow} \in \bar{\Sigma}.$$

The spectrum $\{\lambda_1, \lambda_2\}$ of ρ can then be constructed geometrically as in Fig. 1. It is also directly visible that the spectrum of any $\rho^{\rightarrow}$ off the sphere Σ is not positive.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Note on Relating Pfaffians and Hafnians with Determinants and Permanents

A. DE LUCA AND L. M. RICCIARDI*
Laboratorio di Cibernetica del Consiglio Nazionale delle Ricerche, Naples, Italy
AND
R. VASUDEVAN†
Istituto di Fisica Teorica dell'Università, Naples, Italy

(Received 24 April 1969)

In connection with renormalization problems in quantum field theory, Pfaffians and Hafnians are considered. A noteworthy expansion of determinants and permanents is given, in terms of Pfaffian-like and Hafnian-like objects, that immediately leads to the known reduction of a determinant to a single Pfaffian. Finally, some auxiliary results are presented.

1. INTRODUCTION

At the turn of this century algebraic objects similar to determinants, known by the name of Pfaffians, were studied in mathematical literature. These Pfaffians are related to antisymmetric determinants, and many results regarding these objects have been expounded in detail.[1]

As stated in Ref. 1, these objects, which appear in form of arrays, were introduced by Pfaff in 1815 in connection with the solution of differential equations. The interest in these was revived by Caianiello[2] in connection with renormalization problems in quantum field theory, where one deals with averages of anticommuting operators. The fermion propagators can be arranged in terms of such objects and, then, related to determinants. However, boson propagators correspond to another algebraic object called Hafnian (introduced by Caianiello[1]), which may be arranged in arrays similar to that of a Pfaffian. Though these two objects look similar and possess an expansion rule which resembles that of a Pfaffian, there is an important proviso, that the signature of each term of the expansion of the new object is always positive. This is linked with the fact that boson operators obey commutation rules different from those of fermion operators. This brings to our mind the analogous situation between permanents and determinants: the expansion rule for permanents resembles that of determinants without the parity sign in front of the terms.

There are many theorems concerning Pfaffians and determinants in the literature.[2] Pfaffians have also been pressed into service in the descriptions of problems relating the lattice statistics, e.g., the number of ways dimers can be placed on a 2-dimensional lattice. They have been also found to be of great use in the Ising-model problems and, in other contexts, of combinatorial mathematics.[3]

It is the purpose of this paper to establish relations between determinants and Pfaffians and between permanents and Hafnians in a new fashion by defining the elements of these objects in a particular way. It is well known that the square of the Pfaffian directly yields the value of an antisymmetric determinant, while there is no such relation between the Hafnian and the permanent. In Sec. 2, we show how any determinant can be expressed as a sum of Pfaffian-like objects whose elements are suitably chosen. Similar relations between Hafnians and permanents are also established. In Sec. 3, we derive a result of Brioschi[4] connecting a single Pfaffian with any determinant, starting from our standpoint. Some other auxiliary results are also worked out.

2. A NOTEWORTHY EXPANSION OF A DETERMINANT AND A PERMANENT

It is well known that the determinant of an antisymmetric matrix can be equated to the square of a Pfaffian whose elements are the actual elements of the matrix. It is also well known that there is no simple relation between a permanent and a Hafnian as there is between a Pfaffian and a determinant. In this section we propose to express any determinant of even order as a sum of Pfaffian-like objects whose elements are defined in a particular fashion, e.g., as suitable combinations of the elements of the matrix. Similarly, a

* Present address: Committee on Mathematical Biology, The University of Chicago, Chicago, Ill.
† On leave of absence from the Institute for Mathematical Sciences, Madras, India.

[1] E. R. Caianiello, Nuovo Cimento **10**, 1634 (1953).
[2] E. R. Caianiello, Nuovo Cimento Suppl. **14**, 177 (1959).
[3] E. W. Montroll, in *Applied Combinatorial Mathematics*, E. F. Beckenbach, Ed. (John Wiley & Sons, Inc., New York, 1964), Chap. 4.
[4] F. Brioschi, Crelle's J. **52**, 133 (1855). See also T. Muir, *The Theory of Determinants* (Dover Publications, Inc., New York, 1960), Vol. II, pp. 276–277.

permanent of even order will be expressed as a sum of Hafnian-like objects.

We start with the definition of a determinant $\Delta(A)$ of a given $N \times N$ square matrix $A \equiv \|a_{hk}\|$, $h, k = 1, 2, \cdots, N$, as

$$\Delta(A) = \epsilon^{i_1 i_2 \cdots i_N} a_{i_1 1} a_{i_2 2} \cdots a_{i_N N}, \tag{2.1}$$

where $\epsilon^{i_1 i_2 \cdots i_N}$ is the Ricci indicator of the parity of the permutation of the indices, given by

$$\begin{aligned} \epsilon^{i_1 i_2 \cdots i_N} &= +1, \text{ for an } \textit{even} \text{ permutation of } i_1 i_2 \cdots i_N \\ &\quad \text{with respect to } 1, 2, \cdots, N, \\ &= -1, \text{ for an } \textit{odd} \text{ permutation of the indices} \\ &\quad \text{with respect to the natural order,} \\ &= 0, \text{ if two indices are } \textit{equal.} \end{aligned}$$

The usual summation convention of repeated indices is implied in the above definition.

Let us concern ourselves here with determinants of even order and put $N = 2n$. In this case the expression for the determinant (2.1) can be rewritten in another form given by

$$\Delta(A) = \epsilon^{i_1 \cdots i_{2n}} \prod_{k=1}^{n} (a_{i_{2k-1}, 2k-1} a_{i_{2k}, 2k} - a_{i_{2k}, 2k-1} a_{i_{2k-1}, 2k}), \quad i_{2k-1} < i_{2k}, \text{ for any } k. \tag{2.2}$$

Setting

$$[p, q]^{(r,s)} \equiv a_{pr} a_{qs} - a_{qr} a_{ps}, \quad p, q, r, s = 1, 2, \cdots, 2n, \tag{2.3}$$

Eq. (2.2) can be rewritten in the following more compact form:

$$\Delta(A) = \epsilon^{i_1 \cdots i_{2n}} \prod_{k=1}^{n} [i_{2k-1}, i_{2k}]^{(2k-1, 2k)}, \quad i_{2k-1} < i_{2k}, \text{ for any } k. \tag{2.4}$$

Let us note that, because of (2.3),

$$[p, q]^{(r,s)} = -[q, p]^{(r,s)},$$
$$[p, q]^{(s,r)} = -[p, q]^{(r,s)}.$$

We will now give an algorithm to write the determinant, making use of the Pfaffian expansion.

It is easy to see that if we exchange the indices within a bracket with one in some other bracket, the upper pair of indices also change. Furthermore, for any permutation of pairs of indices the Ricci indicator does not change. If we also switch the upper indices correspondingly, the term does not change. Hence, it is possible to rewrite expression (2.4) in the following manner:

$$\Delta(A) = \eta^{j_1 j_3 \cdots j_{2n-1}} \Theta^{i_1 i_2 \cdots i_{2n}} \prod_{k=1}^{n} [i_{2k-1}, i_{2k}]^{(j_{2k-1}, j_{2k-1}+1)}, \tag{2.5}$$

where $j_1, j_3, \cdots, j_{2n-1}$ are odd and η and Θ are defined as follows:

$$\begin{aligned} \eta^{j_1 j_2 \cdots j_n} &= 0, \quad \text{if two } j\text{'s are equal,} \\ &= +1, \quad \text{otherwise,} \end{aligned} \tag{2.6}$$

and

$$\begin{aligned} \Theta^{i_1 i_2 \cdots i_{2n}} &= \epsilon^{i_1 i_2 \cdots i_{2n}}, \\ &\quad \text{for } i_1 < i_2, i_3 < i_4, \cdots, i_{2n-1} < i_{2n} \\ &\quad \text{and } i_1 < i_3 < i_5 < \cdots < i_{2n-1}, \\ &= 0, \quad \text{otherwise.} \end{aligned} \tag{2.7}$$

If we ignore the upper j indices, the sum of the products of the brackets containing the i indices along the Ricci indicator can be denoted by a Pfaffian, where each element of the Pfaffian represents each bracket with the i indices. Having written out the terms of the Pfaffian, we assign the upper j indices in a particular order to each bracket of each term in the Pfaffian without repetition. Later, we write out the same Pfaffian, but permuting the order of upper j indices in each term, obtaining, thus, $n!$ quantities, whose sum is identical to the determinant, term by term.

It will be worthwhile to observe here that the number of single terms in the development of (2.5) will equal the number of terms in equation (2.1) for the development of the determinant. This is so because each Pfaffian-like quantity contains $(2n - 1)!!$ terms, each of which contains 2^n single terms to yield the total numbers of $n! \times (2n - 1)!! \times 2^n = (2n)!$ terms of determinant. Since by this procedure we have identified the determinant $\Delta(A)$ with the sum of $n!$ Pfaffian-like quantities, with a suitable definition of its elements leading to their equivalence term by term, it requires little effort to extend the same procedure to equate any even order permanent with a sum of $n!$ Hafnian-like quantities with a proper modification of the elements.

As is well known, the permanent $P(A)$ of $N \times N$ even order matrix A is

$$P(A) = \eta^{i_1 i_2 \cdots i_N} a_{i_1 1} a_{i_2 2} \cdots a_{i_N N}, \tag{2.8}$$

with the definition of η given by (2.6). Similar to the procedure adopted in the case of determinant, we re-express (2.8) for a permanent of even $(2n \times 2n)$ order as

$$P(A) = \eta^{i_1 i_2 \cdots i_{2n}} \prod_{k=1}^{n} \{i_{2k-1}, i_{2k}\}^{(2k-1, 2k)}, \quad i_{2k-1} < i_{2k}, \text{ for any } k, \tag{2.9}$$

where, for any integer p, q, r, s, we set

$$\{p, q\}^{(r,s)} \equiv a_{pr} a_{qs} + a_{qr} a_{ps}.$$

By this definition

$$\{p, q\}^{(r,s)} = \{q, p\}^{(r,s)},$$
$$\{p, q\}^{(r,s)} = \{p, q\}^{(s,r)}.$$

A similar algorithm for a permanent, making use of a Hafnian, can be arrived at by the following steps:

The permanent can be expressed as

$$P(A) = \eta^{j_1 j_3 \cdots j_{2n-1}} \zeta^{i_1 i_2 \cdots i_{2n}} \prod_{k=1}^{n} \{i_{2k-1}, i_{2k}\}^{(j_{2k-1}, j_{2k-1+1})}, \tag{2.10}$$

where

$$\begin{aligned} \zeta^{i_1 i_2 \cdots i_{2n}} &= \eta^{i_1 \cdots i_{2n}}, \\ &\quad \text{if} \quad i_1 < i_2, i_3 < i_4, \cdots, i_{2n-1} < i_{2n}, \\ &\quad \text{and } i_1 < i_3 < i_5 < \cdots < i_{2n-1}, \\ &= 0, \quad \text{otherwise.} \end{aligned} \tag{2.11}$$

If we now ignore the upper j indices, the sum of the product of brackets of i indices can be expressed in a Hafnian-like fashion, with the definition of the brackets indicated above. After writing out the terms of the Hafnian, we assign the j indices to each term in a particular order. There are $n!$ possible ways of permuting this order in each term. The sum of these $n!$ Hafnian-like quantities equals the permanent, term by term.

Let us now illustrate how this method works for the single case of a determinant or permanent of a 4 × 4 matrix

$$A \equiv \begin{Vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{Vmatrix}. \tag{2.12}$$

We consider the Pfaffian

$$\begin{vmatrix} [12] & [13] & [14] \\ & [23] & [24] \\ & & [34] \end{vmatrix},$$

whose expansion is

$$[12][34] - [13][24] + [14][23].$$

Then, by means of our algorithm, the determinant $\Delta(A)$ of A can be written, in conformity with the expression obtained (2.5), as

$$\begin{aligned} \Delta(A) = {} & [12]^{(12)}[34]^{(34)} - [13]^{(12)}[24]^{(34)} \\ & + [14]^{(12)}[23]^{(34)} + [12]^{(34)}[34]^{(12)} \\ & - [13]^{(34)}[24]^{(12)} + [14]^{(34)}[23]^{(12)}, \end{aligned} \tag{2.13}$$

where

$$\begin{aligned} [r, s]^{(12)} &\equiv a_{r1}a_{s2} - a_{s1}a_{r2}, \\ [r, s]^{(34)} &\equiv a_{r3}a_{s4} - a_{s3}a_{r4}. \end{aligned}$$

It is remarkable that we can arrange (2.13) in a sum of two Pfaffians:

$$\begin{vmatrix} [12]^{(12)} & [13]^{(12)} & [14]^{(12)} \\ & [23]^{(34)} & [24]^{(34)} \\ & & [34]^{(34)} \end{vmatrix} + \begin{vmatrix} [12]^{(34)} & [13]^{(34)} & [14]^{(34)} \\ & [23]^{(12)} & [24]^{(12)} \\ & & [34]^{(12)} \end{vmatrix}. \tag{2.14}$$

By actual multiplication of the brackets we can easily check that the determinant and the sum of the two Pfaffians (2.14) are identical with each other, term by term. In the case of a permanent of a 4 × 4 matrix, we consider the Hafnian

$$\begin{vmatrix} \{12\} & \{13\} & \{14\} \\ & \{23\} & \{24\} \\ & & \{34\} \end{vmatrix},$$

whose expansion is

$$\{12\}\{34\} + \{13\}\{24\} + \{14\}\{23\}.$$

By means of (2.10) the permanent $P(A)$ of A can be expressed as

$$\begin{aligned} P(A) = {} & \{12\}^{(23)}\{34\}^{(34)} + \{13\}^{(12)}\{24\}^{(34)} \\ & + \{14\}^{(12)}\{23\}^{(34)} + \{12\}^{(34)}\{34\}^{(12)} \\ & + \{13\}^{(34)}\{24\}^{(12)} + \{14\}^{(34)}\{23\}^{(12)}, \end{aligned} \tag{2.15}$$

where

$$\begin{aligned} \{r, s\}^{(12)} &\equiv a_{r1}a_{s2} + a_{s1}a_{r2}, \\ \{r, s\}^{(34)} &\equiv a_{r3}a_{s4} + a_{s3}a_{r4}. \end{aligned}$$

Let us note that the right-hand side of (2.15) can be written just as sum of the two Hafnians

$$\begin{vmatrix} \{12\}^{(12)} & \{13\}^{(12)} & \{14\}^{(12)} \\ & \{23\}^{(34)} & \{24\}^{(34)} \\ & & \{34\}^{(34)} \end{vmatrix} + \begin{vmatrix} \{12\}^{(34)} & \{13\}^{(34)} & \{14\}^{(34)} \\ & \{23\}^{(12)} & \{24\}^{(12)} \\ & & \{34\}^{(12)} \end{vmatrix}. \tag{2.16}$$

Here also one can check that the permanent (2.15) is given, term by term, by the sum of the 2! Hafnian-like quantities (2.16).

The above algorithm for the determinant and the permanent will hold for any even order. If a permanent or determinant of odd order is to be realized, it is a simple matter to make it even by adjoining to the odd matrix another column and a row a_{0k}, with $k = 0, \cdots, N$, and a_{h0}, with $h = 0, \cdots, N$. We have $a_{00} = 1$ and make all elements of the column and of the row zero.

We want to point out at this stage that in (2.5), expressing $\Delta(A)$, each Pfaffian-like object has the property that if the superindices of two brackets are the same, it vanishes. To prove this statement, let us write down (2.5) without the symbol η and with two j indices equal, say $j_{2h-1} = j_{2l-1} = j$. We have

$$\Theta^{i_1 \cdots i_{2n}}[i_1, i_2]^{(j_1, j_1+1)}[i_3, i_4]^{(j_3, j_3+1)} \cdots [i_{2h-1}, i_{2h}]^{(j, j+1)} \cdots [i_{2l-1}, i_{2l}]^{(j, j+1)} \cdots [i_{2n-1}, i_{2n}]^{(j_{2n-1}, j_{2n-1}+1)}. \quad (2.17)$$

In the above expression, let us fix the values of all the indices but (i_{2h-1}, i_{2h}) and (i_{2l-1}, i_{2l}). (2.17), with this condition, can be expressed as

$$\Theta^{\bar{i}_1 \bar{i}_2 \cdots i_{2h-1} i_{2h} \bar{i}_{2h+1} \cdots i_{2l-1} i_{2l} \cdots \bar{i}_{2N}}[\bar{i}_1, \bar{i}_2]^{(j_1, j_1+1)} \times [\bar{i}_3, \bar{i}_4]^{(j_3, j_3+1)} \cdots [i_{2h-1}, i_{2h}]^{(j, j+1)} \cdots [i_{2l-1}, i_{2l}]^{(j, j+1)} \cdots [\bar{i}_{2N-1}, \bar{i}_{2N}]^{(j_{2N-1}, j_{2N})}, \quad (2.18)$$

where the symbol —, used to avoid confusion, means that the corresponding indices are not summed. It is evident that i_{2l-1}, i_{2l}, i_{2h-1}, i_{2h} can range over only four values, say p, q, r, and s. Hence, (2.18), doing the summation explicitly over the four values, will sum as follows:

$$\text{const} \times \Theta^{i_{2h-1} i_{2h} i_{2l-1} i_{2l}}[i_{2h-1}, i_{2h}]^{(j, j+1)} \times [i_{2l-1}, i_{2l}]^{(j, j+1)}, \quad (2.19)$$

where i_{2h-1}, i_{2h}, i_{2l-1}, i_{2l} go through the values p, q, r, and s, subject to the usual conditions implied by Θ. Equation (2.19) is just a Pfaffian given by

$$[p, q]^{(j, j+1)}[r, s]^{(j, j+1)} - [p, r]^{(j, j+1)}[q, s]^{(j, j+1)} + [p, s]^{(j, j+1)}[q, r]^{(j, j+1)},$$

if $p < q < r < s$. In terms of the elements of the original matrix A, this can be expanded as

$$[a_{pj}a_{q,j+1} - a_{qj}a_{p,j+1}][a_{rj}a_{s,j+1} - a_{sj}a_{r,j+1}] - [a_{pj}a_{r,j+1} - a_{rj}a_{p,j+1}][a_{qj}a_{s,j+1} - a_{sj}a_{q,j+1}] + [a_{pj}a_{s,j+1} - a_{sj}a_{p,j+1}][a_{qj}a_{r,j+1} - a_{rj}a_{q,j+1}]. \quad (2.20)$$

It can be easily checked that (2.20) equals zero. Also, it can be shown that this Pfaffian corresponds to a 4×4 determinant with two columns having zero elements only, which means the Pfaffian is zero. Thus, we have seen that even if we do not take into account the restrictions on j indices implied by the definition of η, the nature of expansion of each Pfaffian-like object makes it zero if it contains two or more equal j indices. However, the vanishing of the terms with two j's being equal is not valid for the case of the expression of the permanent in terms of the Hafnians. The presence of positive signs everywhere in the terms of the Hafnians will not make the terms add up to zero identically as in the case of the Pfaffians and determinants. However, the presence of η symbol with its definition given above is such that such repetitions of j indices will make the expression zero.

In view of the above deductions, it would be possible to express the determinant in terms of Pfaffians with a suitable modification of the elements instead of giving a blanket rule for putting in the upper indices, as is done in the algorithms presented above.

We first write the elements of the Pfaffian with brackets containing i indices, ignoring the upper j indices. We give the first line of the Pfaffian one of the upper indices, say j_1. For all the other rows of the Pfaffian, each element is symmetrized with respect to all the other j indices. As an example, we write out a Pfaffian, fixing j_1:

$$\begin{vmatrix} (12)^{(j_1, j_1+1)} & (13)^{(j_1, j_1+1)} \cdots & (1, 2n)^{(j_1, j_1+1)} \\ & \sum_{j_i} (23)^{(j_i, j_i+1)} \cdots & \sum_{j_i} (2, 2n)^{(j_i, j_i+1)} \\ & \cdot & \\ & \cdot & \\ & \cdot & \\ & & \cdots \sum_{j_i} (2n-1, 2n)^{(j_i, j_i+1)} \end{vmatrix}.$$

Now we write the other Pfaffians, described as above, with j_1 ranging through the other values. Hence, it is now possible to write the determinant as a sum of n such Pfaffians, with symmetrized elements for all the rows except the first one. The j for the first row is fixed, and the elements in the rest of the rows are symmetrized with respect to all the other j indices; such Pfaffians are obtained with the n values of j, fixed for the first row. It is not possible to derive such a procedure for the permanents in terms of the Hafnians.

3. REDUCTION OF A DETERMINANT TO A SINGLE PFAFFIAN

In the previous section we equated a determinant to the sum of $n!$ Pfaffians, with suitable algorithms. In this section we will express a determinant in terms of a single Pfaffian. To accomplish this, we recall Eq. (2.5):

$$\Delta(A) = \eta^{j_1 j_3 \cdots j_{2n-1}} \Theta^{i_1 i_2 \cdots i_{2n}} \prod_{k=1}^{n} [i_{2k-1}, i_{2k}]^{(j_{2k-1}, j_{2k-1}+1)}.$$

We will do the summation of j before doing the summation over the i's in the brackets, ignoring the command of the η symbol not to repeat the j values.

The expression now becomes

$$\Theta^{i_1 i_2 \cdots i_{2n}} \prod_{k=1}^{n} \sum_{j_{2k-1}} [i_{2k-1}, i_{2k}]^{(j_{2k-1}, j_{2k-1}+1)} = \Theta^{i_1 i_2 \cdots i_{2n}} \prod_{k=1}^{n} l_{i_{2k-1} i_{2k}} \quad (3.1)$$

The above expression is exactly a single Pfaffian, with its terms given by

$$\begin{aligned} l_{rs} &\equiv [r, s]^{(12)} + [r, s]^{(34)} \\ &\quad + [r, s]^{(56)} + \cdots + [r, s]^{(2n-1, 2n)} \\ &= \sum_{j_{2k-1}=1}^{2n-1} [r, s]^{(j_{2k-1}, j_{2k-1}+1)} \end{aligned} \quad (3.2)$$

When we write out this Pfaffian, we get exactly the $n!$ Pfaffians given by the first algorithm of Sec. 2 and a number of other Pfaffians, whose expansion will contain brackets with repeated upper j indices. We have also deduced, in Sec. 2, that any such expression with repeated j indices becomes zero. Hence, the above single Pfaffian will actually give the value of the determinant. We wish to point out that the elements of the Pfaffian in (3.2) are exactly what Brioschi obtained[4] by using another type of procedure.

We can now easily extend this method to expressing a permanent as a single Hafnian, with elements given by

$$\begin{aligned} \lambda_{rs} &\equiv \{r, s\}^{(12)} + \{r, s\}^{(34)} + \cdots + \{r, s\}^{(2n-1, 2n)} \\ &= \sum_{j_{2k-1}=1}^{2n-1} \{r, s\}^{(j_{2k-1}, j_{2k-1}+1)} \end{aligned} \quad (3.3)$$

However, when expanding the terms of the Hafnian, we have to impose the condition that we avoid all those terms in which upper j indices repeat themselves. This separate injunction is necessitated by the fact that in the expansion of a permanent, as well as of a Hafnian, no parity signs appear and also the brace is a sum of two terms and not a difference, as in the case of bracket for a Pfaffian. The determinant, therefore, is

$$\Delta(A) = \begin{vmatrix} l_{12} & l_{13} & \cdots & l_{1,2n} \\ & l_{23} & \cdots & l_{2,2n} \\ & & \cdots & l_{2n-1,2n} \end{vmatrix}, \quad (3.4)$$

and the permanent is given by

$$P(A) = \begin{vmatrix} \lambda_{12} & \lambda_{13} & \cdots & \lambda_{1,2n} \\ & \lambda_{23} & \cdots & \lambda_{2,2n} \\ & & \cdots & \lambda_{2n-1,2n} \end{vmatrix}, \quad (3.5)$$

with the condition that in the expansion of the Hafnian (3.5) we ignore the terms containing two repeated upper indices.

In conclusion, we will derive some auxiliary results from the first algorithm developed in Sec. 2. We saw that any determinant can be written out as a sum of $n!$ Pfaffian-like quantities. With the definitions given in (2.5), we first write out a Pfaffian with the i brackets, ignoring the upper j indices, and then permute the upper indices over the terms of the Pfaffian in all possible ways, thus obtaining the $n!$ quantities. To any $2n \times 2n$ matrix $A \equiv \|a_{hk}\|$, $h, k = 1, 2, \cdots, 2n$, let us associate the matrix

$$A_{12}^{h,h+1} \equiv \left\| \begin{array}{cccc} a_{11} & \cdots a_{1h} & a_{1,h+1} & \cdots a_{1,2n} \\ a_{21} & \cdots a_{2h} & a_{2,h+1} & \cdots a_{2,2n} \\ a_{31} & \cdots a_{2h} & a_{2,h+1} & \cdots a_{3,2n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ a_{2n,1} & \cdots a_{2h} & a_{2,h+1} & \cdots a_{2n,2n} \end{array} \right\|.$$

In the above we have in the hth and $(h+1)$th columns after the second row, the elements a_{2h}, a_{2h+1} being repeated up to the end of the columns.

If we now calculate the determinant $\Delta(A_{12}^{h,h+1})$ of the previous matrix, making use of the algorithm (2.5), we obtain a sum (with respect to j indices) of $(n-1)!$ quantities:

$$\begin{aligned} \Delta(A_{12}^{h,h+1}) &= \eta^{hj_3 \cdots j_{2n-1}} \\ &\quad \times \Theta^{1i_2 \cdots i_{2n}} [1, i_2]^{(h,h+1)} [i_3, i_4]^{(j_3, j_3+1)} \cdots \\ &\quad [i_{2n-1}, i_{2n}]^{(j_{2n-1}, j_{2n-1}+1)}, \end{aligned}$$

since all the terms containing $[i_{2k}, i_{2k+1}]^{(h,h+1)}$ with $i_{2k} \neq 1$ are vanishing. Because of the properties

$$\Theta^{1 i_2 i_3 \cdots i_{2n}} = (-1)^{i_2} \Theta^{i_3 i_4 \cdots i_{2n}},$$
$$[1 i_2]^{(h,h+1)} = [12]^{(h,h+1)},$$

for any fixed index i_2, the previous expression can be written as

$$[12]^{(h,h+1)} (-1)^{i_2} \Delta_{h,h+1}^{(1,i_2)} = [12]^{(h,h+1)} \alpha_{h,h+1}^{(1,i_2)},$$

where $\Delta_{h,h+1}^{(1i_2)}$ and $\alpha_{h,h+1}^{(1i_2)}$ are, respectively, the *complementary minor* and the *algebraic complement* of the minor determined by the 1, i_2 rows and $h, h+1$ columns. Therefore, we can write

$$\Delta(A_{12}^{h,h+1}) = [12]^{(h,h+1)} \sum_{p=2}^{2n} \alpha_{h,h+1}^{(1,p)}. \quad (3.6)$$

In the case of a matrix,

$$A_{12}^{h,k} \equiv \left\| \begin{array}{cccccc} a_{11} & \cdots & a_{1h} & \cdots & a_{1k} & \cdots a_{1,2n} \\ a_{21} & \cdots & a_{2h} & \cdots & a_{2k} & \cdots a_{2,2n} \\ a_{31} & \cdots & a_{2h} & \cdots & a_{2k} & \cdots a_{3,2n} \\ \cdot & & \cdot & & \cdot & \cdot \\ \cdot & & \cdot & & \cdot & \cdot \\ \cdot & & \cdot & & \cdot & \cdot \\ a_{2n,1} & \cdots & a_{2h} & \cdots & a_{2k} & \cdots a_{2n,2n} \end{array} \right\|,$$

in which in the hth and kth columns, after the second row, the elements a_{2h}, a_{2k} are repeated up to the end of the columns, we have

$$\Delta(A_{12}^{h,k}) = (-1)^{k-\bar{h}+1}\Delta(\bar{A}_{12}^{\bar{h},\bar{h}+1}),$$

where the matrix $\bar{A}_{12}^{h,h+1}$ is obtained from $A_{12}^{h,k}$ by a cyclic permutation of the columns $h+1, h+2, \cdots, k$. By (3.6), we can write

$$\begin{aligned}\Delta(A_{12}^{h,k}) &= (-1)^{k-\bar{h}+1}\Delta(\bar{A}_{12}^{\bar{h},\bar{h}+1}) \\ &= (-1)^{k-\bar{h}+1}[12]^{(\bar{h},\bar{h}+1)}\sum_{p=1}^{2n}\bar{\alpha}_{h,h+1}^{(1,p)} \\ &= [12]^{\bar{h},\bar{k}}\sum_{p=2}^{2n}\alpha_{\bar{h}\bar{k}}^{(1,p)}. \end{aligned} \tag{3.7}$$

Now if we do the sum of all these determinants with h and k ranging on all possible values, with $h < k$, we obtain by means of (3.7)

$$\begin{aligned}\sum_{\substack{h,k\\h<k}}^{1\cdots 2n}\Delta(A_{12}^{h,k}) &= \sum_{\substack{h,k\\h<k}}^{1\cdots 2n}[12]^{(h,k)}\sum_{p=2}^{2n}\alpha_{hk}^{(1,p)} \\ &= \sum_{p=2}^{2n}\sum_{\substack{h,k\\h<k}}^{1\cdots 2n}[12]^{(h,k)}\alpha_{hk}^{(1,p)}. \end{aligned} \tag{3.8}$$

Now the quantity

$$\sum_{\substack{h,k\\h<k}}^{1\cdots 2n}[12]^{(h,k)}\alpha_{hk}^{(1,p)}$$

is always zero, except when $p = 2$, in which case it equals $\Delta(A)$. Therefore, (3.8) becomes

$$\sum_{p=2}^{2n}\delta_{p2}\Delta(A) = \Delta(A).$$

In any case, we can write

$$\Delta(A) = \sum_{\substack{h,k\\h<k}}^{1\cdots 2n}\Delta(A_{12}^{h,k}). \tag{3.9}$$

Hence, the determinant of the original matrix can be written as a sum of $n(2n-1)$ determinants of the matrices $A_{12}^{h,k}$. To derive (3.9), we have fixed the rows 12; but, naturally, this procedure can be generalized whatever the choice of the pair of rows is. Let us observe that the result (3.9) differs from the one that could be obtained by using the Laplace rule. In fact, in this case, in the matrix $A_{12}^{h,k}$ we would have to put in the columns h and k, after the second row, all zero elements.

We can find a result similar to (3.9) in the case of the permanent $P(A)$ of A, but here $A_{12}^{h,k}$ has all zero elements in the h and k columns after the second row; this result is, obviously, a trivial consequence of the Laplace rule for the calculation of a permanent.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Exact Stationary States in Local Relativistic Field Theories*

GERALD ROSEN
Drexel Institute of Technology, Philadelphia, Pa. 19104

(Received 27 May 1969)

For theories featuring n scalar fields and an interaction energy density that converges rapidly for large field magnitudes to the form representative of a linear theory, we present exact physical solutions to the Schrödinger stationary state equation. These solutions show that such quantum field theories are effectively linear with no interaction and scattering between the quanta.

Consider the class of local relativistic field theories based on Lagrangian densities of the form

$$\begin{aligned}\mathfrak{L} &= \tfrac{1}{2}\dot{\phi}\cdot\dot{\phi} - \tfrac{1}{2}(\nabla\phi)\cdot(\nabla\phi) - u \\ &\equiv \tfrac{1}{2}\sum_{i=1}^{n}[(\dot{\phi}_i)^2 - (\nabla\phi_i)^2] - u, \end{aligned} \tag{1}$$

where $\phi = (\phi_1, \cdots, \phi_n)$ denotes an n-tuple of real scalar fields which are coupled through the interaction energy density $u = u(\phi)$, a real 1-tuple function of ϕ. With the momentum density n-tuple for such a theory denoted by $\pi \equiv \partial\mathfrak{L}/\partial\dot{\phi} = \dot{\phi}$, we obtain the canonical Hamiltonian

$$H = H[\phi, \pi] = \int [\tfrac{1}{2}\pi\cdot\pi + \tfrac{1}{2}(\nabla\phi)\cdot(\nabla\phi) + u]\, d^3x. \tag{2}$$

Quantization can be effected in the Schrödinger picture [with $\phi = \phi(\mathbf{x})$ and $\pi = \pi(\mathbf{x})$ independent of time] by evoking the "coordinate-diagonal" representation, in which the n real scalar fields $\phi_i(\mathbf{x})$ are diagonalized for all $\mathbf{x}$, their conjugate momentum densities are represented by the functional differential operators $\pi_i(\mathbf{x}) = -i\hbar\delta/\delta\phi_i(\mathbf{x})$, and the quantum Hamiltonian is the second-order functional differential operator

$$\begin{aligned}\mathbf{H} &= H\left[\phi(\mathbf{x}), -i\hbar\frac{\delta}{\delta\phi(\mathbf{x})}\right] \\ &= \int\left(-\tfrac{1}{2}\hbar^2\frac{\delta}{\delta\phi(\mathbf{x})}\cdot\frac{\delta}{\delta\phi(\mathbf{x})} + \tfrac{1}{2}\phi(\mathbf{x})\right. \\ &\qquad \left.\cdot(-\nabla^2)\phi(\mathbf{x}) + u(\phi(\mathbf{x}))\right) d^3x. \end{aligned} \tag{3}$$

For (3) to give rise to physically realizable dynamics, the interaction energy density must be such that the eigenvalue spectrum of $\mathbf{H}$ is bounded from below. A stationary state of the quantum field is represented by a wave functional of the form

$$\Psi[\phi; t] = [\exp(-iE_\mu t/\hbar)]U_\mu[\phi],$$

where the ϕ-dependent part satisfies the eigenfunctional equation

$$\mathbf{H}U_\mu[\phi] = E_\mu U_\mu[\phi], \tag{4}$$

with E_μ the constant energy eigenvalue. Physically admissible solutions to Eq. (4) are such that $|U_\mu[\phi]|^2$, the relative probability density for locating the state at the n-tuple $\phi = \phi(\mathbf{x})$, vanishes for unbounded field magnitudes:

$$\lim_{\lambda\to\infty}|U_\mu[\lambda\phi]|^2 = 0, \quad \text{for continuous} \quad \phi = \phi(\mathbf{x}) \not\equiv 0. \tag{5}$$

The vacuum-state eigenfunctional solution to (4) is associated with the energy

$$E_0 \equiv \min_\mu\{E_\mu\};$$

once the vacuum-state eigenfunctional $U_0[\phi]$ has been obtained, the general stationary-state eigenfunctional problem reduces to solving the equation

$$\begin{aligned}\hbar^2\int\sum_{i=1}^{n}\left(-\frac{1}{2}\frac{\delta^2\Omega_\mu}{\delta\phi_i(\mathbf{x})^2} - \frac{\delta(\ln U_0[\phi])}{\delta\phi_i(\mathbf{x})}\frac{\delta\Omega_\mu}{\delta\phi_i(\mathbf{x})}\right) d^3x \\ = (E_\mu - E_0)\Omega_\mu, \end{aligned} \tag{6}$$

for

$$\Omega_\mu = \Omega_\mu[\phi] \equiv U_\mu[\phi]/U_0[\phi],$$

because the left-hand side of (6) gives

$$\begin{aligned}\hbar^2\int\sum_{i=1}^{n}\left(-\tfrac{1}{2}U_0^{-1}\frac{\delta^2U_\mu}{\delta\phi_i(\mathbf{x})^2} + \tfrac{1}{2}U_0^{-2}U_\mu\frac{\delta^2U_0}{\delta\phi_i(\mathbf{x})^2}\right) d^3x \\ = U_0^{-2}(U_0\mathbf{H}U_\mu - U_\mu\mathbf{H}U_0), \end{aligned} \tag{7}$$

with $\mathbf{H}$ prescribed by (3).

Approximate physical solutions to Eq. (4) with the Hamiltonian operator (3) have been obtained for certain generic classes of entire analytic $u(\phi)$ by applying the Rayleigh–Ritz procedure for functionalities.[1] Our purpose here is to report a class of nonlinear field theories for which exact physical solutions to Eq. (4) are obtainable in closed form. These theories feature an interaction energy density that

* Work supported by a National Science Foundation grant.

[1] G. Rosen, Phys. Rev. Letters **16**, 704 (1966); Phys. Rev. **173**, 1680 (1968), and works cited therein.

converges rapidly for large field magnitudes to the form representative of a linear theory of uncoupled scalar fields, a $u = u(\phi)$ expressible as

$$u = \tfrac{1}{2}\sum_{i=1}^{n} m_i^2\phi_i^2 + v, \tag{8}$$

where $v = v(\phi)$ is a continuous real 1-tuple function such that

$$\begin{aligned} \lim_{|\phi|\to\infty}[\phi^2 v(\phi)] &= 0, \quad \text{if} \quad n = 1,\\ \lim_{(\phi\cdot\phi)\to\infty}[(\phi\cdot\phi)(\ln\phi\cdot\phi)v(\phi)] &= 0, \quad \text{if} \quad n = 2,\\ \lim_{(\phi\cdot\phi)\to\infty}[(\phi\cdot\phi)v(\phi)] &= 0, \quad \text{if} \quad n \geq 3. \end{aligned} \tag{9}$$

For a continuous $v(\phi)$ that tends to zero for large field magnitudes with an asymptotic behavior consonant with (9) we have the existence of the function

$$\begin{aligned} f(\phi) &\equiv -\int_{-\infty}^{\infty} |\phi - \phi'|\, v(\phi')\, d\phi', \quad \text{for} \quad n = 1,\\ &\equiv -\frac{1}{2\pi}\int [\ln(\phi-\phi')\cdot(\phi-\phi')]v(\phi')\, d^2\phi', \\ &\qquad\qquad \text{for} \quad n = 2,\\ &\equiv \frac{\Gamma(\frac{1}{2}n-1)}{2\pi^{\frac{1}{2}n}}\int \frac{v(\phi')\, d^n\phi'}{[(\phi-\phi')\cdot(\phi-\phi')]^{\frac{1}{2}n-1}}, \\ &\qquad\qquad \text{for} \quad n \geq 3, \end{aligned} \tag{10}$$

which is the singularity-free solution to the Poisson equation in n-dimensional Euclidean ϕ space

$$\sum_{i=1}^{n}\frac{\partial^2 f(\phi)}{\partial\phi_i^2} = -2v(\phi). \tag{11}$$

The asymptotic behavior of $f(\phi)$ for large field magnitudes follows from (9) and (10) as $f(\phi) \propto |\phi|$ if $n = 1$, $f(\phi) \propto [\ln(\phi\cdot\phi)]$ if $n = 2$, and $f(\phi) \propto (\phi\cdot\phi)^{1-\frac{1}{2}n}$ if $n \geq 3$. In terms of the function (10), the $\mu = 0$ vacuum-state solution to (4) with (8) in (3) is

$$U_0[\phi] = \lim_{\epsilon\to 0+} U_0^{(\epsilon)}[\phi], \tag{12}$$

$$U_0^{(\epsilon)}[\phi] \equiv \exp\left(-\frac{1}{2\hbar}\int\sum_{i=1}^{n}\phi_i(\mathbf{x})(-\nabla^2 + m_i^2)^{\frac{1}{2}}\phi_i(\mathbf{x})\, d^3x - \frac{\epsilon}{\hbar^2}\int f(\phi(\mathbf{x}))\, d^3x\right), \tag{13}$$

where the $\epsilon \to 0+$ limit in (12) is understood to be taken as the final step in a computation involving $U_0[\phi]$, in conjunction with a limit representation of the δ function, $\lim \delta_{(\epsilon)}(\mathbf{x}) = \delta(\mathbf{x})$ as $\epsilon \to 0+$, such that[2]

$$\delta_{(\epsilon)}(\mathbf{0}) = \epsilon^{-1}. \tag{14}$$

[2] For example, we have the wavenumber cutoff representation

$$\delta_{(\epsilon)}(\mathbf{x}) \equiv \int_{|\mathbf{k}|\leq K}(\exp i\mathbf{k}\cdot\mathbf{x})\, d^3k/(2\pi)^3 \quad \text{with} \quad K \equiv (6\pi^2/\epsilon)^{\frac{1}{3}}.$$

In order to verify that (12) is an exact solution to (4), we first compute the functional derivatives of (13):

$$\frac{\delta U_0^{(\epsilon)}[\phi]}{\delta\phi_i(\mathbf{x})} = \left(-\frac{1}{\hbar}(-\nabla^2 + m_i^2)^{\frac{1}{2}}\phi_i(\mathbf{x}) - \frac{\epsilon}{\hbar^2}\frac{\partial f(\phi)}{\partial\phi_i}\bigg|_{\phi=\phi(\mathbf{x})}\right)U_0^{(\epsilon)}[\phi], \tag{15}$$

$$\begin{aligned} &\frac{\delta^2 U_0^{(\epsilon)}[\phi]}{\delta\phi_i(\mathbf{x})^2} \\ &= \left(-\frac{1}{\hbar}[(-\nabla^2 + m_i^2)^{\frac{1}{2}}\delta(\mathbf{x})]_{\mathbf{x}=0} - \frac{\epsilon\delta_{(\epsilon)}(\mathbf{0})}{\hbar^2}\frac{\partial^2 f(\phi)}{\partial\phi_i^2}\bigg|_{\phi=\phi(\mathbf{x})} + \frac{1}{\hbar^2}[(-\nabla^2 + m_i^2)^{\frac{1}{2}}\phi_i(\mathbf{x})]^2 + O(\epsilon)\right)U_0^{(\epsilon)}[\phi], \end{aligned} \tag{16}$$

and then make use of (11) and (14) to obtain, in the limit $\epsilon \to 0+$,

$$\int\left[-\tfrac{1}{2}\hbar^2\sum_{i=1}^{n}\frac{\delta^2 U_0[\phi]}{\delta\phi_i(\mathbf{x})^2} + \left(\tfrac{1}{2}\sum_{i=1}^{n}[(-\nabla^2 + m_i^2)^{\frac{1}{2}}\phi_i(\mathbf{x})]^2 + v(\phi(\mathbf{x}))\right)U_0[\phi]\right]d^3x = E_0 U_0[\phi], \tag{17}$$

where the (unobservable) vacuum-state energy appears as the infinite constant

$$E_0 = \tfrac{1}{2}\hbar\sum_{i=1}^{n}[(-\nabla^2 + m_i^2)^{\frac{1}{2}}\delta(\mathbf{x})]_{\mathbf{x}=0}\int d^3x. \tag{18}$$

It is important to observe that (9) and (10) imply dominance of the summation terms in the exponential (13) over the ϵ-proportional term for all $\phi = \phi(\mathbf{x})$; as a consequence, the terms represented by $O(\epsilon)$ in (16) are uniformly small for all $\phi = \phi(\mathbf{x})$.[3] By substituting (12) into (6), we find the excited-state eigenfunctional equation

$$\lim_{\epsilon\to 0+}\int\sum_{i=1}^{n}\left[-\tfrac{1}{2}\hbar^2\frac{\delta^2\Omega_\mu}{\delta\phi_i(\mathbf{x})^2} + \left(\hbar\phi_i(\mathbf{x})(-\nabla^2 + m_i)^{\frac{1}{2}} + \epsilon\frac{\partial f(\phi)}{\partial\phi_i}\bigg|_{\phi=\phi(\mathbf{x})}\right)\frac{\delta\Omega_\mu}{\delta\phi_i(\mathbf{x})}\right]d^3x = (E_\mu - E_0)\Omega_\mu \tag{19}$$

for $E_\mu > E_0$. Since the solutions to (19) are continuous in ϵ about $\epsilon = 0$, the limit $\epsilon \to 0+$ obtains validity in the equation itself, and the solutions for $\Omega_\mu = \Omega_\mu[\phi]$ are polynomial functionals in the n-tuple of scalar fields. Thus, for the 1-particle and 2-particle states we have

$$\begin{aligned} \Omega_1 &= \int \xi(\mathbf{x})\cdot\phi(\mathbf{x})\, d^3x, \\ &[\hbar(-\nabla^2 + m_i^2)^{\frac{1}{2}} - (E_1 - E_0)]\xi_i(\mathbf{x}) = 0 \end{aligned} \tag{20}$$

[3] That the uniform relative smallness of the ϵ-proportional term in (13) is essential to the solution is easily shown by examples. One such example is given in the Appendix.

and

$$\Omega_2 = \tfrac{1}{2}\int \phi(\mathbf{x}) \cdot \zeta(\mathbf{x}, \mathbf{y}) \cdot \phi(\mathbf{y})\, d^3x\, d^3y - \tfrac{1}{2}\hbar^2(E_2 - E_0)^{-1} \times \int \sum_{i=1}^{n} \zeta_{ii}(\mathbf{x}, \mathbf{x})\, d^3x, \quad (21)$$

$$\zeta_{ij}(\mathbf{x}, \mathbf{y}) \equiv \zeta_{ji}(\mathbf{y}, \mathbf{x}),$$

$$[(-\nabla_x^2 + m_i^2)^{\frac{1}{2}} + (-\nabla_y^2 + m_j^2)^{\frac{1}{2}} - \hbar^{-1}(E_2 - E_0)]\zeta_{ij}(\mathbf{x}, \mathbf{y}) = 0,$$

with higher-order polynomial functional solutions to (19) representing multiparticle states. No interaction between the quanta appears in (21) or in the higher-order polynomial functionals that represent the multiparticle states. Hence, Lagrangian densities of the form (1) with an interaction energy density that satisfies (8) and (9) lead to effectively linear quantum field theories with no interaction and scattering between the quanta. This is in agreement with the Rayleigh–Ritz procedure results for a generic entire analytic Class A interaction energy density in the theories with $n = 1$[4] and $n = 2$.[5]

[4] G. Rosen, Phys. Rev. **160**, 1278 (1967); **E165**, 1934 (1968).
[5] G. Rosen, Phys. Rev. **167**, 1395 (1968).

APPENDIX

It is readily seen by the way of example that the uniform relative smallness of the ϵ-proportional term in (13) is essential to the solution. For instance, consider the specialized form of the quantum Hamiltonian (3) for a simple linear field theory with $n = 1$:

$$\mathbf{H} = \int \left(-\tfrac{1}{2}\hbar^2 \frac{\delta^2}{\delta\phi(\mathbf{x})^2} + \tfrac{1}{2}\phi(\mathbf{x})(-\nabla^2 + m_1^2)\phi(\mathbf{x})\right) d^3x$$

and the functional

$$\Lambda^{(\epsilon)}[\phi] \equiv \exp\left[-\int \left(\frac{1}{2\hbar}\phi(\mathbf{x})(-\nabla^2 + m^2)^{\frac{1}{2}}\phi(\mathbf{x}) + \frac{\epsilon}{12\hbar^2}(m^2 - m_1^2)\phi(\mathbf{x})^4\right) d^3x\right], \quad m^2 \geq m_1^2.$$

Clearly $\lim \Lambda^{(\epsilon)}[\phi]$ as $\epsilon \to 0+$ is an eigenfunctional of $\mathbf{H}$ only if $m^2 = m_1^2$, but a computation that ignores the actual dominance of the $O(\epsilon)$ terms for large field magnitudes would give

$$\lim_{\epsilon \to 0+} \left(\mathbf{H} - \tfrac{1}{2}\hbar[(-\nabla^2 + m^2)^{\frac{1}{2}}\delta(\mathbf{x})]_{\mathbf{x}=0} \int d^3x\right) \Lambda^{(\epsilon)}[\phi]$$

equal to zero for all $m^2 \geq m_1^2$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Remarks on Morse Theory in Canonical Quantization

NORMAN E. HURT
Courant Institute of Mathematical Sciences, New York University, New York, New York 10012
and
Department of Mathematics, University of Massachusetts, Amherst, Massachusetts 01002

(Received 20 October 1968)

The Morse index is studied in the theory of canonical quantization of dynamical systems. Standard results on the Hodge index and vanishing theorems are translated into the language of quantizable dynamical systems.

INTRODUCTION

From asymptotic theory in quantum mechanics it has been noted by several authors[1-5] that the Morse index arises in the same way as the Bohr–Sommerfeld quantum numbers. In this paper it is shown by means of the techniques of canonical quantization,[6,7] that the Morse index occurs naturally. For the classes of homogeneous quantizable dynamical systems, the Morse index can be evaluated.

The language of modern algebraic geometry is natural for the presentation of the results. In Sec. 1 we review the elementary objects of algebraic geometry[8(a)] and Hermitian (Kähler) geometry,[9-11] and we set the vocabulary in terms of the theory of G-structures. Another aim of this paper is to act as a foundation for the treatment of Kodaira–Spencer deformation theory (of G-structures) in the theory of quantizable dynamical systems. (This is completed in Sec. 4.) In Sec. 2 we review the theory of canonical quantization lifted to manifolds modeled on Banach spaces. Here, examples arise from the theory of algebraic Lie groups. We have not gone into detail on blowing up (or compactification), which is a mathematical method for the construction of examples (proofs) in this theory, that has been used recently in the Pham–Leray analysis of Feynman integrals. In Sec. 3, the Morse index is introduced and studied for Hermitian (Kähler) symmetric spaces, which are used in canonical quantization construction. In Sec. 4, the Hodge index and vanishing theorems are presented in the language of quantizable dynamical systems.

1. PRELIMINARY[8-20]

A. Cohomology

Let M be a paracompact manifold (modeled on a Banach space). The cohomology is given by covariant functor $H^*(M, F)$ in sheaf F with $H^0(M, F) \cong \Gamma(M, F)$ = sheaf of sections of M which are valued in F. If F is fine, then $H^q(M, F) = 0$ for all $q \geq 1$. A resolution of sheaf F is the exact sequence of sheaves

$$0 \to F \to A^0 \to A^1 \to \cdots \to A^n \to \cdots, \quad (1.1)$$

such that $H^q(M, A^n) = 0$ for $q \geq 1, n \geq 0$, e.g., when A^n is fine. From (1.1), the induced sequence

$$0 \to \Gamma(M, F) \to \Gamma(M, A^0) \to \Gamma(M, A^1) \to \cdots$$

is called a complex. There is a natural isomorphism

$$H^n(\Gamma(M, A)) \cong H^n(M, F), \quad n \geq 0. \quad (1.2)$$

For the case where M is a C^∞ manifold, modeled on a separable Hilbert space, take for A^p the **R**- (or **C**)-module of $\Lambda^p T^*(M) \to M$, which is the vector

[1] V. P. Maslov, U.S.S.R. Comp. Math. Math. Phys. **1**, No. 1, 123 (1961); No. 2 744 (1961), and references in Ref. 3.
[2] M. V. Fedoryuk, U.S.S.R. Comp. Math. Math. Phys. **3**, No. 1, 162; (1962); **4**, No. 2, 66 (1964).
[3] V. I. Arnol'd, J. Fnal. Anal. Appl. **1**, 1 (1967).
[4] M. C. Gutzwiller, J. Math. Phys. **8**, 1979 (1967).
[5] J. B. Keller, Ann. Phys. **4**, 180 (1958).
[6] N. E. Hurt, Nuovo Cimento **55A**, 534 (1968).
[7] N. E. Hurt, Nuovo Cimento **58B**, 361 (1968).
[8] (a) F. Hirzebruch, *Topological Methods in Algebraic Geometry* (Springer-Verlag, New York, 1966); (b) Symp. Intern. Topological Algebra (Universidad de Mexico, Mexico City, 1956), p. 129; (c) F. Hirzebruch and A. Borel, Am. J. Math. **80** (1958) 458ff.
[9] (a) A. Weil, *Introduction a l'etude des varietes kähleriennes* (Hermann & Cie, Paris, 1958); (b) Am. J. Math. **74**, 865 (1952).
[10] S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic Press Inc., New York, 1962).
[11] I. Goldberg, *Curvature and Homology* (Academic Press Inc., New York, 1962).
[12] J. Eells, Bull. Am. Math. Soc. **72**, 751 (1966).
[13] S. Lang, *Introduction to Differentiable Manifolds* (Interscience Publishers, Inc., New York, 1962).
[14] (a) R. S. Palais, "Lectures on Differential Topology of Infinite Dimensional Manifolds," Brandeis University, Waltham, Mass., 1964; and papers in Topology; (b) *Seminar on the Atiyah–Singer Index Theorem* (Princeton University Press, Princeton, N.J., 1965).
[15] (a) S. Kodaira, Proc. Natl. Acad. Sci. (U.S.) **39**, 1269, 1273 (1953); (b) Ann. Math. **60**, 28 (1954).
[16] (a) R. Bott, Ann. Math. **66**, 203 (1957); (b) **70**, 313 (1959); (c) R. Bott and H. Samelson, Am. J. Math. **80**, 964 (1958); (d) R. Bott and S. Chern, Acta Math. **114**, 71 (1965); B. Kostant, Ann. Math. **74**, 329 (1961).
[17] (a) E. Calabi and E. Vesentini, Ann. Math. **71**, 472 (196); (b) E. Calabi, Am. Math. Soc. Proc. Symp. Pure Math. **3**, 155 (1961).
[18] (a) A. Andreotti and E. Vesentini, Acta Math. **112**, 249 (1964); (b) I.H.E.S. **25** (1965).
[19] (a) P. A. Griffiths, Trans. Am. Math. Soc. **109**, 1 (1963); (b) Acta Math. **110**, 115 (1963); (c) J. Math. Mech. **14**, 117 (1965); (d) Am. J. Math. **88**, 366 (1966).
[20] (a) S. Kobayashi, Tohoku Math. J. **8**, 29 (1956); (b) S. Kobayashi and J. Hano, Trans. Am. Math. Soc. **94**, 233 (1960); (c) S. Kobayashi, Ann. Math. **74**, 570 (1961); (d) S. Kobayashi and T. Nagano, J. Math. Mech. **16**, 875ff (1964).

bundle of p-covectors over M with the fiber being the Banach space of continuous alternating p-linear forms on the tangent space of M. The space $\mathfrak{U}^p = \mathfrak{U}(\Lambda^p T^*(M))$ of C^∞-sections forms a complex $\Gamma(M, A)$ of C^∞ p-forms on M. Let $H^*(\Gamma(M, A)) = \text{Ker } (d)/\text{Im } (d)$, where $d:\mathfrak{U}^p \to \mathfrak{U}^{p+1}$. Then (1.2) is just the generalized deRham isomorphism

$$H^*(M, \mathbf{R} \text{ (or } \mathbf{C})) \cong \text{Ker } (d)/\text{Im } (d). \tag{1.3}$$

B. G-structures

Let M be a C^∞ manifold modeled on a Banach space $\mathbf{E}$. Let $GL(\mathbf{E}) \to B \to M$ be the principal bundle of the tangent bundle $T(M)$ of M, where, geometrically, B is the bundle of all tangent frames on M. Let B/G be the bundle $B \times_G GL(\mathbf{E})$, where $G = G(\mathbf{E})$ is a closed linear subgroup of $GL(\mathbf{E})$, with fiber $GL(\mathbf{E})/G$. A *G-structure* on M is a cross section $s:M \to B/G$, i.e., M has the G-structure iff the structural group of tangent bundles of M can be reduced to G.

C. Examples

M has the *Sp*-structure iff M is a *Hamiltonian dynamical system*; i.e., M has the antisymmetric nondegenerate bilinear form (2-form Ξ). Let $i(X)$ be the inner product; then the 1-parameter (pseudo-) group H_t of diffeomorphisms of M with *Sp*-structure is called Hamiltonian iff $i(X)\Xi = \omega_0$ for some closed 1-form ω_0, where X is the infinitesimal generator of H_t. By the Cartan identity, $\pounds(X) = i(X)\,d + di(X)$, H_t is Hamiltonian iff $\pounds(X)\Xi = 0$, where $\pounds$ is the Lie derivative. The sheaf of germs of the Hamiltonian vector fields forms a Lie-algebra sheaf with respect to the Lie bracket.[21] This is naturally interpretable in terms of the Poisson bracket.

M has the *SU*-structure iff M is *almost complex*. Furthermore, M is torsionless iff M admits the Kähler metric whose Ricci tensor vanishes (see below). Alternatively, M is almost complex if it carries a real C^∞ tensor field J of type (1, 1) such that $J^2 = -I$.

M has the $\mathbf{1} \times U$-structure iff M is an *almost contact* manifold (cf. Refs. 22–24).

M has the O-structure iff M has the C^∞ *Riemann structure* to be defined. Let $\xi:E \to M$ be the C^∞ vector bundle whose fibers are Hilbert spaces. For the vector bundle $SL^2(E, \mathbf{R}) \to M$ whose fiber over any point m of M is space of continuous symmetric bilinear forms on fiber $\xi^{-1}(m) = E_m$, a C^∞ *Riemann structure* on ξ is a C^∞ section of $SL^2(R, \mathbf{R}) \to M$ which is positive definite at every point m and which, since it is the inner product on E_m, determines its Hilbert-space structure.[12–14(a)] If M has a continuous assignment of a norm to each tangent space compatible with its Banach-space structure, locally uniform on M, then M is called a Finsler manifold.[14(a)] Finsler structures arise naturally in fibration of dynamical systems.[25] Taking M to be modeled on a separable Hilbert space $\mathbf{E}$, a *Riemann structure* on M is a C^∞ Riemann structure on the tangent bundle $T(M) \to M$.

M has U-structure iff M is *Hermitian*. Alternatively, M is Hermitian if M (connected) is almost complex and there is a real bihomogeneous, everywhere-positive, and nondegenerate (1, 1)-form Ξ on $T(M)$, which, when identified with its projection on M, is given locally in the family $\omega^1, \cdots, \omega^n$ of (1, 0)-forms by

$$\Xi = \tfrac{1}{2}(-1)^{\frac{1}{2}} \sum h_{\alpha\beta}\omega^\beta \wedge \bar{\omega}^\alpha, \tag{1.4}$$

where $h_{\alpha\beta}$ is $\mathbf{C}$-valued, of class C^∞, and where $h_{\alpha\beta} = \bar{h}_{\beta\alpha}$. The *fundamental* 2-form is called Ξ and it determines a C^∞ Riemann structure on M,

$$ds^2 = \sum h_{\alpha\beta}\bar{\omega}^\alpha \cdot \omega^\beta, \tag{1.5}$$

for which J is an isometry. If $d\Xi = 0$ and if the almost-complex structure is integrable, then M is called *Kähler*. Note that if the symplectic manifold M admits a Riemann structure h such that J is an isometry, then M is Hermitian, admits Ξ as the fundamental 2-form, and J is determined by h, $\Xi:h(X, Y) = \Xi(X, JY)$.

D. Metric and Curvature

Given any holomorphic (or C^∞) vector bundle $\xi:E \to M$, we may define a metric or norm on E, i.e., a positive-definite Hermitian scalar product $h(u, v)$: $E \to \mathbf{R}$, where $u, v \in \xi^{-1}(m)$, which is a Hermitian metric on each fiber and depends analytically (or C^∞) on $m \in M$. Any real Hermitian (1, 1) form A [$\in A^{1,1}(\text{End } (E))$, see below] which is positive definite at each point defines a Hermitian metric on E. Locally on the covering $\{U_\alpha\}$ of M, let h_U be the representation of $h(u, v)$. In terms of h_U, we define the connection and curvature of the vector bundle.

More generally,[26] for any holomorphic (or C^∞) G-bundle E over M,

$$G \to E \to M, \tag{1.6}$$

[21] I. M. Singer and S. Sternberg, J. Anal. Math. **15**, 1 (1965).
[22] (a) W. M. Boothby and H. C. Wang, Ann. Math. **68**, 721 (1958); (b) W. M. Boothby, Am. Math. Soc. Proc. Symp. Pure Math. **3**, 144 (1961); (c) Proc. Am. Math. Soc. **13**, 276 (1962).
[23] J. W. Gray, Ann. Math. **99**, 421 (1959).
[24] S. Takizawa, Tohoku Math. J. **15**, 227 (1963).
[25] (a) G. Reeb, Colloq. Top. Strasbourg (1951); Compt. Rend; (b) Acad. Roy. Bel. Class Sci. Mem. **27**, No. 9 (1952).
[26] M. F. Atiyah, Trans. Am. Math. Soc. **85**, 181 (1957).

let $Q(E)_m = T(E)/G$ be the space of vector fields on E_m which are tangent to E and invariant by G; let $\mathbf{g}_m$ be the usual representation of Lie algebra as vector fields on E_m invariant by G, which is the bundle associated to E by adjoint representation of G on $\mathbf{g}_m$, $L(E) = E \times_{\mathrm{ad}\, G} \mathbf{g}$. From (1.6) we have the exact sequence of (holomorphic) vector bundles over M,

$$0 \to \mathbf{g}_m \to Q(E)_m \to T(M)_m \to 0,$$

which is equivalent to

$$0 \to L(E) \to Q(E) \xrightarrow{\lambda} T(M) \to 0. \tag{1.7}$$

The holomorphic (or C^∞) *connection* on E is a homomorphism $\omega\colon T(M) \to Q(E)$ such that $\omega \circ \lambda = \mathbf{1}$. The *curvature* form of connection ω is a $\mathbf{g}$-valued 2-form on E given by

$$\Xi = d\omega + \tfrac{1}{2}[\omega, \omega] \in H^1(M, T(M) \otimes_{\mathcal{O}} L(E)). \tag{1.8}$$

Sequence (1.7) gives the exact sequence of vector bundles

$$0 \to \mathrm{Hom}\,(T, L) \xrightarrow{\rho} \mathrm{Hom}\,(Q, L) \xrightarrow{\lambda} \mathrm{End}\,(L) \to 0$$

and also the associated exact cohomology sequence

$$\xrightarrow{\rho^*} H^0(M, Q \otimes_{\mathcal{O}} L) \to H^0(M, L \otimes L)$$
$$\xrightarrow{\delta^*} H^1(M, T \otimes_{\mathcal{O}} L)\colon \quad \omega \mapsto \mathbf{1} \mapsto \delta^*\mathbf{1} = c(E), \tag{1.9}$$

where $\mathcal{O} = \Omega^0$ is the sheaf of germs of holomorphic functions on M and $\Omega^1 = T^*(M)$. Thus, the obstruction to the existence of a connection is the class of extensions of holomorphic fibrations $c(E)$. In the C^∞ case, this is always trivial.

For the above case, $E \to M$, which is a Hermitian ($\mathbf{C}^*$-) bundle [L is a trivial bundle, $\Xi(E) \in H^1(M, \Omega^1)$], and $c(E) = [(-1)^{\frac{1}{2}}/2\pi]\Xi(E)$ is representative; locally, the connection of the principal bundle associated with E is a system of $(1,0)$-forms $\omega_U = h_U^{-1}\partial h_U$. The curvature form of this connection is the system of $(1,1)$ forms

$$\Xi_U = \bar\delta\omega_U = \bar\delta(h_U^{-1}\partial h_U),$$

which is the obstruction of this connection to be holomorphic. In components, we have $\Xi^\rho_\alpha = R_{\alpha\beta\bar\gamma} \times dz^\beta \wedge dz^{\bar\gamma}$ and $[2\pi(-1)^{\frac{1}{2}}]^{-1}R_{\alpha\bar\beta}\, dz^\alpha \wedge d\bar z^\beta$ is representative of the first Chern class, where $R_{\alpha\bar\beta} = -R^\rho_{\alpha\beta\rho}$ is the Ricci tensor with

$$(2\pi(-1)^{\frac{1}{2}})^{-1} \sum R_{\alpha\bar\beta}\, dz^\alpha \wedge d\bar z^\beta = -[2\pi(-1)^{\frac{1}{2}}]\partial\bar\partial \log\det h_{\alpha\bar\beta}.$$

If the tangent bundle $T(M)$ is given a Hermitian metric h, then h induces a Hermitian metric on M, usually represented as

$$ds^2 = 2\sum h_{\alpha\bar\beta}\, dz^\alpha \cdot d\bar z^\beta.$$

The corresponding ∂-connection is given by

$$\omega = \sum_{\bar\gamma} h^{\beta\bar\gamma}\partial h_{\bar\gamma\alpha} = \sum C^\beta_{\alpha\gamma}\, dz^\gamma,$$

the curvature by

$$\Xi = \bar\delta\omega^\beta_\alpha = \frac{\partial C^\beta_{\alpha\rho}}{\partial \bar z^\sigma}\, d\bar z^\sigma \wedge dz^\rho,$$

and the torsion tensor by

$$S^\beta_{\alpha\gamma} = \tfrac{1}{2}\{C^\beta_{\alpha\gamma} - C^\beta_{\gamma\alpha}\};$$

thus $d\omega = (-1)^{\frac{1}{2}}\,\mathrm{Re}\,({}^tS \wedge \bar\omega)$. S vanishes iff the Hermitian metric is Kähler.

E. Sheaves

Let Ω^p, $\mathfrak{U}^{p,q}$ be sheaves of germs of holomorphic p-forms [respectively, C^∞ (p,q) forms] on M. Recall that d splits as $d = \partial + \bar\delta$ and the kernel of $\bar\delta\colon \mathfrak{U}^{p,q} \to \mathfrak{U}^{p,q+1}$ is the sheaf Ω^p. The sheaf of germs of holomorphic sections of $E \to M$ is a coherent, locally free, analytic sheaf. The sheaf of germs of holomorphic p-forms on M, $\Omega^p(E)$, with coefficients in E, is the sheaf of germs of holomorphic sections of $E \otimes \Lambda^p T^*$,

$$\Omega^p(E) = \Omega^p \otimes E \otimes \Lambda^p T^* = \Omega(E \otimes \Lambda^p T^*),$$

where $\otimes = \otimes_{\mathcal{O}}$. Similarly,

$$\mathfrak{U}^{p,q}(E) = \mathfrak{U}^{p,q} \otimes E \otimes \Lambda^p T^* \otimes \Lambda^q \bar T^*.$$

The exact sequence

$$0 \to \Omega^p(E) \to \mathfrak{U}^{p,0}(E) \to \mathfrak{U}^{p,1}(E) \to \cdots$$

is due to Grothendieck and Dolbeault. Let

$$A^{p,q}(E) = H^0(M, \mathfrak{U}^{p,q}(E)).$$

The Dolbeault–Serre isomorphism is

$$H^{p,q}(M, E) = H^q(M, \Omega^p(E)) \cong H^*(\Gamma(M, A^{p,q}(E)) = \frac{Z^{p,q}(E)}{\bar\delta A^{p,q-1}(E)},$$

where the cohomologies are given with respect to an arbitrary family Φ of closed or compact subsets of M. Recall that if M is Stein, then $H^q(M, \Omega^p(E)) = 0$ for all $q \neq 0$.

Assume that $E \to M$ has a positive-definite Hermitian scalar product [which is the case, e.g., if $T(M)$ has a Hermitian metric]. We review briefly the standard operations:

(i) exterior multiplication by curvature Ξ, where $e(\Xi) = \Xi \wedge \varphi$ for $\varphi \in A^{p,q}(E)$, is given by

$$e(\Xi)\colon A^{p,q}(E) \to A^{p+1,q+1}(E);$$

(ii) anti-isomorphism for E to dual E^* is given by

$$\#: A^{p,q}(E) \to A^{q,p}(E^*)$$

for $\#\varphi_U = h_U \bar{\varphi}_U$;

(iii) star isomorphism is given by

$$*: A^{p,q}(E) \to A^{n-q,n-p}(E);$$

(iv) covariant differentiation on local sections of E is given by $D = D' + \bar{\delta}$, where $D' = \partial + e(\omega)$; the Ricci identity is $D^2 = (D'\bar{\delta} + \bar{\delta}D') = e(\Xi)$; the formal adjoint $\mathfrak{D} = -*D'*$ gives $\square = \bar{\delta}\mathfrak{D} + \mathfrak{D}\bar{\delta}$ (see below); so there is differential operator

$$\partial\bar{\delta} + \bar{\delta}\partial: A^{p,q}(E) \to A^{p+1,q+1}(E),$$

where $(\partial\bar{\delta} + \bar{\delta}\partial)\varphi = e(\Xi)\varphi$ for $\varphi \in A^{p,q}(E)$;

(v) the formal adjoint to d is $\delta = -*d*$; let $D = d + \delta$ (elliptic), then the *Laplacian* adjoint is defined by $D^2 = \Delta = \delta d + d\delta$ (strongly elliptic), where $\delta = \mathfrak{D} + \overline{\mathfrak{D}}$; $\mathfrak{D}$ is formal adjoint to $\bar{\delta}$, where $\mathfrak{D} = -*\delta* = -*\#\bar{\delta}\#*$ and is defined by

$$\mathfrak{D}: A^{p,q+1}(E) \to A^{p,q}(E);$$

let $D = \bar{\delta} + \mathfrak{D}$ (elliptic) then the *Laplace–Beltrami* operator (strongly elliptic) is given by

$$D^2 = \square = \bar{\delta}\mathfrak{D} + \mathfrak{D}\bar{\delta}: A^{p,q}(E) \to A^{p,q}(E);$$

an (almost) Hermitian manifold is called *pseudo-Kähler* if $\square = 2\Delta$; these operators are formal adjoints with respect to the scalar product (1.10) below.

For $\varphi, \psi \in A^{p,q}(E)$,

$$\langle \varphi, \psi \rangle \, dM = {}^t\varphi \wedge *\#\psi$$

is a **C**-valued Hermitian scalar form, where dM is the volume element of ds^2. With respect to the compact family Φ on M,

$$(\varphi, \psi) = \int_M \langle \varphi, \psi \rangle \, dM \tag{1.10}$$

is finite and defines a complex pre-Hilbert structure on $A^{p,q}$. Let $A^{p,q}$ also denote the completion with respect to norm $\|\varphi\| = (\varphi, \varphi)^{\frac{1}{2}}$.

The product $\wedge$ gives the map

$$A^{p,q}(E) \times A^{r,s}(E^*) \to A^{p+r,q+s}(\mathbf{1}).$$

By the Dolbeault–Serre isomorphism, the inner product (1.10) induces the *Serre duality*

$$H^{p,q}(M, E) \times H^{n-p,n-q}(M, E^*) \to \mathbf{C}.$$

The subspace $\mathcal{H}^{p,q}(E)$ of elements $\varphi \in A^{p,q}(E)$ for which $\square\varphi = 0$ is called the space of *complex harmonic* forms. Since $\square\varphi = 0$ iff $\bar{\delta}\varphi = \mathfrak{D}\varphi = 0$, by the Dolbeault–Serre isomorphism we have the Kodaira isomorphism

$$H^{p,q}(M, E) \cong \mathcal{H}^{p,q}(M, E).$$

Kodaira proved that $\mathcal{H}^{p,q}(M, E)$, hence $H^{p,q}(M, E)$, is finite dimensional, since $\square$ is a strongly elliptic operator.

F. Kähler Metric

If a Hermitian metric on the tangent bundle of M is a Kähler metric, i.e., ds^2 is a Kähler metric, then the fundamental (1, 1)-form Ξ represents an element of $H^2(M, \mathbf{R})$ and, in terms of Ξ, two more operations are defined as follows:

$$L: A^{p,q}(E) \to A^{p+1,q+1}(E),$$

given by $L\varphi = \Xi \wedge \varphi$, and

$$\Lambda = *^{-1}L*,$$

which is the formal adjoint of L. For M compact, by the Kodaira isomorphism we have that L is multiplicative on harmonic forms.

2. REVIEW OF CANONICAL QUANTIZATION

A. Quantizable Dynamical Systems

Let M be a C^∞ or a complex analytic manifold (suitably modeled), let Ξ be a C^∞ or analytic symplectic 2-form on M, and let H_t be a 1-parameter (pseudo-) group of diffeomorphism such that $\pounds(X)\Xi = 0$, where X is the infinitesmal generator of H_t. The triple (M, Ξ, H_t) is called a *dynamical system* (developed by Poincaré, Birkhoff, and Reeb[25(b)]). From Refs. 6 and 7, *canonical quantization* is the construction of the exact sequence

$$T \to E \to M, \tag{2.1}$$

where $E(M, T)$ is the principal toral T bundle over M which carries a "regular" (or "dynamic") contact form. A contact form on $E(M, T)$ is called "dynamic" if it defines a connection on $E(M, T)$. A dynamical system for which (2.1) holds is called a *quantizable dynamical system* (QDS).

B. Review

We review the following facts.[22–24] From exact sequences of Abelian groups

$$0 \to \mathbf{Z} \xrightarrow{\lambda} \mathbf{R} \to \mathbf{S}^1 = \mathbf{R}^+ \to 0,$$
$$0 \to \mathbf{Z} \to \mathbf{C} \to \mathbf{C}^* \to 0, \tag{2.2}$$

there are exact sequences of sheaves of Abelian groups

$$0 \to \mathbf{Z} \to R = \mathfrak{U}(\mathbf{R}) \to S^1 = \mathfrak{U}^*(\mathbf{S}^1) \to 0, \tag{2.3}$$
$$0 \to \mathbf{Z} \to C = \Omega(\mathbf{C}) \to C^* = \Omega^*(\mathbf{C}^*) \to 0 \tag{2.4}$$

for sheaves of germs of C^∞ (respectively, holomorphic) functions on M, valued in $\mathbf{Z}, \mathbf{R}, \mathbf{S}^1$ (respectively, $\mathbf{Z}$, $\mathbf{C}$, $\mathbf{C}^*$). In associated cohomology sequences, $H^1(M, G^*$ (respectively, $\mathbf{G}^*$)) represents the isomorphism classes of G^*- (respectively, flat G^*-) bundles over M. Since R is a fine sheaf, the cohomology sequence associated with (2.3),

$$0 \to H^1(M, S^1) \xrightarrow{\delta^*} H^2(M, \mathbf{Z}) \to 0,$$
$$\xi \mapsto \chi(\xi),$$

gives a bijection of representative ξ and class $\delta^*\xi = \chi(\xi)$, that is, the Euler–Poincaré class of the $\mathbf{S}^1$-bundle ξ. Then the constant-coefficient (flat) sequence

$$H^1(M, \mathbf{S}^1) \to H^2(M, \mathbf{Z}) \to H^2(M, \mathbf{R}),$$
$$\chi(\xi) \mapsto \Xi, \tag{2.5}$$

associates $\chi(\xi)$ with the curvature form Ξ of connection on the $\mathbf{S}^1$-bundle structure. [Recall that the Cartan structure equation (1.8) reduces to $d\omega = \Xi$ for the Abelian $\mathbf{S}^1$ case and that we identify Ξ with its projection on M.] Thus, by the deRham isomorphism (1.10), Ξ from (2.5) represents a cohomology class, that is, $\Xi \in H^2(\Gamma(M, \mathfrak{U})) \cong H^2(M, \mathbf{R})$. The converse is evident. Thus:

Proposition 1: There exists a canonical quantization for a C^∞ dynamical system (M, Ξ, H_t), i.e., a C^∞ toral bundle $E(M, T)$ over M, and a C^∞ dynamic contact form ω on E such that $d\omega = \Xi$, iff Ξ represents an integral cohomology class of M.

C. Existence of Canonical Quantization

In the complex analytic case, the cohomology sequence from (2.4),

$$\xrightarrow{\exp^*} H^1(M, C^*) \xrightarrow{\delta^*} H^2(M, \mathbf{Z}) \xrightarrow{\lambda^*} H^2(M, C),$$
$$\xi \mapsto c_1(\xi), \tag{2.6}$$

maps the representative ξ of an analytic $\mathbf{C}^*$-bundle structure to the class $\delta^*\xi = c_1(\xi)$, which is the Chern class of ξ. For a C^∞ dynamic contact structure with the curvature form $\Xi = d\omega$, the form $\Theta = (2\pi(-1)^{\frac{1}{2}})\Xi$ is a complex-valued closed (1, 1)-form which represents the Chern class $\delta^*\xi$ in the sense of the deRham isomorphism, that is,

$$\Theta \in H^2(M, \mathbf{Z}) \to H^2(M, C) \cong H^2(\Gamma(M, \mathfrak{U}, \mathbf{C})).$$

In this case, $\lambda^*\Xi$ is represented by the (0, 2) component Θ^{02} in sense of Dolbeault and Kodaira, namely

$$\Theta^{02} \in H^2(M, C) \cong H^2(\Gamma(M, \mathfrak{U}, \mathbf{C})).$$

From Sec. 1C [cf. Refs. 15(a), 8(a), and 19(d)] for any real (1, 1)-form representing the (1, 1) cohomology class $\delta^*\xi \in H^2(M, \mathbf{Z})$, there exists a C^∞ $\mathbf{C}^*$-bundle E such that $c_1(E) = \delta^*\xi$ and conversely. Furthermore, given a C^∞ $\mathbf{C}^*$-bundle E represented by ξ and a C^∞ connection form θ on E, then there exists a global 1-form $\varphi \in \mathfrak{U}^1(M, \mathbf{C})$ on M such that $d\varphi = \Xi - d\theta$ [since closed 2-forms Ξ and $d\theta$ represent the same cohomology class $2\pi(-1)^{\frac{1}{2}}c_1(\xi) \in H^2(M, \mathbf{C})$]. Thus, $\omega = \theta + \varphi \in \mathfrak{U}^1(E, \mathbf{C})$ is a connection in E whose curvature $d\omega = \Xi$ is of type (1, 1), which means that E is a holomorphic bundle and that ω is a C^∞ dynamic contact structure on E. Ξ is holomorphic if $\Xi^{11} = \Xi^{02} = 0$ and $d''\Xi = 0$ (which implies $d''\omega = \Xi^{11} = 0$). In summary, we have the following proposition.

Proposition 2. There exists a canonical quantization for the complex analytic dynamical system (M, Ξ, H_t) iff $(2\pi(-1)^{\frac{1}{2}})^{-1}\Xi$ determines an integral cocycle on M. If $H^1(M, C) = 0$, e.g., if M is a Stein manifold, then ω is a global dynamic contact form on E.

D. Abelian Varieties

The standard examples are Abelian varieties. [Note that $H^1(M, C^*)$ is the class of divisors with respect to linear equivalence.] A complex manifold, where $\dim_{\mathbf{C}} M = n$, is called (complex) parallelizable, if there exists n globally defined linearly independent holomorphic vector fields X_i on M. A compact (complex) parallelizable manifold has the complex Lie group as the universal covering space of M; so M is holomorphically isomorphic to $M/\mathfrak{D}$, where $\mathfrak{D}$ is the fundamental group of M. Thus M is isomorphic to the complex (compact) quotient space $G^{\mathbf{C}}/H$, where $G^{\mathbf{C}}$ is the complex Lie group and H is the discrete subgroups; that is, the connected component H^0 of the identity is invariant by $G^{\mathbf{C}}$.[27–29] If the Lie algebra generated by the X_i is Abelian, then M is Kähler with a locally flat metric. Let V be a complex vector space with real form V_0, with $\dim_{\mathbf{R}} V_0 = 2n$. Let $\mathfrak{Z}$ be the lattice of V, i.e., the discrete additive subgroup of V_0 with the $2n$ generators of V_0. Then $V/\mathfrak{Z}$ is a complex torus [locally flat, (complex) parallelizable]. A compact (complex) parallelizable manifold is Kähler iff it is a complex ("multi"-) torus. According to Tits,[29] if $G^{\mathbf{C}}/H$ is compact, $N(H_0)$ is a parabolic subgroup of $G^{\mathbf{C}}$ and there is one and only one fibration

$$N(H_0)/H \to G^{\mathbf{C}}/H \to G^{\mathbf{C}}/N(H_0)$$

with complex parallizable fiber over a flag manifold; when H is connected, this reduces to Wang's theorem (cf. Sec. 3C).

[27] (a) H. C. Wang, Am. J. Math. **76**, 1 (1954); (b) Proc. Am. Math. Soc. **5**, 771 (1954).
[28] J. P. Serre, Sem. Boubaki, Exp. **100** (1953/54).
[29] J. Tits, Commun. Math. Hel. **37**, 111 (1962).

A complex torus whose rank is equal to its complex dimension is called an *Abelian variety*. This is equivalent to the existence of a Riemann form on the torus, i.e., the existence of a nondegenerate Hermitian form whose imaginary part is integral on $\mathfrak{Z}$. A complex torus is an Abelian variety iff it is a Hodge variety. The complex torus with Riemann form is a nonsingular algebraic variety embedded in complex projective space, where the imbedding isomorphism is given by the theta function. In the future, polarized Abelian varieties, Picard varieties, duality, cusp forms, etc. will be studied in the context of QDS.[30]

3. MORSE INDEX[10,12,14,16,19,31-33]

A. Definition

Let M be a C^∞ (complete Riemann) manifold modeled on a Banach space (e.g., a separable Hilbert space). Let $f: M \to \mathbf{R}$ be a differentiable function defined on M. A *critical point* of f is a point $m \in M$ such that the differential $Xf = df_m(X) = 0$, for $df_m \in T(M)^*_m$, where X is the directional derivative in $T(M)_m$. If M admits a Riemann structure (norm) $\langle\ ,\ \rangle$, then $df_m = \langle \nabla f_m, \rangle$, where $\nabla f \in \mathfrak{U}(T(M))$ is the gradient field of f. Also, m is a critical point, if, for trajectory $\gamma_t(m)$,

$$\frac{d}{dt}\gamma_t(m)\bigg|_{t=0} = -\nabla f(\gamma_t(m))\bigg|_{t=0} = 0.$$

In terms of the connection ω,

$$\nabla\gamma = D'\gamma = \partial\gamma + \omega \wedge \gamma.$$

A critical point m is *nondegenerate* if the Hessian $d^2f(m)$ is a nondegenerate bilinear form. The *Morse index* of m is the supremum of dimensions of linear subspaces of $T(M)_m$ on which $d^2f(m)$ is negative definite.

Let f be the energy function

$$E(\gamma) = \tfrac{1}{2}\int_0^1 \|\partial\gamma\|^2\, dt = \tfrac{1}{2}\|\partial\gamma\|_0^2,$$

where γ is the path $[0, 1] \to M$ [space of paths $= \mathfrak{S}(M)$]. γ is a critical point of $E(\gamma)$, $dE(\gamma) = 0$, iff γ is a geodesic, that is, $\nabla\partial\gamma = 0$. Define the linear transformation, in terms of the curvature tensor R, as

$$K_\gamma(\xi) = R(\partial\gamma, \xi)\partial\gamma, \tag{3.1}$$

where

$$v = \frac{d}{dt}\gamma(0) = \partial\gamma(0)$$

is the velocity vector and ξ (parallel) $\in T\mathfrak{S}(M)_\gamma$ along γ. Then the Hessian of $E(\gamma)$ is

$$d^2E(\gamma)(\xi, \eta) = \langle \nabla\xi, \nabla\eta\rangle_0 + \langle \partial\gamma, R\circ(\xi, \partial\gamma, \eta)\rangle_0 = \langle A_\gamma \circ \xi, \eta\rangle,$$

where $A_\gamma = 1 + (1 - \Delta)^{-1} \circ (K_\gamma - 1)$ is a self-adjoint Fredholm operator and $\Delta = \nabla^2$ is the Laplacian operator. There is a decomposition $T_\gamma = T^0_\gamma + T^-_\gamma + T^+_\gamma$ of eigenspaces of A_γ corresponding to zero, negative, and positive eigenvalues λ, respectively; these are

$$T^\pm_\gamma = \operatorname{Ker}(-\Delta + K_\gamma),$$
$$T_\gamma = \sum_{\lambda \gtrless 0} \operatorname{Ker}((\lambda - 1)\Delta + K_\gamma - \lambda), \tag{3.2}$$

where Nullity $(\gamma) = \dim T^0_\gamma$ and M-index$(\gamma) = \dim T^-_\gamma$ are finite since $(\lambda - 1)\Delta + K_\gamma - \lambda$ is an elliptic differential operator. The M-index(γ) of $d^2E(\gamma)$ is equal to the number of points $\gamma(t)$, with $0 < t < 1$, such that $\gamma(t)$ is conjugate to $\gamma(0)$ along γ, where each one is counted with its multiplicity; in terms of Jacobi fields,

$$\nabla^2 J + R(v, J)v = (\Delta - K_\gamma)J = 0.$$

More generally, a closed connected submanifold N of M is a *nondegenerate critical manifold* of f if every $n \in N$ is a critical point of f whose tangent space $T(N)_n \subset T(M)_n$ coincides with the null space of the Hessian of f at n.

B. Decompositions

For (Hermitian) Riemann symmetric (involutive automorphism) spaces M, there is considerable simplification. The conjugate points are the points $\gamma(\pi n/(e_i)^{\frac{1}{2}})$, where n is a nonzero integer and e_i is any positive eigenvalue of K_γ.[31] If we realize M as the coset spaces of Lie groups, then every point of M lies on a maximal torus, maximal tori are geodesically imbedded in M, and all maximal tori are "conjugate" to each other by inner automorphisms. For Hermitian (therefore, Kähler) homogeneous symmetric spaces, the coset spaces of (semi-) simple Lie groups are of two types (in the irreducible case):

(A) G is a noncompact simple Lie group with center the identity and K is a maximal-connected compact subgroup (a one-dimensional center), so that G/K is complex and analytically equivalent to the irreducible bounded homogeneous symmetric domains of Cartan,

(B) G is a compact group and K is a subgroup of maximal rank (a one dimensional centralizer). These

[30] P. Cartier, Am. Math. Soc. Proc. Symp. Pure Math. 9, 361 (1966).
[31] J. Milnor, *Morse Theory* (Princeton Univ. Press, Princeton, N.J., 1963).
[32] (a) H. I. Eliasson, J. Diff. Geom. 1, 169 (1967); (b) Preprint.
[33] M. Takeuchi, J. Fac. Sci. Univ. Tokyo 12, 81 (1965).

spaces were classified by Cartan as follows: $\mathrm{BI}_{m,m'}$ $U(m+m')/U(m)\times U(m')$, BII_m $SO(2m)/U(m)$, BIII_m $Sp(m)/U(m)$, BIV_m $SO(m+2)/SO(m)\times(SO(2)=T^1)$, BV $E_6/\mathrm{Spin}(10)\times(SO(2)=T^1)$, BVI $E_7/E_6\times(SO(2)=T^1)$. These spaces are rational algebraic varieties. Spaces of type A can be imbedded naturally in open subsets of spaces of type B and there is a one-to-one correspondence between the two types.[3]

Let M be a symmetric space of noncompact type. M is generated by products of geodesic isometries. The largest connected group of isometries $I_0(M)=G$ is semisimple and contains no compact normal subgroups except for $\{e\}$. Let K denote the isotropy group of G. Let $\mathbf{g}_0$, $\mathbf{k}_0$ denote the corresponding Lie algebras and let $\mathbf{m}_0$ be the orthogonal complement of $\mathbf{k}_0$ in $\mathbf{g}_0$ with respect to the Killing form B of $\mathbf{g}_0$, that is, $\mathbf{g}_0=\mathbf{k}_0+\mathbf{m}_0$, where B is strictly positive definite on $\mathbf{m}_0$ and negative definite on $\mathbf{k}_0$. Let $\exp:\mathbf{g}_0\to G$, $\mathrm{Exp}:T(M)_0\to M$ (diffeomorphism), and $\pi:G\to G/K$, thus, $d\pi:\mathbf{m}_0\xrightarrow{\sim}T(M)_0$ and $\pi(\exp X)=\mathrm{Exp}\,(d\pi(X))$ for $X\in\mathbf{m}_0$. Let $\mathbf{g}$ denote the complexification of $\mathbf{g}_0$ with Lie group $G^{\mathbf{C}}$, where $\mathbf{g}_0$ is noncompact form of $\mathbf{g}$. $G^{\mathbf{C}}$ (center reduced to identity) has a connected closed complex subgroup U with no closed normal nondiscrete subgroups of $G^{\mathbf{C}}$. $M^u=G^u/U$ is a complex compact symmetric Hermitian space (the compact form or Cartan dual of M). $\mathbf{g}_0^u=\mathbf{k}_0+(-1)^{\frac{1}{2}}\mathbf{m}_0$ is the compact real form of $\mathbf{g}$. Thus $M^u=G^u/K$, $K=G^u\cap U$.

If $l=\mathrm{rank}\,M$, then M contains flat, totally geodesic submanifolds of dimension l, which are called *hyperplanes*. Let P be the hyperplane containing origin 0 and let $\mathbf{h}_{\mathbf{m}_0}$ be the maximal Abelian subalgebra of vectors $X\in\mathbf{m}_0$ for which $\mathrm{Exp}\,(d\pi(X))\subset P$. The *Cartan* subalgebra of M is $\mathbf{h}_{\mathbf{m}_0}$. Let $A_{\mathbf{m}_0}$ be the analytic subgroup corresponding to $\mathbf{h}_{\mathbf{m}_0}$. An element $H\in\mathbf{h}_{\mathbf{m}_0}$ is called *regular* if its centralizer $\mathbf{z}_H$ in $\mathbf{m}_0$ equals $\mathbf{h}_{\mathbf{m}_0}$. The *critical* or *singular* elements are those H for which $\dim\mathbf{z}_H>\dim\mathbf{h}_{\mathbf{m}_0}$. Let $\mathbf{h}_{\mathbf{m}_0}^*$ denote the set of regular elements in $\mathbf{h}_{\mathbf{m}_0}$. The connected components of $\mathbf{h}_{\mathbf{m}_0}^*$ are called the *Weyl chambers* of $\mathbf{h}_{\mathbf{m}_0}$. Fix a Weyl chamber D in $\mathbf{h}_{\mathbf{m}_0}$ and let Δ be the set of linear maps $\alpha:\mathbf{h}_{\mathbf{m}_0}\to\mathbf{C}$. By the transformation exp linear forms α are in one-to-one correspondence with characters $\mathrm{Hom}\,(A_{\mathbf{m}_0},\mathbf{C}^*)$, and $\mathrm{Hom}\,(A_{\mathbf{m}_0},\mathbf{C}^*)\cong H^1(A_{\mathbf{m}_0}^{\mathbf{R}},\mathbf{Z})$ by $\alpha\to(2\pi(-1)^{\frac{1}{2}})^{-1}\alpha$. The dual space of $\mathbf{h}_{\mathbf{m}_0}$ is

$$\Delta_{\mathbf{R}}=\{\alpha\in\Delta\mid\alpha(\mathbf{h}_{\mathbf{m}_0})\subset\mathbf{R}\}.$$

The element $\alpha\in\Delta_{\mathbf{R}}$ is called *positive* if $\alpha(H)>0$ for all $H\in D$. For $\alpha\in\Delta_{\mathbf{R}}$ let

$$\mathbf{g}_0^\alpha=\{X\in\mathbf{g}_0\mid[H,X]=\alpha(H)X\text{ for }H\in\mathbf{h}_{\mathbf{m}}\}.$$

Then α is called a *root* of M with respect to $\mathbf{h}_{\mathbf{m}_0}$ if $\dim\mathbf{g}_0^\alpha\neq0$. (The algebra $\mathbf{n}_0^+=\sum_{\alpha>0}\mathbf{g}_0^\alpha$ is nilpotent and we have the Iwasawa decomposition $\mathbf{g}_0=\mathbf{k}_0+\mathbf{h}_{\mathbf{m}_0}+\mathbf{n}_0^+$ or $G=KA_mN^+$, as discussed below.)

Let $\mathbf{h}_0$ be the maximal Abelian subalgebra of $\mathbf{g}_0$ which contains $\mathbf{h}_{\mathbf{m}_0}$ and let $\mathbf{g}$ be as above. Thus $\mathbf{h}$, the subspace spanned by $\mathbf{h}_0$, is a Cartan subalgebra of $\mathbf{g}$. Let $\Sigma^{(+)}$ be the set of (positive) roots of $\mathbf{g}$ with respect to $\mathbf{h}$. There is a direct-sum decomposition

$$\mathbf{g}=\mathbf{h}\oplus\sum_{\alpha\in\Sigma}\mathbf{g}^\alpha.$$

According to Weyl, for every $\alpha\in\Sigma$, there are vectors $e_\alpha\in\mathbf{g}^\alpha$ and $H_\alpha\in\mathbf{h}$ such that $B(e_\alpha,e_{-\beta})=\delta_{\alpha,-\beta}$, $H_\alpha=[e_\alpha,e_{-\alpha}]$, $B(H_\alpha,X)=\alpha(X)$ for $X\in\mathbf{h}$, etc. The restriction of the Killing form B induces a nondegenerate bilinear form $(\,,\,)$ on $(-1)^{\frac{1}{2}}\mathbf{h}_0$ and also on $\mathbf{h}^*=\mathbf{h}_{\mathbf{m}_0}+(-1)^{\frac{1}{2}}(\mathbf{h}\cap\mathbf{k}_0)$. All roots are real on $\mathbf{h}^*$; for any root α on $\mathbf{h}^*$ there exists an $H_\alpha\in\mathbf{h}^*$ such that $\alpha(X)=(H_\alpha,X)$ for all $X\in\mathbf{h}^*$. Let $(\alpha,\beta)=(H_\alpha,H_\beta)$. Let H_β be the element of $\mathbf{h}$ such that

$$\alpha(H_\beta)=2(\alpha,\beta)(\beta,\beta)^{-1}.$$

The Weyl chamber of $\mathbf{g}$ is

$$D^{(0)}(\mathbf{g})=\{\alpha\in\mathbf{h}^*\mid(\alpha,\beta)=\alpha(H_\beta)\geq0\,(>0),\quad\text{for all}\quad\beta\in\Sigma^+\},$$

and it is divided by the hyperplanes

$$P_\alpha=\{\alpha\in\mathbf{h}^*\mid(\alpha,\beta)=0,\beta\in\Sigma^+\}.$$

$D(\mathbf{g})$ is the fundamental domain for the Weyl group $W(\mathbf{g})$ which is generated by reflections $r_\alpha(\beta)=\beta-n_{\beta\alpha}\alpha$ about the hyperplanes for $\alpha,\beta\in\Sigma$ and

$$n_{\alpha\beta}=2(\alpha,\beta)(\beta,\beta)^{-1}.$$

$W(\mathbf{g})$ is the quotient of the normalizer of $\mathbf{h}$ by the centralizer of $\mathbf{h}$ in $G^{\mathbf{C}}$, that is,

$$W(G^{\mathbf{C}})=N_{G^{\mathbf{C}}}(H)/Z_{G^{\mathbf{C}}}(H)=N_{G^{\mathbf{C}}}(H)/H.$$

An element $\alpha\in\mathbf{h}^*$ is called *integral* if

$$2(\alpha,\beta)(\beta,\beta)^{-1}\in\mathbf{Z}\quad\text{for}\quad\beta\in\Sigma.$$

α is integral on a discrete lattice $\mathbf{Z}(\mathbf{g})$ of maximal rank. Let $\mathbf{h}^{\#}=(\mathbf{Z}(\mathbf{g}))_{\mathbf{Q}}$. A linear form α on $\mathbf{h}$ is called a *weight* of a representation $\rho:G^{\mathbf{C}}\to V$ (where V is a vector space) if there is a nonzero vector $v\in V$ such that $\rho(H)v=\alpha(H)v$, for $H\in\mathbf{h}$. Thus α is a weight if $\alpha(H_\beta)$ is integral for all simple roots β_i, for $i=1,\cdots,l$. For each weight α, we set

$$V^\alpha=\{v\in v\mid\rho(H)v=\alpha(H)v,\text{ for all }H\in\mathbf{h}\};$$

this gives the decomposition of the representation space into weight spaces $V=\oplus V^\alpha$. (The weights of the adjoint representation of $\mathbf{g}$, other than 0, are just roots of $\mathbf{g}$ and we have the decomposition $\mathbf{g}=\oplus\mathbf{g}^\alpha$, as

given above with $\mathbf{g}^0 = \mathbf{h}$.) The weight α is called *dominant* if $\alpha > r(\alpha)$ for every $r \in W(\mathbf{g})$. The weights Λ_i, $i = 1, \cdots, l$, such that

$$\Lambda_i(H_{\beta_j}) = (\Lambda_i, \beta_j)(\beta_j, \beta_j)^{-1} = \delta_{ij}$$

for all simple roots $\beta_j, j = 1, \cdots, l$, are the *fundamental dominant weights* [they form a basis for $\mathbf{Z}(\mathbf{g})$]. $2\Sigma_1^l \Lambda_i = \delta =$ total sum of positive roots. The highest weight with respect to the lexicographic ordering on $\mathbf{h}^*$ of an irreducible representation of $\mathbf{g}$ is a dominant integral form, by the fundamental theorem, giving a one-to-one correspondence between irreducible representations and dominant integral forms. If $\mathbf{g}$ is simple, the adjoint representation is simple and the corresponding highest weight is the maximal root of $(\mathbf{g}, \mathbf{h})$.

The tensor field J, defining a complex structure, induces a decomposition of $\mathbf{m} = \mathbf{n}^+ + \mathbf{n}^-$ corresponding to the eigenvalues $\pm(-1)^{\frac{1}{2}}$. Furthermore, there exists an element Z in the center of $\mathbf{k}_0$ such that $[Z, Y] = \pm(-1)^{\frac{1}{2}} Y$ for any $Y \in \mathbf{n}^{\pm}$. ($J = \operatorname{ad} Z$ is a complex structure on $\mathbf{m}_0$.) There is a subset $\Pi_1 \subset \Pi =$ set of simple roots, such that $(Z, \alpha_j) = 0(1)$ for $\alpha \in (\notin)\, \Pi_1$. Then $\mathbf{n}^{\pm}$ is just the sum of positive (negative) eigenspaces of $\operatorname{ad} Z$ and is nilpotent. $\mathbf{k}$ is the 0 eigenspace of $\operatorname{ad} Z$. Let

$$\Sigma_K = (\Pi)_Z \cap \Sigma$$

and

$$\Psi = \Sigma - \Sigma_K.$$

Then

$$\mathbf{n}^{\pm} = \sum_{\alpha \in \Psi} \mathbf{C} e_{\pm\alpha}.$$

A root α is called *complementary* if $e_\alpha \in \mathbf{n}^+ + \mathbf{n}^-$. Roots may be ordered such that the roots of Ψ are all positive. $\mathbf{u} = \mathbf{k} + \mathbf{n}^+$ contains a Borel subalgebra of $\mathbf{g}$ and is the normalizer of $\mathbf{n}^+$ in $\mathbf{g}$ and $U^{\mathbf{C}} = K^{\mathbf{C}} \cdot N^+$ (where the semidirect product is meant).

C. Homogeneous (QDS)

Let $M = G^{\mathbf{C}}/U = G/K$ be a complex (respectively, real compact) form of a (Kähler) C-space. Let $2u =$ rank $G -$ rank K. Every C-space is a principal toral T^{2u}-bundle, where $T^{2u} \cong \hat{U}/U \to G/U \to G/\hat{U}$, over a C-space of positive Euler characteristic. T^{2u} is the identity component of the center of the group of all complex analytic homeomorphisms of M. M is Kähler iff M has a positive Euler characteristic; thus, M is a rational algebraic variety. In Ref. 7 we used the rational splitting of Wang, Borel, and Weil. In review, let $\tilde{\mathbf{k}}_0 = \mathbf{k}_0 \otimes_{\mathbf{R}} \mathbf{g}$, such that $\mathbf{g} = \tilde{\mathbf{g}}_0$. From the above discussion, the root system $\Sigma_{\mathbf{K}} \subset \Sigma$ of $\tilde{\mathbf{k}}_0$ is described as follows. For a Cartan subalgebra $\mathbf{h}$ of $\mathbf{g}$, we have $\mathbf{h} = \mathbf{c} + \mathbf{h}_K$ and $\mathbf{c} = \mathbf{p} \oplus \bar{\mathbf{p}}$, where $\mathbf{p} = 0$ iff rank $G =$ rank K. Also $\tilde{\mathbf{k}}_0 = \mathbf{h}_{\mathbf{K}} \oplus (\sum_{\alpha \in \Sigma_K} \mathbf{g}^\alpha)$ and $\mathbf{n}^{\pm}$ is given above; $\mathbf{u} = \bar{\mathbf{p}} \oplus \tilde{\mathbf{k}}_0 \oplus \mathbf{n}^-$ and $\bar{\mathbf{p}} \oplus \mathbf{h}_K$ lies on no rational hyperplane. If $\alpha \in \mathbf{h}^{\#}$, then $\alpha(H) = 0$ for all $H \in \bar{\mathbf{p}} \oplus \mathbf{h}_K$ implies $\alpha = 0$; Z is given above. There is a maximal root λ such that $Z = H_\lambda(e_\lambda)$ and $\mathbf{u} = \mathbf{h} \oplus \sum_{(\alpha,\lambda) \geq 0} \mathbf{g}^\alpha$ [the nonnegative eigenspaces of $\operatorname{ad}(H_\lambda)$]:

$$\hat{\mathbf{u}} = \{X \in U \mid B(X, H_\lambda) = 0\} = \mathbf{z}(e_\lambda)$$
$$= \{X \in \mathbf{u} \mid [X, H_\lambda] = 0\}; \quad \mathbf{k}_0 = \mathbf{g} \cap \mathbf{u},$$
$$\text{and} \quad [\mathbf{u}, H_\lambda] = \{H_\lambda\}.$$

Since $\mathbf{u}$ contains a Borel subgroup of $\mathbf{g}$, it is parabolic. Thus, M is a compact simply connected Hodge manifold.

For a homogeneous QDS, i.e., a homogeneous contact manifold, there is a unique fundamental field Z, which is given in terms of the contact form on $G^{\mathbf{C}}/U$. Let ω^* be lifted to $G^{\mathbf{C}}$; then there is a unique $Z \in \mathbf{g}$ with $\omega^*(X) = B(Z, X)$, for $X \in \mathbf{g}$, where B is a Killing form and where $B(X, Z) = 0$ for $X \in \mathbf{u}$. From the preceding discussion, we have that there is unique maximal root λ and $B(e_\lambda, X)$ defines an $\operatorname{ad}(\hat{U})$-invariant linear form on $G^{\mathbf{C}}/\hat{U}$, which we realize as a contact form. Thus, the Boothby structure theorem[22(b)] is the following.

Proposition: If $G^{\mathbf{C}}$ is a complex simple Lie group and U is a parabolic subgroup, such that

$$\mathbf{u} = \mathbf{h} \oplus \sum_{(\alpha,\lambda) \geq 0} \mathbf{g}^\alpha$$

and

$$\hat{\mathbf{u}} = \mathbf{z}(e_\lambda) = \{X \in \mathbf{g} \mid B(X, H_\lambda) = [X, e_\lambda] = 0\},$$

where λ is a maximal root, then

$$\hat{U}/U \to E = G^{\mathbf{C}}/U \to M = G^{\mathbf{C}}/\hat{U} \tag{3.3}$$

is a homogeneous QDS, where E is a Kähler C-space. Conversely, if $E(M, T)$ is a homogeneous QDS with E a Kähler C-space, then $E(M, T)$ is given by the sequence (3.3).

As a corollary we gave in Ref. 7 the classification of homogeneous QDS for compact complex simple Lie groups. The irreducible (Kähler) symmetric spaces S of type B (and their noncompact duals) of quaternionic type are put in one-to-one correspondence with homogeneous C-space QDS by the result of Wolf[34(a)]:

Proposition: Let G^u/K be a compact complex simply connected irreducible Riemann symmetric space of

[34] (a) J. A. Wolf, J. Math. Mech. **14**, 1033 (1965); (b) J. A. Wolf and A. Koranyi, Am. J. Math. **87**, 899 (1965).

type B, G^u compact form of $G^{\mathbf{C}}$; then

$$K/L = A_1/T^1 \to E = G^u/L = L_1 \times T^1 \to M = S = G^u/K = L_1 \times A_1,$$

where $A_1 = Sp(1)$ gives quaternionic structure, $\mathbf{l} = \mathbf{u} \cap \mathbf{g}^u = \mathbf{l}_1 \oplus \{(-1)^{\frac{1}{2}} H_\lambda\}$ is the real form of

$$\mathbf{h} \oplus \sum_{(\alpha,\lambda)=0} \mathbf{g}^\alpha, \mathbf{l}_1 = \hat{\mathbf{u}} \cap \mathbf{g}^u = \{H \in \mathbf{g}^u \wedge \mathbf{h} \mid \lambda(H) = 0\} \oplus \sum_{\substack{\alpha>0 \\ (\alpha,\lambda)=0}} \mathbf{g}^u \cap (\mathbf{g}^\alpha + \mathbf{g}^{-\alpha}),$$

and λ is the maximal root. Realizing $\mathbf{g}^u$ as $\mathbf{k}_0 + \mathbf{m}_0$, such that the noncompact dual is $\mathbf{g} = \mathbf{k}_0 + (-1)^{\frac{1}{2}}\mathbf{m}_0$ with Lie group G, we have then $E = G/L$ and $M = G/K$ gives the noncompact dual fibering.

D. Morse Index

In review, for symmetric spaces $M = G^{\mathbf{C}}/U = G/K$, the critical points are those $\alpha \in \mathbf{h}^\#$ for which $(\alpha, \beta) = 0$ for some $\beta \in \Sigma^+$. The Morse index Ind α of α is the number of roots $\beta \in \Sigma^+$ such that $(\alpha, \beta) < 0$. Define mappings $\mathfrak{I}_s(\Lambda) = s(\Lambda + \frac{1}{2}\delta) - \frac{1}{2}\delta$, for $\Lambda \in \mathbf{Z}(\mathbf{g})$, and $\mathfrak{I}: D(\tilde{\mathbf{k}}_0) \to D(\mathbf{g}) \cup \{0\}$, given by $\mathfrak{I}(\Lambda + \frac{1}{2}\delta) = 0$ if $\Lambda + \frac{1}{2}\delta$ is singular in $\mathbf{h}^\#$. If $\Lambda + \frac{1}{2}\delta$ is nonsingular, then there exists a unique $s \in W(\mathbf{g})$, such that

$$s(\Lambda + \tfrac{1}{2}\delta) \in D^0(\mathbf{g}),$$

and $\mathfrak{I}(\Lambda) = \mathfrak{I}_s(\Lambda) \in D(\mathbf{g})$. An element $s \in W(\mathbf{g})$ has the expression $s = s_{i_1} \cdots s_{i_r}$ of length r. If this length is a minimum, $n(s)$, then the length is the index of s. Let

$$\Phi_s = s(\Sigma^-) \cap \Sigma^+ = \{\alpha \in \Sigma^+ \mid s^{-1}\alpha \in \Sigma^-\}.$$

Thus, Ind $(s) = n(s) =$ Ind Φ_s, where the index here is the cardinality of the set Φ_s, i.e., the number of roots which "change sign" under s. If α is nonsingular, then there exists a unique s_α such that $s_\alpha(\alpha) \in D^0(\mathbf{g})$ and Ind $\alpha =$ Ind s_α.

Remarks. The minimum path from $D^0(\mathbf{g})$ to $s^{-1}D^0(\mathbf{g})$ meets $n(s)$ hyperplanes, which is the number of linear forms in $s^{-1}D^0(\mathbf{g})$ for $\alpha \in \Sigma^+$ which take negative values [that is, $\alpha(X) = \alpha(s^{-1}X)$, for $X \in \mathbf{h}^\#$]. s^{-1} is given as $s^{-1} = s_{\alpha_{j_1}} \cdots s_{\alpha_{j_q}}$, where $s_{\alpha_{i_j}}$ is a reflection across the root plane of the simple root α_{i_j} and

$$W^1 = \{s \in W(\mathbf{g}) \mid s^{-1}\Sigma_K^+ \subset \Sigma^+\} = \{s \in W \mid \Phi_s \subset \Psi = \Sigma^+ - \Sigma_K^+\} = W(\mathbf{g})/W(\tilde{\mathbf{k}}_0).$$

Let Λ be the highest weight of irreducible representation ρ of $G^{\mathbf{C}}$; then the weight λ of ρ_Λ is *extreme* iff $\lambda + \alpha$ or $\lambda - \alpha$ is not a weight of ρ_Λ. W^1 is set of all extreme weights of ρ_Λ. Let $\mathbf{n}_s^\pm = \{e_{\pm\alpha} \mid \alpha \in \Phi_s\}_{\mathbf{C}}$, for $s \in W$, with group $N_s^\pm$. The generalized Bruhat decomposition is

$$G^{\mathbf{C}} = \bigcup_{s \in W^1} N_{s^{-1}s}^+ s^{-1} K^{\mathbf{C}} N^+$$

and the cellular decomposition is

$$G^{\mathbf{C}} = \bigcup_{s \in W^1} V_s,$$

where $V_s = N_{s^{-1}s}^+ s^{-1} \cdot 0$. Thus, V_s is homeomorphic to $N_{s^{-1}s}^+$ and has complex dimension $n(s)$.

Every closed invariant form η on a compact form of the Kähler C-space $G/\hat{U}$ may be written in the form

$$\eta_\rho = \sum_{\alpha \in \Psi} \rho(H_\alpha) \omega^\alpha \wedge \bar{\omega}^\alpha$$

at the origin, where $H_\alpha \in \mathbf{h}^*$ is orthogonal to

$$\mathbf{h}_\gamma^\# = \{H_\alpha \mid \alpha \in \Sigma - \Sigma_K\}_{\mathbf{C}}$$

and where the $\omega^{\pm\alpha}(\omega^{-\alpha} = \bar{\omega}^\alpha)$ are duals of the root vectors $e_\alpha \in \mathbf{g}^\alpha$. The metric is

$$ds^2 = \sum_{\alpha \in \Psi} \rho(\alpha) \omega^\alpha \cdot \bar{\omega}^\alpha$$

and, as usual, $(2\pi(-1)^{\frac{1}{2}})^{-1}\eta_\rho = c_1(E_\rho)$ represents the first Chern class. Thus we may refer to the Morse index in terms of number of negative eigenvalues of the curvature form, Ricci tensor, or first Chern class.

Let $\mathbf{g} = \mathbf{k} + \mathbf{m}$ for a C-space G/K and let $\mathbf{h}_\mathbf{m}$ be the maximal Abelian subalgebra of $\mathbf{m}$; now $X \in \mathbf{m}$ is conjugate to the origin 0 if $d \exp X = 0$ or is such that

$$\det\left(\sum_0^\infty \frac{[(\operatorname{ad} X)^2]^n}{(2n+1)!}\right) = \prod_{\alpha \in \Sigma^{>0}} \frac{\sin \alpha((-1)^{\frac{1}{2}}H)}{\alpha((-1)^{\frac{1}{2}}H)} = 0,$$

for $H \in \mathbf{h}^*$ [that is, $R_0(X, \cdot) = -(\operatorname{ad} X)^2$]. Thus $X =$ ad $(k)H$ is conjugate to 0 iff $\alpha(H) \in \pi(-1)^{\frac{1}{2}} \times (\mathbf{Z} - 0)$. The $T_\mathbf{m} = \exp \mathbf{h}_\mathbf{m}$ are the geodesically imbedded tori. For $x \in \mathbf{h}_\mathbf{m}$, the Morse index of the geodesic segment $\tilde{x}(t) = \exp(tx)$, $0 \le t \le 1$, is Ind $(\tilde{x}) = \sum n_\alpha \|\alpha(x)\|$, where $\alpha \in \Sigma^{>0}$ restricts the index n_α times to some real linear form $\pm\alpha_i$ on $\mathbf{h}_\mathbf{m}$. Thus n_α is the multiplicity of the critical point of the geodesic at the hyperplane. The greatest integer less than $|\ |$ is denoted by $\|\ \|$. From Bott,[16(b)] the space of geodesics $\mathcal{S}_M$ is a collection of nondegenerate critical manifolds $\mathcal{O}_M$. Two geodesics lie in the same $V \in \mathcal{O}_M$ iff they are conjugate under the Weyl group. Let Λ be the maximal root. If $x_\gamma \in \mathbf{h}_\mathbf{m}$ with $\tilde{x}_\gamma \in \mathcal{S}M$, then $x \in x_\gamma + \Lambda$ has a unique critical manifold $V_x \subset \mathcal{S}M$. V_x is homeomorphic to $K^0/Z_x(K)$, where K^0 is the identity component of K and $Z_x(K)$ is the centralizer of x in K. Bott (cf. Milnor[31]) evaluates the Morse index for $U(n)$, $SU(n)$, and lower bounds $|\nu|$ of positive integers $\in$ Ind (γ) for various cases (to prove

Bott periodicity). From the discussion in Sec. 3A, the conjugate points of the geodesic $t \to \exp(tA)$, where $A \in \mathbf{su}(n) = T(SU(n))_I$, are determined by the positive eigenvalues of $K_A(W) = R(A, W)V = \frac{1}{4}[[A, W]A]$. If $W = (w_{ij})$ and

$$A = \begin{pmatrix} i\pi k_1 & \cdots & \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ & \cdots & i\pi k_n \end{pmatrix},$$

then $K_A(W) = \frac{1}{4}(\pi^2(k_i - k_j)^2 w_{ij})$ has eigenvalues for $e = \frac{1}{4}\pi^2(k_i - k_j)$ and conjugate points $t = n\pi/(e)^{\frac{1}{2}}$, $n \in \mathbf{Z}_+$. The index of the geodesic is $\mathrm{Ind}\,(\gamma[-I, I]) = \sum_{k_i > k_j}(k_i - k_j - 2)$. For $U(n)$, let n_i be the number of times the integer k_i occurs. Then, $\mathrm{Ind}\,(\gamma) = \sum_{i>j} n_i n_j (k_j - k_i - 1)$. The critical manifold in this case is $V_{x\gamma} = U(2n)/U(n) \times U(n)$, $|\nu| = 2(n+1)$. In other cases, $V_{x\gamma} = SO(2n)/U(n)$, $|\nu| = 2(n-1)$; $V_{x\gamma} = Sp(n)/U(n)$, $|\nu| = 2(n+1)$; $V_{x\gamma} = U(2n)/Sp(n)$, $|\nu| = 4n - 2$; etc.

E. Homogeneous Bundles

Associated to any C-space $M = G^{\mathbf{C}}/U$ there is a natural principal fiber bundle $U \to G^{\mathbf{C}} \to G^{\mathbf{C}}/U$ (with G-invariant connection ω). U is the semidirect product $U = K^{\mathbf{C}} \cdot N^-$; thus, every representation $\rho \in \mathrm{Hom}\,(K^{\mathbf{C}}, GL(m, \mathbf{C}))$ can be extended naturally to a holomorphic representation of U over $\mathbf{C}^{\mathrm{m}}$ (T^{m} or the finite $\mathbf{C}$-module $E_\rho u$), say, $\rho^u \in \mathrm{Hom}\,(U, GL(m, \mathbf{C}))$. Define a homogeneous principal-bundle space $\mathbf{E}_\rho u = G \times_u \mathbf{C}^{\mathrm{m}}(T^{\mathrm{m}}$ or $E)$ by the equivalence relation $u \circ (g, \xi) \sim (gu^{-1}, \rho(u)\xi)$ for $g \in G$, $\xi \in \mathbf{C}^{\mathrm{m}}$ (etc.), and $u \in U$. Thus, $\mathbf{E}_\rho u$ is a homogeneous holomorphic vector bundle over $G^{\mathbf{C}}/U$ with fiber $\mathbf{C}_{\mathrm{m}}$ (etc.). Let $\hat{U}$ be the kernel of ρ. Thus $\hat{U}$ is a closed subgroup of U. The principal fiber bundle

$$\mathbf{C}^{\mathrm{m}}\ (\text{etc.}) \to \mathbf{E}_\rho u \to G^{\mathbf{C}}/U$$

is equivalent to the principal fiber bundle

$$\hat{U}/U \to G/\hat{U} \to G/U.$$

In Sec. 4 we examine the sheaf of germs of holomorphic cross sections of $\mathbf{E}_\rho u \to M$, and therefore $G/\hat{U} \to M$, which is $\Omega(\mathbf{E}_\rho u)$. We have omitted a discussion of the case of $\mathbf{E}_\rho \circ J$ defined by automorphic forms of type ρ, although it is very pertinent, especially for a discussion of Abelian varieties, cusp forms, etc.[35–39]

[35] M. Ise, Am. J. Math. **86**, 70 (1964).
[36] S. Ramanan, Topology **5**, 159 (1966).
[37] (a) A. Borel, Proc. Natl. Acad. Sci. (U.S.) **40**, 1140 (1954); (b) Ann. Math. **71**, 509 (1960).
[38] Y. Matsushima and S. Murakami, Ann. Math. **78**, 365 (1963); Proc. Sem. Diff. Geometry Kyoto, 82 (1965); Osaka Math. J. **14**, 1 (1962); Osaka J. Math. **2**, 1 (1965).
[39] M. S. Raghunathan, Am. J. Math. **87**, 103 (1965); J. Math. Mech. **13**, 97 (1964).

4. HODGE INDEX AND VANISHING THEOREMS[8,9,15–20,23,33–39]

A. Hodge Index of Inertia

Given a real-valued symmetric bilinear form $A(x, y)$, the associated quadratic form is $q(x) = A(x, x)$, where A is determined by $q: A(x, y) = \frac{1}{2}[q(x + y) - q(x) - q(y)]$. The *index of inertia* (Hodge index) of a ("normalized") real-valued symmetric bilinear form A on a real finite-dimensional manifold is defined as $\mathrm{Ind}\,A = s - t$, where s is the number of positive eigenvalues and t is the number of negative eigenvalues; A is then said to have *signature*, or type, (s, t).

The natural bilinear form for compact manifolds is that induced from the bilinear form of the Serre duality. It is immediately evident that (1.10) induces the cup product

$$H^q(M, \Omega^p(E)) \times H_\Phi^{n-q}(M, \Omega^{n-p}(E^*)) \to H_\Phi^{2n}(M, \mathbf{C}\,(\text{or } \mathbf{R})), \quad (4.1)$$

which is the bilinear map $u \cup [M]$, where $u \in H_\Phi^{2n}(M, \mathbf{C}\ (\text{or } \mathbf{R}))$.

The standard object which we are treating is the Hermitian holomorphic bundle $E \to M$. Any real Hermitian (1, 1) form

$$A = \tfrac{1}{2}(-1)^{\frac{1}{2}} \sum a_{\alpha\beta}\omega^\alpha \wedge \bar{\omega}^\beta,$$

where $a_{\alpha\beta} + \bar{a}_{\beta\alpha} = 0$, defines an endomorphism of $T(M)$ given by $(Ae_\alpha, e_\beta) = a_{\alpha\beta}$, where the e_κ are dual to ω_κ, for $\kappa = 1, \cdots, n$. Since $A = {}^t\bar{A}$, the frames may be chosen such that

$$A = \tfrac{1}{2}(-1)^{\frac{1}{2}} \sum \lambda_\alpha \omega^\alpha \wedge \bar{\omega}^\alpha,$$

where the λ_κ are real. Then, if A has signature (s, t), we may assume that A is of the form

$$A = \tfrac{1}{2}(-1)^{\frac{1}{2}}\left\{ \sum_{1 \le \alpha \le s} \lambda_\alpha \omega^\alpha \wedge \bar{\omega}^\alpha - \sum_{n-t \le \beta \le n} \lambda_\beta \omega^\beta \wedge \bar{\omega}^\beta \right\}, \lambda_\kappa > 0.$$

For the case $E \to M$, $A \in A^{1,1}(\mathrm{End}\,E)$ is said to have signature (s, t) if the corresponding real (1, 1)-form $A\{e\} = \frac{1}{2}(-1)^{\frac{1}{2}}\langle Ae, e\rangle$ has signature (s, t) for all $e \neq 0$ in E. The natural real Hermitian (1, 1) form is the curvature Ξ of the metric connection, because we see from the discussion in Secs. 1C and 2A that the Chern classes can be calculated in terms of the curvature of the bundle. For QDS, E is $T^*(M)$ with the Kähler (Hodge) metric identified with the Kähler metric on M. The operations L and Λ are defined, and there are two Hermitian forms $A(\varphi, \varphi) = \langle A\varphi, \varphi\rangle$ and $A^{\#}(\varphi, \varphi) = \langle \Lambda A \wedge \varphi, \varphi\rangle$ which satisfy

$$A^{\#}(\varphi, \varphi) + (\mathrm{Tr}\,A)(\varphi, \varphi) - A(\varphi, \varphi) = A(*\bar{\varphi}, *\bar{\varphi}) = A(\#\varphi, \#\varphi), \quad (4.2)$$

for $\varphi \in A^{0,q}(E)$ and the metric (1.10). From Ref. 19(d) we quote

Lemma: If A has signature (s, t), then there exist Hermitian metrics on M such that $A(\varphi, \varphi)$ is negative definite on $A^{0,q}$ for $q > n - t$ and $A^{\#}(\varphi, \varphi)$ is positive definite on $A^{0,q}$ for $q < s$. For $A = \Xi$, we have

$$\Xi^{\#}(\varphi, \varphi) = (-1)^{\frac{1}{2}}\langle \Lambda e(\Xi)\varphi, \varphi\rangle.$$

B. Signature and Eigenvalues

Define the signature of E to be the signature (s, t) of Ξ. If Ξ has signature $(n, 0)$, E is said to be *weakly positive*. E is called *weakly negative* if E^* is weakly positive. E is called *negative* if, for $A = \Xi$ in (4.2), $\Xi^{\#}(\varphi, \varphi)$ is negative definite for $0 \leq q \leq n - 1$. This agrees with Kodaira's definition that a $\mathbf{C}^*$-bundle E is positive if $c_1(E) \in H^{1,1}(M, \mathbf{Z})$ is positive, i.e., iff its induced element from sequence (2.5), namely $\lambda^* c_1(E) \in H^{1,1}(M, \mathbf{R})$, is a fundamental class of the Kähler metric. M is called a *Hodge* manifold if $H^{1,1}(M, \mathbf{Z})$ contains at least one positive element, i.e., the fundamental 2-form is integral $\Xi \in H^2(M, \mathbf{Z})$. We noted previously[6,7] and in Secs. 2 and 3 of this paper that Hodge manifolds are "canonical" QDS; the homogeneous QDS which we constructed are Hodge and therefore algebraic.

For homogeneous C-spaces the Cartan–Killing form gives a Hermitian geometry on a Kähler C-space M; this is a Kähler geometry iff M is a Hermitian symmetric space. According to Koszul, Kähler C-spaces are Einstein spaces, i.e., the Ricci tensor is given by $R_{\alpha\beta} = (R/2n)h_{\alpha\beta}$, where $R = 2\sum R^{\alpha}_{\alpha}$ is the scalar curvature of the metric h of M. In this case, $E = T^*(M)$ has fiber $\mathbf{C}^N$, $N = \frac{1}{2}n(n + 1)$. Then the Nakano identity

$$((-1)^{\frac{1}{2}}(e(\Xi)\Lambda - \Lambda e(\Xi))\varphi, \varphi) = (-1)^{\frac{1}{2}}(\Lambda e(\Xi)\varphi, \varphi), \quad \varphi \in A^{0,q}(E)$$

reduces to

$$\Xi^{\#}(\varphi, \varphi) = (-1)^{\frac{1}{2}}(\Lambda e(\Xi)\varphi, \varphi) = \frac{-R}{2n}(\varphi, \varphi) + \frac{1}{(q-1)!}\sum R_{\cdots\varphi\ldots\varphi\cdots}.$$

A new Hermitian form may be defined as

$$\Xi^{\#}(\varphi, \varphi) + (R/2n)(\varphi, \varphi).$$

Then we define E to be *sufficiently negative* if this form is negative definite on $E^{0,q}$, $0 \leq q \leq n - 1$. Thus one is led to investigate the eigenvalues of the equation

$$P(x_0)\colon \varphi_{\alpha\beta} \to \sum_{\alpha,\beta} R_{\tau}{}^{\alpha\beta}{}_{\gamma}(x_0)\varphi_{\alpha\beta}, \quad \varphi_{\alpha\beta} = \varphi_{\beta\alpha}. \tag{4.2'}$$

$P(x_0)$ has N eigenvalues; let $\delta(x_0)$ be the smallest. If $\delta(x_0) \leq 0$, then

$$\Xi^{\#}(\varphi, \varphi)_{x_0} + \frac{R}{2n}(\varphi, \varphi)_{x_0} \geq \delta(x_0)\frac{(q+1)}{2}$$

and

$$R = 2\,\mathrm{Tr}\,P(x_0) \geq n(n+1)\delta(x_0).$$

Therefore, if M is a Kähler–Einstein space with $R < 0$, then

$$\delta(x_0) \leq R/n(n+1) < 0.$$

From Ref. 17a, $H^q(M, \Omega(E)) = 0$, for

$$0 \leq q < \inf_{x_0}[(R/n)\delta(x_0) - 1] = \gamma(D) - 1 = M\text{-index}$$

(see below).

C. Theorems in Terms of QDS

Call the QDS negative if $E(M, T)$ is negative. The standard theorems can now be restated in terms of QDS. From Refs. 15, 8(b), 8(c), 19, and 20(b), we have the following proposition.

Proposition: If a compact Kähler (C-space G/K with $\mathbf{p} = 0$) QDS is a Hodge QDS, then the fact that the Ricci tensor is positive definite implies that the first Chern class is positive definite; thus the Kähler QDS is algebraic and simply connected. The canonical bundle $\mathbf{K} = \Lambda^n T^*(M)$ of symmetric spaces of type A are positive definite, which implies $\mathbf{K}^{-1}$ is positive definite for symmetric spaces of type B. Thus G/K, the canonical bundle of C-space QDS, is a negative line bundle.

Proposition: If $E(M, T)$ is a negative QDS (where M is a Kähler C-space), then

$$H^q(M, \Omega(E(M, T)) = 0 = H^{n-q}(M, \Omega(\mathbf{K} \times E^*(M, T)), \quad 0 \leq q \leq n - 1,$$

where $\mathbf{K}$ is a canonical bundle.

The Hodge index of inertia is given by the quadratic form (1.10) as

$$I(\varphi, \eta) = \int_M \varphi \wedge \eta, \quad \text{for} \quad \varphi, \eta \in H^n(M, \mathbf{R}) \tag{4.3}$$

[cf. Eq. (4.1)]. When we identify indices, we have, by definition,[8(a)] the following proposition:

Proposition: The index of inertia of a compact Kähler QDS is given by the index of (4.3) and is

$$\tau(E(M, T)) = s - t = \sum_{p,q}(-1)^q h^{p,q}(E(M, T)) = \sum_{0}^{n}\chi^p(E(M, T)) = \chi_1(E(M, T)),$$

where $h^{p,q} = \dim H^{p,q}$ and $\chi^p(E \otimes \Lambda^p T^*) = \chi^p(E)$ is the Euler–Poincaré characteristic. τ has been evaluated for spaces type B by Hirzebruch[8(a,b)]; see Ref. 19(b).

This is generalized by Hirzebruch and Thom to any compact oriented manifold M of dimension $4n$ and holomorphic vector bundle E. Here, the signature is identically zero, if $\dim M$ is not divisible by four. In the "Atiyah–Singer" formulation:

$$\begin{aligned}\text{signature } I(\varphi, \eta) &= \text{signature } (\varphi, \eta)\\ &= \text{analytic index } D = \tau(D)\\ &= \dim \operatorname{Ker} D - \dim \operatorname{Ker} D^*\\ &= \sum (-1)^p \dim H^p(M, E)\\ &= \chi(M, E)\end{aligned}$$

where $D = \bar{\partial} + \mathfrak{D}$ (or $D = d + \delta$) and the topological index is equal to the analytic index.

From Refs. 31, 16(a), and 19, we have the proposition:

Proposition: If $E(M, T)$, a compact Kähler QDS, has curvature form Ξ with α positive and β negative eigenvalues, then there exists μ_0 such that, for $\mu \geq \mu_0$,

$$H^q(M, \Omega E^\mu(M, T)) = 0,$$
$$0 \leq q \leq \beta, \quad n - \alpha \leq q \leq n,$$

where E^μ is the μth symmetric power and β is the M-index of the Chern class $c_1(E)$, i.e., h or Ξ. A negative line bundle has only negative eigenvalues. In this case, if $\beta \geq 1$, then $H^0(M, \Omega(E)) = 0$.

From calculation on eigenvalues of (4.2), we quote the results of Calabi and Vesentini [17(a)] and Borel.[37(b)]

Proposition: For every compact symmetric complex manifold M whose universal covering space is the Cartan domain D (type A), then $H^q(M, \Omega(E = T(M))) = 0$ for all $0 \leq q \leq \gamma(D) - 1$, where

	$\mathrm{AI}_{m,m'}$	AII_m	AIII_m	AIV_m		
$\gamma(D) - 1$:	$m + m' - 1$	$2m - 3$	m	$m - 1$		
				AV	AVI	
				11	17	

$\mathbf{g}$ simple has unique simple root $\gamma_1 \in \Psi$, $\gamma(D) = (\gamma_1, \gamma_1)^{-1}$.

Let Λ be an element of the interior of the space of maximal weights. Then from Borel[37(a)] we have the following proposition:

Proposition: The QDS $E_\Lambda(M, T)$ corresponding to Λ is positive in the sense of Kodaira.

Let M be the (irreducible) Kähler C-space G^C/U. Then from (1.7), the Atiyah exact sequence for the case $U \to G^C \to M$, one finds that

$$H^q(M, \Omega(E = T(M))) = H^{q+1}(M, \Omega(L)) = 0$$

and

$$H^q(M, \Omega(L)) = 0,$$

for all q. Thus, we have the following proposition.[16(a),19(b)]

Proposition: We have

$$H^q(M, \Omega(E = T(M))) = 0,$$

for $q > 0$. For M an irreducible compact symmetric Kähler manifold,

$$\dim_C H^0(M, \Omega(E)) = \dim_C G^C(M) = \dim_R G(M),$$

where $G^C(M)$ is the complex Lie group of all holomorphic transformations of M onto itself and where $G(M)$ is the compact Lie group of all isometric transformations.

Note: The space of Kodaira–Spencer deformations is $H^1(M, \Omega(E))$ or $H^1(M, \Omega(L))$; see below.

Furthermore, from Bott[16(a)] [cf. Refs. 17(a), 19, 29, 30], if $\rho \in \operatorname{Hom}(K^C, GL(M, \mathbf{C}))$ is an irreducible representation [and thus is represented by the maximal (dominant) weight Λ], $\mathbf{E}_\rho u$ is the homogeneous vector bundle, and $\frac{1}{2}\delta$ is the half-sum of the positive roots of G^C, then we have the following result.

Proposition: For every QDS represented by $M = G^C/U$, if $\Lambda + \frac{1}{2}\delta$ is singular [i.e., there exists a root α such that $(\Lambda + \frac{1}{2}\delta, \alpha) = 0$], then

(i) $H^i(M, \Omega(\mathbf{E}_\rho u)) = 0$ for all $q \geq 0$;

(ii) if $\Lambda + \frac{1}{2}\delta$ is regular (see Sec. 3C), there exists a unique $s \in W(\mathbf{g})$ such that $s(\Lambda + \frac{1}{2}\delta) \in D^0(\mathbf{g})$ and $H^q(M, \Omega(\mathbf{E}_\rho u)) = 0$ for $q \neq \operatorname{Ind} s$; the action of G^C on $H^q(M, \Omega(\mathbf{E}_\rho u))$ (i.e., the "induced representation") is irreducible and has the highest weight $s(\Lambda + \frac{1}{2}\delta) - \frac{1}{2}\delta$. If the highest weight of ρ is not the highest weight of any irreducible representation of G^C, then $H^0(M, \Omega(\mathbf{E}_\rho u)) = 0$.

These results have been presented, since they lead directly to the study of Kodaira–Spencer deformation theory,[16–21,23,35–39] which has recently been studied by

Hermann[40] with quantum physics applications. Details in this direction, a study of calculus of variations for quantum field theory, and relations to Morse theoretic techniques will be studied in later publications.

[40] R. Hermann, Commun. Math. Phys. 2, 251 (1966); 3, 53, 75, (1966).

ACKNOWLEDGMENTS

The author was a NATO postdoctoral fellow 1966/67 at Université Libre de Bruxelles during which period this research was completed. Presently the author is a Visiting Member at Courant Institute of Mathematical Sciences where this paper materialized.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

On the Inverse Problem for the Klein–Gordon s-Wave Equation*

ANTONIO DEGASPERIS
Physics Department, Indiana University, Bloomington, Indiana 47401

(Received 3 March 1969)

The s-wave Klein–Gordon equation of a particle in an electriclike, spherically symmetrical, local, and energy-independent potential is studied with the aim of deriving exact relationships between the potential and the experimental data. We obtain the value of the potential and of all its derivatives at the origin as explicit functionals of the phase shift of the particle and the antiparticle and of their bound-state parameters. Furthermore, we find a set of infinitely many nonlinear sum rules for the phase shift and the bound-state parameters of both the particle and the antiparticle; these sum rules embody the requirement that the potential be electriclike. This restriction on the potential is also briefly discussed in the framework of the Gel'fand–Levitan approach to the inverse problem.

1. INTRODUCTION

This paper is one of a series on the inverse problem in quantum potential scattering theory. The aim of this investigation is to derive exact relationships between the parameters which characterize the scattering process and the interaction which causes it. In the first paper expressions for the value of the potential and of its derivatives at the origin were derived in terms of the s-wave phase shift and bound-state parameters corresponding to that potential via the Schrödinger equation.[1] In the second paper, variational bounds for the potential were established in terms of the s-wave phase shift and bound-state parameters.[2,3] In I it was shown that different techniques can be used to derive explicit relationships between the potentials and the scattering parameters, and we emphasized the convenience and the elegance of the method based on the asymptotic expansions of the scattering parameters and their analyticity properties. In this paper we apply this method to the simplest relativistic case, the s-wave radial Klein–Gordon[4] equation[5]

$$\left(\frac{d^2}{d\rho^2} - M^2 + [E - V(\rho)]^2\right)\varphi(E;\rho) = 0. \qquad (1.1)$$

Here M is the mass of the scattered or bound particle and the interaction is completely described by a single function of the modulus of the radius $\boldsymbol{\rho}$, the spherically symmetric, local, and energy-independent potential $V(\rho)$ (regarded as time-component of a 4-vector), which is assumed to be holomorphic (and vanishes asymptotically). This assumption says that the value of the potential and of all its derivatives at the origin determines the potential everywhere.

As a consequence of the relativistic kinematics, Eq. (1.1) has positive-energy as well as negative-energy solutions; moreover, the potential is supposed to be such that Eq. (1.1) does not possess complex energy solutions.[6] The stationary KG equation is not a usual eigenvalue equation, but it could be considered a Schrödinger equation with an energy-dependent interaction. As is well known, we are,

* This work has been partially supported by the U.S. Army Research Office, Durham, North Carolina.

[1] F. Calogero and A. Degasperis, J. Math. Phys. 9, 90 (1968). Hereafter referred to as I.

[2] F. Calogero, O. D. Corbella, A. Degasperis, and B. M. De Stefano, J. Math. Phys. 9, 1002 (1968).

[3] Simultaneously with this work, an analogous investigation has been devoted by O. D. Corbella to the potential scattering of Dirac particles (J. Math. Phys.) (to be published).

[4] KG will be used as an abbreviation for Klein–Gordon.

[5] Throughout this paper we use units such that $h = c = 1$.

[6] The appearance of these unphysical solutions has been investigated for a square-well potential by L. I. Schiff, H. Snyder, and J. Weinberg, Phys. Rev. 57, 315 (1940).

Hermann[40] with quantum physics applications. Details in this direction, a study of calculus of variations for quantum field theory, and relations to Morse theoretic techniques will be studied in later publications.

[40] R. Hermann, Commun. Math. Phys. 2, 251 (1966); 3, 53, 75, (1966).

ACKNOWLEDGMENTS

The author was a NATO postdoctoral fellow 1966/67 at Université Libre de Bruxelles during which period this research was completed. Presently the author is a Visiting Member at Courant Institute of Mathematical Sciences where this paper materialized.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

On the Inverse Problem for the Klein–Gordon s-Wave Equation*

ANTONIO DEGASPERIS
Physics Department, Indiana University, Bloomington, Indiana 47401

(Received 3 March 1969)

The s-wave Klein–Gordon equation of a particle in an electriclike, spherically symmetrical, local, and energy-independent potential is studied with the aim of deriving exact relationships between the potential and the experimental data. We obtain the value of the potential and of all its derivatives at the origin as explicit functionals of the phase shift of the particle and the antiparticle and of their bound-state parameters. Furthermore, we find a set of infinitely many nonlinear sum rules for the phase shift and the bound-state parameters of both the particle and the antiparticle; these sum rules embody the requirement that the potential be electriclike. This restriction on the potential is also briefly discussed in the framework of the Gel'fand–Levitan approach to the inverse problem.

1. INTRODUCTION

This paper is one of a series on the inverse problem in quantum potential scattering theory. The aim of this investigation is to derive exact relationships between the parameters which characterize the scattering process and the interaction which causes it. In the first paper expressions for the value of the potential and of its derivatives at the origin were derived in terms of the s-wave phase shift and bound-state parameters corresponding to that potential via the Schrödinger equation.[1] In the second paper, variational bounds for the potential were established in terms of the s-wave phase shift and bound-state parameters.[2,3] In I it was shown that different techniques can be used to derive explicit relationships between the potentials and the scattering parameters, and we emphasized the convenience and the elegance of the method based on the asymptotic expansions of the scattering parameters and their analyticity properties. In this paper we apply this method to the simplest relativistic case, the s-wave radial Klein–Gordon[4] equation[5]

$$\left(\frac{d^2}{d\rho^2} - M^2 + [E - V(\rho)]^2\right)\varphi(E;\rho) = 0. \qquad (1.1)$$

Here M is the mass of the scattered or bound particle and the interaction is completely described by a single function of the modulus of the radius $\boldsymbol{\rho}$, the spherically symmetric, local, and energy-independent potential $V(\rho)$ (regarded as time-component of a 4-vector), which is assumed to be holomorphic (and vanishes asymptotically). This assumption says that the value of the potential and of all its derivatives at the origin determines the potential everywhere.

As a consequence of the relativistic kinematics, Eq. (1.1) has positive-energy as well as negative-energy solutions; moreover, the potential is supposed to be such that Eq. (1.1) does not possess complex energy solutions.[6] The stationary KG equation is not a usual eigenvalue equation, but it could be considered a Schrödinger equation with an energy-dependent interaction. As is well known, we are,

* This work has been partially supported by the U.S. Army Research Office, Durham, North Carolina.

[1] F. Calogero and A. Degasperis, J. Math. Phys. 9, 90 (1968). Hereafter referred to as I.

[2] F. Calogero, O. D. Corbella, A. Degasperis, and B. M. De Stefano, J. Math. Phys. 9, 1002 (1968).

[3] Simultaneously with this work, an analogous investigation has been devoted by O. D. Corbella to the potential scattering of Dirac particles (J. Math. Phys.) (to be published).

[4] KG will be used as an abbreviation for Klein–Gordon.

[5] Throughout this paper we use units such that $h = c = 1$.

[6] The appearance of these unphysical solutions has been investigated for a square-well potential by L. I. Schiff, H. Snyder, and J. Weinberg, Phys. Rev. 57, 315 (1940).

therefore, forced to introduce an indefinite metric in the space of the two-component representation of the KG solutions and to write down the usual orthogonality and completeness relations in this twofold set of solutions. The physical interpretation lies in the fact that the KG equation describes two different processes, particle scattering as well as antiparticle scattering, whereas in the nonrelativistic picture two different equations are required for these two processes. In fact, it is well known that the infinite-light-velocity limit of the KG equation turns out to be a pair of Schrödinger equations with potentials $V(\rho)$ and $-V(\rho)$, respectively, for the particle and for the antiparticle.

We now consider the inverse problem. Let the experimental data be characterized by the scattering phaseshift $\eta(k)$ for all the energies (k is the modulus of the linear momentum), the binding energies E_n, and the norm of the bound-state functions C_n^{-1} (see I). A 1-to-1 correspondence between these quantities and the potential $V(\rho)$ can be proved in the nonrelativistic case using the Gel'fand–Levitan integral equation.[7] However, the KG equation determines the dynamics of both the particle and the antiparticle, so that the question arises: Are experiments with particles alone sufficient, or do we have to perform experiments with both particles and antiparticles to get all the information on the potential? In other words, we have to find out what the minimum experimental information is for completely reconstructing the potential. This aspect of the inverse problem does not appear in the previous literature on the KG equation. The papers by Corinaldesi[8] and Verde[9] are devoted to a relativistic generalization of the Gel'fand–Levitan[10] approach to the inverse problem, while a paper by De Alfaro[11] gives the mathematical tools for conveniently inserting the experimental data in the Gel'fand–Levitan formalism. In this formalism the input data are condensed in the so-called spectral function. As we show in Appendix B, the latter can be expressed in terms of the phase shifts $\eta_+(k)$ and $\eta_-(k)$, respectively, of the particle and of the antiparticle, of the energies $E_n > 0$ ($E_m < 0$) of the bound states of the particle (antiparticle), and of the normalization constants C_n of the bound-state functions. In Appendix A we sketch briefly the mathematical situation connected with the Gel'fand–Levitan integral equation, and we show that, for the construction of the potential, the quantities $\{\eta_+(k), \eta_-(k), E_n, C_n\}$ are redundant. That is to say, if arbitrarily given, they generally do not lead to a potential in the KG equation (1.1). To avoid this redundancy, we start from the more general equation

$$\left(\frac{d^2}{d\rho^2} + k^2 + V_1(\rho) + EV_2(\rho)\right) y(k, E; \rho) = 0, \quad (1.2)$$

which reduces to the KG equation (1.1) with the particular choice

$$V_2(\rho) = -2V(\rho), \quad V_1(\rho) = V^2(\rho), \quad k^2 = E^2 - M^2. \quad (1.3)$$

In Sec. 2 we apply the method of asymptotic expansion to Eq. (1.2), and in Sec. 3 we give the exact relations between the potential and its derivatives at the origin in terms of the parameters $\{\eta_+(k), \eta_-(k), E_n, C_n\}$. Furthermore, we obtain in a quite natural way the relations that these parameters must satisfy in order to be caused by a single potential via Eq. (1.1). These relations turn out to be an infinite number of nonlinear sum rules. Appendices C and D are devoted to developing the mathematical technique used in the asymptotic expansion method, in complete analogy with Paper I. Finally, we note that the present paper contains the main results of I, because the usual Schrödinger equation can be obtained by putting $V_2(\rho) = 0$ and $V_1(\rho) = -V(\rho)$ in Eq. (1.2).

2. THE ASYMPTOTIC EXPANSION METHOD

Every solution $y(k, E; \rho)$ of Eq. (1.2), regarded as a function of the linear momentum k in the complex k plane, has the two branch points $k = \pm iM$ because of the relativistic expression of the total energy as a double-valued function of the momentum

$$E(k) = (k^2 + M^2)^{\frac{1}{2}}. \quad (2.1)$$

The complex k plane will be considered cut along the imaginary axis from iM to $+i\infty$ and from $-iM$ to $-i\infty$, so that the solution[12] $y(k, E; \rho)$ is defined on the corresponding two-sheet Riemann surface. We will frequently use the fact that a function $F(k, E)$, which is an entire function of the variable E as in our present situation,[9] can be written in the following convenient way to handle the discontinuity introduced in the k plane by the relation (2.1):

$$\begin{aligned} F(k, E) &= \tfrac{1}{2}[F(k, E) + F(k, -E)] \\ &\quad + E[F(k, E) - F(k, -E)]/2E \\ &= F_1(k) + EF_2(k), \end{aligned} \quad (2.2)$$

[7] R. Jost and W. Kohn, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd. **27**, No. 9 (1953).
[8] E. Corinaldesi, Nuovo Cimento **11**, 468 (1954).
[9] M. Verde, Nucl. Phys. **9**, 255 (1958).
[10] I. M. Gel'fand and B. M. Levitan, Iz. Akad. Nauk SSSR **15**, 309 (1951).
[11] V. De Alfaro, Nuovo Cimento **10**, 675 (1958).
[12] We assume, of course, that the boundary conditions do not introduce any other cut in the k plane.

where the functions $F_1(k)$ and $F_2(k)$ have no branch point at $k = \pm iM$. This notation has the advantage of explicitly showing the analytic structure caused by the relativistic kinematics; furthermore, the dispersion-relations technique shall be applied directly to the functions $F_1(k)$ and $F_2(k)$.

We now consider the "Jost solutions," i.e., those solutions of Eq. (1.2) defined by the ingoing- and outgoing-wave asymptotic conditions

$$\lim_{\rho\to\infty} e^{\pm ik\rho} f(\pm k, E; \rho) = 1. \tag{2.3}$$

The Jost function

$$f(k, E) \equiv f(k, E; 0) \tag{2.4}$$

is a holomorphic function in the half-plane $\operatorname{Im} k \leq 0$, excluding the relativistic cut from $-iM$ to $-i\infty$ on the imaginary axis. The zeros, if any, of the Jost function can be there only on the imaginary axis and, by assumption, with $\operatorname{Im} k > -M$. It is well known, furthermore, that the zeros which lie on the positive-energy Riemann sheet $\{k_n = -ip_n \mid n = 1, \cdots, N^+\}$ and on the negative-energy Riemann sheet $\{k_l = -ip_l \mid l = N^+ + 1, \cdots, N^+ + N^-\}$ correspond, respectively, to the bound states of the particle and of the antiparticle

$$f(-ip_n, E_n) = 0,\ M > p_n > 0,\ |E_n| = +(M^2 - p_n^2)^{\frac{1}{2}},$$
$$-M < E_{N^+ + N^-}, \cdots, E_{N^+ + 1} < 0 < E_{N^+}, \cdots, E_1 < M. \tag{2.5}$$

On the other hand, on the real axis the phase of the Jost function is just the scattering phase shift

$$f(k, E) = |f(k, E)| \exp i\eta(k, E)$$
$$= |f(k, E)| \exp i[\eta_1(k) + E\eta_2(k)], \quad \operatorname{Im} k = 0, \tag{2.6}$$

so that the scattering phase shift $\eta_+(k)$ for the particle and $\eta_-(k)$ for the antiparticle are

$$\eta_\pm(k) = \eta_1(k) \pm |E|\, \eta_2(k), \quad \operatorname{Im} k = 0. \tag{2.7}$$

As we show in Appendix B, the analyticity properties of the Jost function can be used to obtain an expression for the modulus $|f(k, E)|$ as a functional of the phase shift $\eta(k, E)$ and of the energies E_n of the bound states:

$$|f(k, E)|^2 = \left(\frac{E + M}{E - M}\right)^{N^+ - N^-} \prod_{n=1}^{N^+ + N^-} \left(\frac{E - E_n}{E + E_n}\right)\left(1 + \frac{p_n^2}{k^2}\right)$$
$$\times \exp\left(\frac{-2}{\pi} P\int_0^\infty \frac{q\, dq}{q^2 - k^2}\left\{\eta_+(q) + \eta_-(q) + \frac{E(k)}{E(q)}[\eta_+(q) - \eta_-(q)]\right\}\right), \quad \operatorname{Im} k = 0. \tag{2.8}$$

As in I, we introduce the function

$$g(k, E; \rho) = \frac{d}{d\rho}\{\ln [e^{ik\rho} f(k, E; \rho)]\}$$
$$= g_1(k, \rho) + Eg_2(k, \rho); \tag{2.9}$$

then Eq. (1.2) implies the following system of two nonlinear equations:

$$\frac{d}{d\rho} g_1(k, \rho) + V_1(\rho) - 2ikg_1(k, \rho) + g_1^2(k, \rho) + (k^2 + M^2)g_2^2(k, \rho) = 0, \tag{2.10a}$$

$$\frac{d}{d\rho} g_2(k, \rho) + V_2(\rho) + 2g_1(k, \rho)g_2(k, \rho) - 2ikg_2(k, \rho) = 0, \tag{2.10b}$$

with the boundary conditions

$$g_1(k, \infty) = g_2(k, \infty) = 0. \tag{2.11}$$

The property of holomorphy of the solution $f(k, E; \rho)$ for $\operatorname{Im} k < 0$,[9] excluding the relativistic cut, implies the meromorphy of the functions $g_1(k, \rho)$ and $g_2(k, \rho)$ in the half-plane $\operatorname{Im} k < 0$. The assumed holomorphy of the potentials then leads to the asymptotic expansions

$$g_1(k, \rho) = \sum_{n=0}^{N} g_1^{(n)}(\rho)(-2ik)^{-n-1} + O(k^{-N-2}), \tag{2.12a}$$

$$g_2(k, \rho) = \sum_{n=0}^{N} g_2^{(n)}(\rho)(-2ik)^{-n-1} + O(k^{-N-2}). \tag{2.12b}$$

Inserting these expansions in the system (2.10), we obtain the recursion relations

$$\frac{d}{d\rho} g_1^{(n+1)}(\rho) + g_1^{(n+2)}(\rho) + \sum_{m=0}^{n} g_1^{(n-m)}(\rho)g_1^{(m)}(\rho) + M^2 \times \sum_{m=0}^{n} g_2^{(n-m)}(\rho)g_2^{(m)}(\rho) - \tfrac{1}{4}\sum_{m=0}^{n+2} g_2^{(n-m+2)}(\rho)g_2^{(m)}(\rho) = 0, \tag{2.13a}$$

$$\frac{d}{d\rho} g_2^{(n+1)}(\rho) + g_2^{(n+2)}(\rho) + 2\sum_{m=0}^{n} g_1^{(n-m)}(\rho)g_2^{(m)}(\rho) = 0, \quad n \geq 0. \tag{2.13b}$$

These equations, together with the initial conditions

$$g_1^{(0)}(\rho) = \tfrac{1}{4}[V_2^2(\rho) - 4V_1(\rho)],$$
$$g_1^{(1)}(\rho) = \frac{d}{d\rho} V_1(\rho) - V_2(\rho)\frac{d}{d\rho} V_2(\rho), \tag{2.14a}$$
$$g_2^{(0)}(\rho) = -V_2(\rho),$$
$$g_2^{(1)}(\rho) = \frac{d}{d\rho} V_2(\rho), \tag{2.14b}$$

are sufficient to express by recursion all the real functions $g_l^{(n)}(\rho)$ in terms of the potentials and of their derivatives, using only the derivative operation. On the other hand, it is easy to prove by induction that the pth derivative of the potentials has the following expression:

$$\frac{d^p}{d\rho^p} V_l(\rho) = F_l^{(p)}[g_1^{(i)}(\rho), g_2^{(i)}(\rho); M^2], \quad l = 1, 2. \tag{2.15}$$

Here, the functions $F_l^{(p)}$, $l = 1, 2$, are sums of products of the arguments shown, the functions $g_1^{(i)}(\rho)$ and $g_2^{(i)}(\rho)$ entering in the expressions (2.15) for i up to p, with the obvious restriction that each term of these sums must have the dimensions $\{k^{p+3-l}\}$. We want to direct attention to the fact that all these products appear in each sum (2.15) multiplied by a numerical coefficient which is independent of the potentials. In fact, these coefficients come just from the structure of Eqs. (2.10) and could be computed once and for all. The values of these coefficients for p up to 5, given in Appendix D, have been obtained by computer with a program in FORMAC language.

The next step of this method consists in taking the expressions (2.15) for ρ equal to zero, because, as we show in Appendix C, the values at the origin of the functions $g_l^{(i)}(\rho)$,

$$g_l^{(i)} \equiv g_l^{(i)}(0), \quad l = 1, 2, \tag{2.16}$$

are explicit functionals of the scattering and bound-state parameters of the particle and antiparticle, $\{\eta_+(k), \eta_-(k), E_n, C_n\}$. To construct these functionals, we use the fact that the function

$$\begin{aligned} g(k, E) &\equiv g(k, E; 0) \\ &= ik + \frac{1}{f(k, E)} \frac{d}{d\rho} f(k, E; \rho)\Big|_{\rho=0} \\ &= g_1(k) + E g_2(k) \end{aligned} \tag{2.17}$$

is such that, for real k,

$$\operatorname{Im} g(k, E) = k(1 - |f(k, E)|^{-2}), \quad \operatorname{Im} k = 0. \tag{2.18}$$

This follows from the Wronskian relation

$$f(k, E; \rho) \frac{d}{d\rho} f(-k, E; \rho) - f(-k, E; \rho) \frac{d}{d\rho} f(k, E; \rho) = 2ik, \tag{2.19}$$

together with the reflection property

$$f(k, E; \rho) = f^*(-k, E; \rho), \quad \operatorname{Im} k = 0. \tag{2.20}$$

Furthermore, the definition (2.17) implies that the two branches of the function $g(k, E)$ have simple poles corresponding to the binding energies E_n (see Appendix C) in such a way that the function

$$\tilde{g}(k, E) = g(k, E) + \tfrac{1}{2} \sum_{n=1}^{N^+ + N^-} \frac{C_n}{k^2 + p_n^2} \left(1 + \frac{E}{E_n}\right) \tag{2.21}$$

has no pole ($\operatorname{Im} k \leq 0$). The real positive numbers C_n, which are simply related to the residues of the function (2.17) at the poles, are given by the expression (A2b), as we show in Appendix C. On the other hand, the properties (2.18) and (2.21) can be written down for the functions $g_l(k)$, $l = 1, 2$, as follows:

$$\begin{aligned} &\operatorname{Im} g_1(k) = \operatorname{Im} \tilde{g}_1(k) = k(1 - S_1(k)), \\ &\operatorname{Im} g_2(k) = \operatorname{Im} \tilde{g}_2(k) = -kS_2(k), \\ &\operatorname{Im} k = 0, \end{aligned} \tag{2.22}$$

where we have introduced for simplicity the function

$$S(k, E) = S_1(k) + ES_2(k) = |f(k, E)|^{-2}, \tag{2.23}$$

which is a known functional of the phase shift and the bound-state energies [see Eq. (2.8)]. Therefore, the real part of the functions $\tilde{g}_1(k)$ and $\tilde{g}_2(k)$ is also a known functional of the scattering phase shift and of the binding energies, because of the dispersion relations

$$\operatorname{Re} \tilde{g}_1(k) = \frac{1}{\pi} P \int_{-\infty}^{+\infty} dq \, \frac{q^2}{q^2 - k^2} [S_1(q) - 1], \quad \operatorname{Im} k = 0, \tag{2.24a}$$

$$\operatorname{Re} \tilde{g}_2(k) = \frac{1}{\pi} P \int_{-\infty}^{+\infty} dq \, \frac{q^2}{q^2 - k^2} S_2(q), \quad \operatorname{Im} k = 0. \tag{2.24b}$$

We note, incidentally, also for future reference, that the function $S(k, E)$ and, hence, the functions $S_1(k)$ and $S_2(k)$, are even,

$$S_l(k) = S_l(-k), \quad l = 1, 2, \quad \operatorname{Im} k = 0, \tag{2.25}$$

while the phase shift (2.7) is odd, as implied by the reflection property (2.20):

$$\eta_l(k) = -\eta_l(-k), \quad l = 1, 2, \quad \operatorname{Im} k = 0. \tag{2.26}$$

Furthermore, the property (2.25), together with the expressions (2.22) and (2.24), implies

$$\operatorname{Re} \tilde{g}_l(k) = \operatorname{Re} \tilde{g}_l(-k), \quad \operatorname{Im} \tilde{g}_l(k) = -\operatorname{Im} \tilde{g}_l(-k), \quad l = 1, 2, \quad \operatorname{Im} k = 0. \tag{2.27}$$

The integrals (2.24) are convergent because, at large k,

$$S_1(k) = 1 + O(k^{-2}), \quad S_2(k) = O(k^{-2}), \quad \text{Im}\, k = 0, \tag{2.28}$$

and they imply for $\tilde{g}_l(k)$ the asymptotic behavior

$$\text{Re}\, \tilde{g}_l(k) = O(k^{-2}), \quad \text{Im}\, \tilde{g}_l(k) = O(k^{-1}), \quad \text{Im}\, k = 0, \quad l = 1, 2, \tag{2.29}$$

which is already a consequence of expressions (2.12), (2.17), and (2.21). It should be emphasized that the function $\tilde{g}(k, E)$ is known once the phase shift $\eta(k, E)$ and the bound-state energies E_n are given, because of Eqs. (2.8), (2.23), (2.22), and (2.24). To know the function $g(k, E)$, it is further necessary to fix the values of the bound-state parameters C_n. The dependence on these parameters is explicitly shown by Eq. (2.21).

We are now in the position to go back to the quantities (2.16) that define the asymptotic expansion

$$g_l(k) = \sum_{n=0}^{N} g_l^{(n)}(-2ik)^{-n-1} + O(k^{-N-2}) \tag{2.30}$$

and to find their expression in terms of the scattering parameters $\{\eta_+(k), \eta_-(k), E_n, C_n\}$. As in Paper I, we use a dispersion-relation technique of Roberts,[13] that we develop in Appendix C. We find that the coefficients $g_l^{(n)}$, $l = 1, 2$, can be expressed in terms of these four kinds of "building-block" quantities

$$J_1^{(2n)} = \frac{-2^{2n+1}}{n}\left(\frac{(-1)^n 4n}{\pi(2n-1)!} \times \int_0^\infty dk\, D_k^{2n-1}[\eta_1(k)] + \sum_{m=1}^{N^+ + N^-} p_m^{2n}\right), \quad n \geq 1 \tag{2.31}$$

$$J_1^{(2n+1)} = 2^{2n-1}\left(\frac{(-1)^n 4}{\pi(2n+1)!} \times \int_0^\infty dk\, D_k^{2n+1}[k^{-1}S_1(k)] + \sum_{m=1}^{N^+ + N^-} C_m p_m^{2n-2}\right) = g_1^{(2n-1)}, \quad n \geq 1, \tag{2.32}$$

$$J_2^{(2n+1)} = \frac{-2^{2n+1}}{n}\left(\frac{(-1)^n 2}{\pi(2n)!} \times \int_0^\infty dk[D_k^{2n+1} + (2nM)^2 D_k^{2n-1}][\eta_2(k)] + \sum_{m=1}^{N^+ + N^-} E_m p_m^{2n}\right), \quad n \geq 1, \tag{2.33a}$$

$$J_2^{(1)} = -\frac{4}{\pi}\int_0^\infty dk\, D_k[\eta_2(k)] - 2\sum_{n=1}^{N^+ + N^-} E_m + 2M(N^+ - N^-), \tag{2.33b}$$

$$J_2^{(2n)} = 2^{2n-1}\left(\frac{(-)^n 4}{\pi(2n+1)!} \times \int_0^\infty dk\, D_k^{2n+1}[k^{-1}S_2(k)] + \sum_{m=1}^{N^+ + N^-} \frac{C_m}{E_m} p_m^{2n-2}\right) = g_2^{(2n-1)}, \quad n \geq 1. \tag{2.34}$$

It turns out that the even coefficients $g_l^{(2n)}$ are expressed as sums of products of $J_1^{(2m)}$, $J_2^{(2m+1)}$ and of positive powers of M^2 (see Appendix C where we give such explicit expression for n up to 2), while the odd coefficients $g_l^{(2n-1)}$ are given simply by the relations (2.32) and (2.34). We finally substitute the coefficients $g_l^{(n)}$ with these expressions into the equations (2.15) with vanishing ρ to get our relationships between the potentials, the scattering phase shifts and the bound-state parameters. We condense our results in the formulas

$$\frac{d^n}{d\rho^n} V_p(\rho)\Big|_{\rho=0} = \sum \Gamma_p(n, m, l_1, l_2; s_2, \cdots, s_{l_1+1}; t_1, \cdots, t_{l_2}) M^{2m} \times \prod_{j=2}^{l_1+1} [J_1^{(j)}]^{s_j} \prod_{i=1}^{l_2} [J_2^{(i)}]^{t_i}, \quad p = 1, 2. \tag{2.35}$$

This sum extends over the integral values of m between $m = 0$ and $m = \frac{1}{2}n$ if n is even and $m = \frac{1}{2}(n-1)$ if n is odd, and over all the nonnegative integral values of $l_1, l_2, s_2, \cdots, s_{l_1+1}, t_1, \cdots, t_{l_2}$ satisfying the dimensional condition

$$2m + \sum_{j=2}^{l_1+1} j s_j + \sum_{i=1}^{l_2} i t_i = n + 1 + \delta_{p1}, \quad p = 1, 2, \tag{2.36}$$

with the convention that $s_{l_1+1}[t_{l_2}]$ can be zero only if all the $s_j[t_i]$ are zero. We formally indicate this case putting $l_1 = 0$ ($l_2 = 0$). However, the vanishing of both s_{l_1+1} and t_{l_2} can never occur because of the maximum value of m. Thus, the maximum value of l_1 and l_2 appearing in Eq. (2.35) is

$$l_{q\,\max} = n + q - \delta_{p2}, \quad q = 1, 2, \quad p = 1, 2. \tag{2.37}$$

We note that the numerical coefficients $\Gamma_p(n, m, l_1, l_2; s_2, \cdots, s_{l_1+1}; t_1, \cdots, t_{l_2})$ are independent of the potentials and could be computed once and for all. They are given for n up to 3 in Table I.

[13] M. J. Roberts, Proc. Phys. Soc. (London) **80**, 1290 (1962).

TABLE I. The coefficients $\Gamma_p(n, m, l_1, l_2; s_2, \cdots s_{l_1+1}; t_1, \cdots t_{l_2})$ of Eq. (2-35) for $n \leq 3$. Only the nonvanishing coefficients are given.

$\Gamma_1(0,0,1,0;1;\cdots) = \frac{1}{4}$,	$\Gamma_2(0,0,0,1;\cdots;1) = 2$	
$\Gamma_1(1,0,0,2;\cdots;1\,1) = 2$,	$\Gamma_1(1,0,2,0;0\,1;\cdots) = 1$,	$\Gamma_2(1,0,0,2;\cdots;0\,1) = 1$
$\Gamma_1(2,0,1,0;2;\cdots) = -\frac{1}{8}$,	$\Gamma_1(2,0,1,1;1;2) = -7$,	$\Gamma_1(2,0,3,0;0\,0\,1;\cdots) = \frac{1}{4}$
$\Gamma_1(2,0,0,1;\cdots;4) = \frac{44}{3}$,	$\Gamma_1(2,0,0,3;\cdots;1\,0\,1) = \frac{10}{3}$,	$\Gamma_1(2,0,0,2;\cdots;0\,2) = \frac{5}{4}$
$\Gamma_1(2,1,0,1;\cdots;2) = \frac{80}{3}$,	$\Gamma_2(2,0,0,1;\cdots;3) = \frac{20}{3}$,	$\Gamma_2(2,0,0,3;\cdots;0\,0\,1) = \frac{4}{3}$
$\Gamma_2(2,0,1,1;1;1) = -3$,	$\Gamma_2(2,1,0,1;\cdots;1) = \frac{32}{3}$	
$\Gamma_1(3,0,0,2;\cdots;3\,1) = 30$,	$\Gamma_1(3,0,0,3;\cdots;0\,1\,1) = 6$,	$\Gamma_1(3,0,0,4;\cdots;1\,0\,0\,1) = 4$
$\Gamma_1(3,0,1.2;1;1\,1) = -\frac{31}{2}$,	$\Gamma_1(3,0,2,0;1\,1;\cdots) = -1$,	$\Gamma_1(3,0,2,1;0\,1;2) = -16$
$\Gamma_1(3,0,4,0;0\,0\,0\,1;\cdots) = 1$,	$\Gamma_1(3,1,0,2;\cdots;1\,1) = 40$,	$\Gamma_2(3,0,0,4;\cdots;0\,0\,0\,1) = 1$
$\Gamma_2(3,0,1,2;1;0\,1) = -1$,	$\Gamma_2(3,0,2,1;0\,1;1) = -8$	

3. REMARKS ON THE KG EQUATION

Provided the solution of the Gel'fand–Levitan equation exists and is unique (see Appendix A), the two potentials appearing in Eq. (1.2) can be obtained from the scattering phase shifts and the bound-state parameters $\{\eta_+(k), \eta_-(k), E_n, C_n\}$ by the method discussed in Appendix A. This means that special relationships between the potentials $V_1(\rho)$ and $V_2(\rho)$ must be reflected in a relationship between the scattering and bound-state parameters, and vice versa. The simplest example is the case $V_2(\rho) = 0$, which corresponds to the relationships $\eta_+(k) = \eta_-(k)$, $N^+ = N^- = N$, $E_m = -E_{N+m}$ and $C_m = C_{N+m}$, for $m = 1, 2, \cdots, N$.[14]

The other case of interest is the KG equation of a spinless particle in an electriclike potential; this equation is obtained with the relation (1.3). In the Gel'fand–Levitan formalism, using two comparison potentials $V_{0_1}(\rho)$ and $V_{0_2}(\rho)$ satisfying the same relationship as $V_1(\rho)$ and $V_2(\rho)$, we express the two previous cases as a relation between the two components of the solution of the integral equation (A6). In the case $V_2(\rho) = 0$, we obtain $K_1(\rho, \rho) = 0$ so that the expressions (A8) reduce to the well-known nonrelativistic formula

$$V_1(\rho) - V_{01}(\rho) = -2\frac{d}{d\rho} K_2(\rho, \rho). \tag{3.1}$$

In the case of the KG equation, the following expression can be easily obtained:

$$K_1(\rho, \rho)\frac{d}{d\rho} \arctan K_1(\rho, \rho) = 2K_2(\rho, \rho), \tag{3.2}$$

while the potential is given by

$$V(\rho) - V_0(\rho) = 2K_2(\rho, \rho)/K_1(\rho, \rho). \tag{3.3}$$

[14] When $V_2(\rho) = 0$, Eq. (1.2) does not contain any relation between the momentum k and the energy E, so that we can formally define as energy any function of k. If we choose the nonrelativistic definition $E = k^2/2M$, Eq. (1.2) with $V_1(\rho) = -V(\rho)$ reduces to the Schrödinger equation, so that the results of Paper I are implicitly contained in this paper.

Before going to the asymptotic-expansion method developed in Sec. 2, we mention a simple theorem that can be proved with the Gel'fand–Levitan formalism. Let us call "physically equivalent potentials" two potentials that produce, via the KG equation (1.1), the same phase shifts $\eta_+(k)$ and $\eta_-(k)$, the same energies E_n of the bound states, but different normalization constants C_n. Then, the theorem states that the parameters C_n cannot change continuously in the set of the "physically equivalent" potentials. To prove it, we make use of the partial derivative of the potentials $V_1(\rho)$ and $V_2(\rho)$ with respect to the parameters C_n, following the same procedure used by Newton,[15] to show that there is no direction $(\alpha_1, \alpha_2 \cdots \alpha_{N^+ + N^-})$ in the space of the $N^+ + N^-$ variables C_n, such that

$$\sum_{n=1}^{N^+ + N^-} \alpha_n \frac{\partial}{\partial C_n} [V_1(\rho) - \tfrac{1}{4}V_2^2(\rho)] = 0. \tag{3.4}$$

Using the following expressions for the partial derivatives,

$$\frac{\partial V_2(\rho)}{\partial C_n} = \frac{1}{E_n}\frac{d}{d\rho}[\varphi(E_n; \rho)]^2,$$

$$\frac{\partial V_1(\rho)}{\partial C_n} = \left(1 + \frac{V_2(\rho)}{E_n}\right)\frac{d}{d\rho}[\varphi(E_n; \rho)]^2 + \frac{1}{2E_n}[\varphi(E_n; \rho)]^2 \frac{d}{d\rho} V_2(\rho), \tag{3.5}$$

and the relations (1.3) which define the KG equation, we get the condition

$$\sum_{n=1}^{N^+ + N^-} \alpha_n [\varphi(E_n; \rho)]^2 \left(1 - \frac{V(\rho)}{E_n}\right) = 0, \tag{3.6}$$

which can never be satisfied.

We now apply the results obtained in the previous section to the KG equation. The expressions of the

[15] R. G. Newton, Phys. Rev. **101**, 1588 (1956).

potential $V(\rho)$ and of its derivatives at the origin, in terms of the phase shifts and bound-state parameters, are given by the formula (2.35), for $p = 2$:

$$\frac{d^n}{d\rho^n} V(\rho)\Big|_{\rho=0} = -\tfrac{1}{2}\sum \Gamma_2(n, m, l_1, l_2; s_2 \cdots, s_{l_1+1}; t_1 \cdots t_{l_2}) \times \prod_{j=2}^{l_1+1} [J_1^{(j)}]^{s_j} \prod_{i=1}^{l_2} [J_2^{(i)}]^{t_i}. \quad (3.7)$$

The potential and its first derivative for vanishing ρ are, for example,

$$V(0) = -\frac{2}{\pi}\int_0^\infty dk k \frac{d}{dk}\left(\frac{k}{E}[\eta_+(k) - \eta_-(k)]\right) - 2\sum_{n=1}^{N^+ + N^-} E_n + 2M(N^+ - N^-), \quad (3.8)$$

$$V'(0) = -\frac{1}{\pi}\int_0^\infty dk k^2 \frac{d^2}{dk^2}\left(\frac{k^2}{E}(|f(k, E)|^{-2} - |f(k, -E|^{-2})\right) - \sum_{n=1}^{N^+ + N^-} \frac{C_n}{E_n}. \quad (3.9)$$

The same formula (2.35), for $p = 1$, gives us the square of the potential and its derivatives at the origin in terms of the scattering and bound-state parameters. Substituting in these expressions the derivatives of the potentials given by Eq. (3.7), we get nontrivial relations between the phase shifts, the binding energies, and the normalization constants. To calculate these sum rules, it is more convenient to use the fact that the function $g_1^{(0)}(\rho)$, defined by the asymptotic expansion (2.12a) and explicitly given by the formula (2.14a), vanishes in our case. We set $g_1^{(0)}(\rho) = 0$ in Eq. (D2a) and let ρ go to zero, obtaining

$$g_1^{(n)} = (-1)^{n+1} A_1^{(n)}[g_1^{(0)}, g_1^{(1)}, \cdots, g_1^{(n-2)}; g_2^{(0)}, g_2^{(1)}, \cdots, g_2^{(n)}; M^2], \quad n \geq 1. \quad (3.10)$$

These relations can, in principle, be solved recursively, giving all the coefficients $g_1^{(n)}$ in terms of the coefficients $g_2^{(m)}$ for m up to n. Our sum rules are now obtained using the expressions (2.32), (2.34), (C49), (C47), (C51), (2.31), and (2.33). For example, the first four of such sum rules read

$$J_1^{(2)} - 4[J_2^{(1)}]^2 = 0, \quad (3.11)$$

$$J_1^{(3)} + J_2^{(1)} J_2^{(2)} = 0, \quad (3.12)$$

$$J_1^{(4)} + 8J_2^{(3)}J_2^{(1)} + 20[J_2^{(1)}]^4 - 14J_1^{(2)}[J_2^{(1)}]^2 - \tfrac{1}{4}[J_1^{(2)}]^2 + 3[J_2^{(2)}]^2 + 64M^2[J_2^{(1)}]^2 = 0, \quad (3.13)$$

$$J_1^{(5)} + 3J_2^{(4)}J_2^{(1)} + 4J_2^{(3)}J_2^{(2)} + 16J_2^{(2)}[J_2^{(1)}]^3 + 6J_1^{(2)}J_2^{(1)}J_2^{(2)} - 24M^2 J_2^{(2)}J_2^{(1)} = 0. \quad (3.14)$$

These equations can be used to express the energies and the normalization constants of the bound states in terms of the phase shifts. If only one positive-energy bound state exists, the normalization constant is

$$C_1 = \frac{2}{3\pi}\left[2\left(\frac{2}{\pi}\int_0^\infty dk D_k[\eta_2(k)] + E_1 - M\right) \times \int_0^\infty dk D_k^3[k^{-1}S_2(k)] - \int_0^\infty dk D_k^3[k^{-1}S_1(k)]\right] \times \left(\frac{4}{E_1\pi}\int_0^\infty dk D_k\eta_2(k) - \frac{2M}{E_1} + 1\right)^{-1}. \quad (3.15)$$

ACKNOWLEDGMENTS

The author wishes to thank Professor F. Calogero who suggested the problem, Dr. O. D. Corbella for the useful conversations, Dr. A. Rambaldi and Dr. A. Morgana who used the FORMAC language to compute the coefficients given in Table II. Also he wants to acknowledge the hospitality of the Department of Physics of Indiana University where this paper has been completed and thank Professor R. G. Newton for many stylistic improvements of the manuscript.

APPENDIX A

In this Appendix we summarize the Gel'fand–Levitan approach to the inverse problem for the wave equation (1.2). The main features of this equation are well known: the energy spectrum consists of a continuous part $|E| \geq M$ plus a finite number of discrete energy values $-M < E_l < M$. The corresponding solutions $\varphi(E; \rho)$ may be arranged in such a way to form a basis in a two-component function space

$$\boldsymbol{\varphi}(E; \rho) = \varphi(E; \rho)\begin{pmatrix} 1 \\ E + \frac{1}{2}V_2(\rho) \end{pmatrix}, \quad \lim_{\rho\to 0} \rho^{-1}\varphi(E; \rho) = 1, \quad (A1)$$

satisfying the quasi-orthogonality relations[16]

$$\int_0^\infty \boldsymbol{\varphi}^+(E; \rho)\boldsymbol{\Sigma}_1\boldsymbol{\varphi}(E'; \rho)\, d\rho = \frac{\pi}{k}\frac{E}{|E|}|f(k, E)|^2 \delta(E' - E), \quad |E| > M, \quad (A2a)$$

$$\int_0^\infty \boldsymbol{\varphi}^+(E_m; \rho)\boldsymbol{\Sigma}_1\boldsymbol{\varphi}(E_n; \rho)\, d\rho = 2E_n C_n^{-1}\delta_{mn}, \quad -M < E_m, \quad E_n < M, \quad (A2b)$$

[16] The relation (A2a) is not correct for $E' = -E$ because the scalar product is given by

$$\int_0^\infty \boldsymbol{\varphi}^+(-E; \rho)\boldsymbol{\Sigma}_1\, \boldsymbol{\varphi}(E; \rho)\, d\rho = (2Ek)^{-1}|f(k, E)f(k, -E)| \times \sin[\eta_+(k) - \eta_-(k)].$$

This, however, has no effect on the considerations that follow.

where Σ_1 is the Pauli matrix $\left(\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right)$ and $f(k, E)$ is the Jost function [see definitions (2.3) and (2.4)]. Note that the equation (A2b) defines the bound-state parameters C_n. The completeness relations are[17]

$$\int_{-\infty}^{+\infty} d\tau(E)\boldsymbol{\varphi}(E;\rho)\boldsymbol{\varphi}^{+}(E;\rho')\Sigma_1 = \delta(\rho-\rho')\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \tag{A3}$$

where the spectral function is

$$\frac{d\tau(E)}{dE} = \frac{k}{\pi}\frac{E}{|E|}|f(k,E)|^{-2}, \qquad \text{if } |E| > M,$$
$$\frac{d\tau(E)}{dE} = \sum_{n=1}^{N^+ + N^-}\frac{C_n}{2E_n}\delta(E-E_n), \quad \text{if } |E| < M. \tag{A4}$$

The connection between the spectral function (A4) and the potentials $V_1(\rho)$ and $V_2(\rho)$ is given by the Gel'fand–Levitan integral equation. The best way to derive this equation is the dispersion-relations technique of Verde[9]; so we refer to that paper for a detailed derivation of the integral equation. We limit ourselves to writing down the extension of Verde's results[18] to Eq. (1.2). Let $V_{01}(\rho)$ and $V_{02}(\rho)$ be two given potentials. The regular solution is

$$\varphi(E;\rho) = \cos\left[\tfrac{1}{2}\int_0^{\rho}(V_2(x) - V_{02}(x))\,dx\right] \times\left[\varphi_0(E;\rho) + \int_0^{\rho} dt\boldsymbol{\varphi}_0^{+}(E;t)\Sigma_1\mathbf{k}(\rho,t)\right], \tag{A5}$$

where the two-component kernel function $\mathbf{k}(\rho, t)$ is the solution of the Gel'fand–Levitan integral equation

$$\mathbf{k}(x,t) = \tilde{\mathbf{k}}(x,t) + \int_0^{x} dy\tilde{\mathbf{K}}(t,y)\mathbf{k}(x,y) \equiv \begin{pmatrix}K_1(x,t)\\ K_2(x,t)\end{pmatrix}, \qquad x \geq t \geq 0. \tag{A6}$$

The term $\tilde{\mathbf{k}}(x, t)$ and the 2×2 matrix function $\tilde{\mathbf{K}}(x, t)$ are obtained from the spectral function through the integrals

$$\tilde{\mathbf{k}}(x,t) = \int_{-\infty}^{+\infty}(d\tau_0(E) - d\tau(E))\varphi_0(E;x)\boldsymbol{\varphi}_0(E,t),$$
$$\tilde{\mathbf{K}}(x,t) = \int_{-\infty}^{+\infty}(d\tau_0(E) - d\tau(E))\boldsymbol{\varphi}_0(E;x)\boldsymbol{\varphi}_0^{+}(E;t)\Sigma_1. \tag{A7}$$

The differences with the comparison potentials are found to be

$$V_1(\rho) - V_{01}(\rho) = \left[\frac{d}{d\rho}\arctan K_1(\rho,\rho)\right]^2 - V_{02}(\rho)\frac{d}{d\rho}\arctan K_1(\rho,\rho) + \frac{d}{d\rho}\left[K_1(\rho,\rho)\frac{d}{d\rho}\arctan K_1(\rho,\rho) - 2K_2(\rho,\rho)\right], \tag{A8a}$$

$$V_2(\rho) - V_{02}(\rho) = -2\frac{d}{d\rho}\arctan K_1(\rho,\rho). \tag{A8b}$$

In Appendix B we show that the modulus of the Jost function for real k is a known functional of the phase shift $\eta(k, E)$ and of the binding energies, so that the spectral function (A4) is in a 1-to-1 correspondence with the scattering and bound-state parameters $\{\eta_+(k), \eta_-(k), E_n, C_n\}$. Furthermore, the expressions (A8) show that these parameters allow us to construct the two potentials, provided the solution of the integral equation (A6) exists and is unique. However, the conditions that the phase shift and the bound-state parameters have to satisfy, in order that the Gel'fand–Levitan equation (A6) has a unique solution, are still not known. Because of the non-self-adjoint-ness of the differential equation (1.2), the standard method used for the Schrödinger and Dirac equations cannot be applied. However, it is possible to prove, using the behavior for large k of the regular solution $\varphi(E; \rho)$, that Eq. (A6) can be transformed into a Fredholm integral equation for every fixed x. On the other hand, the kernel $\tilde{\mathbf{k}}^*(x, t)$ of the adjoint integral equation is given by $\Sigma_1\tilde{\mathbf{k}}(x, t)\Sigma_1$, so that if $\mathbf{k}(x, t)$ is a solution of the homogeneus integral equation [i.e., Eq. (A6) without the term $\tilde{\mathbf{k}}(x, t)$], then $\Sigma_1\mathbf{k}(x, t)$ is a solution of the adjoint homogeneous equation. This means that it remains to prove only that such a solution of the homogeneous integral equation is identically zero for every positive x. We note that in the nonrelativistic case, in addition to the usual conditions on the k dependence of the phase shift and on the values of the bound-state energy, the condition $C_n > 0$ guarantees the uniqueness of the solution of the Gel'fand–Levitan equation; this condition is quite natural because the numbers C_n^{-1} are the norm of the bound-state wavefunctions. We do not expect such a general condition in our case. In fact, we find, for example,

[17] The completeness relations can be obtained by standard methods constructing the Green's function for the nonstationary equation

$$((\partial^2/\partial t^2) - iV_2(\rho)(\partial/\partial t) - V_1(\rho))\Psi(\rho,t) = ((\partial^2/\partial\rho^2) - M^2)\Psi(\rho,t).$$

[18] Our results do not agree completely with the formulas given by Verde, Ref. 9. In addition to Eq. (A5), he obtains a second equation for the function $\varphi(E; \rho)$ that we believe to be incorrect. Furthermore, we do not find the mass-dependent term appearing in his relation between the potential and the kernels $k_1(\rho, \rho)$ and $k_2(\rho, \rho)$. Our corresponding formula is Eq. (3.3).

that, for $V_{01}(\rho) = V_{02}(\rho) = 0$, to the spectral function

$$\frac{d}{dE}\tau(E) = \frac{kE}{\pi|E|}(k^2 + p_1^2)^{-1}\left[k^2 + p_1^2 + \frac{C_1}{4p_1} + \frac{C_1 E}{4p_1 E_1}\right], \quad |E| > M,$$

$$\frac{C_1}{2E_1}\delta(E - E_1), \quad |E| < M, \quad E_1 > 0, \qquad \text{(A9)}$$

corresponds an integral equation (A6) that has a unique solution only if $0 < C_1 < 8p_1E_1^2$. The corresponding potentials for $C_1 > 8p_1E_1^2$ have a simple pole at a certain value of ρ. The function $g(k, E)$, introduced in Sec. 2, is, in this case,

$$g(k, E) = -\frac{iC_1}{4p_1}\frac{1}{k + ip_1}\left(1 + \frac{E}{E_1}\right). \qquad \text{(A10)}$$

In the nonrelativistic limit this corresponds to the Wigner-Eckart potential, and the constant C_1 can take on any positive value. (Its upper bound $8p_1E_1^2$ goes to infinity in the nonrelativistic limit.)

APPENDIX B

In this Appendix we prove the relation (2.8) between the modulus and the phase of the Jost function. This formula, for the case in which there are no bound states, can be found in a paper by Corinaldesi.[8] To take into account the contributions of the bound-state energies, we follow the same procedure used in the nonrelativistic case,[19] with the only difference being that now the Jost function is defined in a two-sheet Riemann surface. To look at the asymptotic behavior of the Jost function we note that the definitions (2.4) and (2.9) imply

$$f(k, E) = \exp\left(\int_\infty^0 g_1(k, \rho)\, d\rho + E\int_\infty^0 g_2(k, \rho)\, d\rho\right), \quad \text{Im } k \leq 0, \qquad \text{(B1)}$$

which together with (2.12) and (2.14) give the well-known relativistic property (Parzen[20] theorem)

$$\lim_{k\to\infty} f(k, E) = \exp\left(\tfrac{1}{2}i\int_0^\infty V_2(\rho)\, d\rho \lim_{k\to\infty}\frac{E(k)}{k}\right), \quad \text{Im } k \leq 0. \qquad \text{(B2)}$$

We begin by introducing the function $P(k) = f(k, E)f(k, -E)$, which is holomorphic in the lower half-plane with simple zeros[21] on the imaginary axis, as shown by Eqs. (2.5), so that the function

$$\tilde{P}(k) = \prod_{n=1}^{N^+ + N^-}\left(\frac{k - ip_n}{k + ip_n}\right)P(k) \qquad \text{(B3)}$$

is regular without any zero for $\text{Im } k \leq 0$. We then apply the Cauchy theorem to the regular function $\ln \tilde{P}(k)$, for which the limit (B2) guarantees the asymptotic behavior

$$\lim_{k\to\infty} \ln \tilde{P}(k) = 0. \qquad \text{(B4)}$$

Taking as integration contour a large semicircle in the lower half k plane and the real axis, we get

$$\ln|f(k, E)f(k, -E)| = -\frac{1}{\pi}P\int_{-\infty}^{+\infty}\frac{dk'}{k' - k}[\eta_+(k') + \eta_-(k')] + \sum_{n=1}^{N^+ + N^-}\ln\left(1 + \frac{p_n^2}{k^2}\right), \quad \text{Im } k = 0, \qquad \text{(B5)}$$

$$\eta_+(k) + \eta_-(k) = \frac{1}{\pi}P\int_{-\infty}^{+\infty}\frac{dk'}{k' - k}\ln|f(k', E')f(k', -E')| + 2\sum_{n=1}^{N^+ + N^-}\arctan\frac{p_n}{k}, \quad \text{Im } k = 0. \qquad \text{(B6)}$$

Let us consider now the function $R(k, E) = f(k, E)/f(k, -E)$, which has simple zeros and poles because of the zeros of the Jost function, so that the function

$$\tilde{R}(k, E) = \prod_{n=1}^{N^+ + N^-}\left(\frac{E_n k - ip_n E}{E_n k + ip_n E}\right)R(k, E) \qquad \text{(B7)}$$

is regular in the two Riemann sheets for $\text{Im } k \leq 0$ and satisfies the relation $\tilde{R}(k, E)\tilde{R}(k, -E) = 1$. This means that the function $(k/E)\ln \tilde{R}(k, E)$ has the same analytic properties as the function $\ln \tilde{P}(k)$ except for the high-energy limit

$$\lim_{k\to\infty}\frac{k}{E}\ln \tilde{R}(k, E) = 2i\left(\sum_{n=1}^{N^+ + N^-}\arctan\frac{p_n}{E_n} + \tfrac{1}{2}\int_0^\infty V_2(\rho)\, d\rho\right) = i\gamma. \qquad \text{(B8)}$$

We now use the same contour used before to write down the Cauchy theorem for the function

[19] See the review article by R. G. Newton, J. Math. Phys. **1**, 319 (1960).

[20] G. Parzen, Phys. Rev. **80**, 261 (1950).

[21] As in the nonrelativistic case, these zeros are simple because the expression (C2) holds.

$(k/E)\ln\tilde{R}(k) - i\gamma$; for the real part we obtain

$$\frac{k}{E}\ln\left|\frac{f(k,E)}{f(k,-E)}\right| = -\frac{2}{\pi}P\int_{-\infty}^{+\infty}\frac{dk'}{k'-k}\left[\frac{k'}{2E'}(\eta_+(k')-\eta_-(k')) - \tfrac{1}{2}\int_0^\infty V_2(\rho)\,d\rho\right] + \frac{k}{E}\left[\sum_{n=1}^{N^+N^-}\ln\left(\frac{E-E_n}{E+E_n}\right) - (N^+-N^-)\ln\left(\frac{E-M}{E+M}\right)\right], \quad \mathrm{Im}\,k=0. \quad \text{(B9)}$$

It is possible, however, to avoid the subtraction constant appearing in the dispersion integral by using the reflection property (2.26) of the phase shifts

$$\ln\left|\frac{f(k,E)}{f(k,-E)}\right| = -\frac{E}{\pi}P\int_{-\infty}^{+\infty}\frac{dk'}{(k'-k)E'}[\eta_+(k')-\eta_-(k')] + \sum_{n=1}^{N^+N^-}\ln\left(\frac{E-E_n}{E+E_n}\right) + (N^+-N^-)\ln\left(\frac{E+M}{E-M}\right), \quad \mathrm{Im}\,k=0. \quad \text{(B10)}$$

For the imaginary part, we get the dispersion relation

$$\eta_+(k)-\eta_-(k) = \frac{E}{\pi}P\int_{-\infty}^{+\infty}\frac{dk'}{(k'-k)E'}\ln\left|\frac{f(k',E')}{f(k',-E')}\right| + 2\sum_{n=1}^{N^+N^-}\arctan\frac{p_nE}{kE_n}, \quad \mathrm{Im}\,k=0. \quad \text{(B11)}$$

We note finally that the relations (B5) and (B10) are equivalent to the functional expression (2.8) quoted in Sec. 2.

We also remark that the asymptotic normalization of the phase shifts

$$\lim_{k\to\infty}\eta_\pm(k) = \eta_\pm(\infty) = \pm\int_0^\infty V_2(\rho)\,d\rho, \quad \text{(B12)}$$

which follows from (B2) and (2.6), implies the relation between the phase shift at zero energy and the number of bound states

$$\lim_{k\to 0^+}\eta_\pm(k) = \eta_\pm(0) = N^\pm\pi. \quad \text{(B13)}$$

This is the form taken by Levinson's theorem in the present case. Considering that the modulus of the Jost function is an even function of k, Eq. (B13) is simply the $k=0$ limit of the dispersion relations (B6) and (B11).

APPENDIX C

We now derive the expressions of the asymptotic constants $g_l^{(n)}$, $l=1,2$, $n=0,1,\cdots,\infty$, defined by Eqs. (2.16), (2.12), and (2.9), in terms of the scattering phase shift $\eta_\pm(k)$ and the bound-state parameters E_n and C_n [see Eqs. (2.5)–(2.7), and (A2b)]. We begin by proving the holomorphy of the functions $\tilde{g}_l(k)$, $l=1,2$, for $\mathrm{Im}\,k\leq 0$ stated in Sec. 2. As we have seen, their definition is

$$\tilde{g}(k,E) = ik + [f(k,E)]^{-1}\frac{d}{d\rho}f(k,E;\rho)|_{\rho=0} + \frac{1}{2}\sum_{n=1}^{N^+N^-}\left(1+\frac{E}{E_n}\right)\frac{C_n}{k^2+p_n^2} = \tilde{g}_1(k) + E\tilde{g}_2(k), \quad \text{(C1)}$$

which is equivalent to Eq. (2.9) together with Eq. (2.21). In the lower half of the k plane, excluding the relativistic cut, the meromorphy of the function $\tilde{g}(k,E)$ easily follows from the known holomorphy of the function $f(k,E;\rho)$ and its derivative. What remains to be proved is the fact that the poles owing to the zeros of the Jost function (2.5) are exactly canceled by the poles explicitly contained in the sums in the definition (C1). Thus, it is required to show that

$$C_n = 2ip_n\frac{\partial}{\partial\rho}f(-ip_n,E_n;\rho)|_{\rho=0}\left(\frac{\partial}{\partial k}f(k,E)|_{\substack{k=-ip_n\\E=E_n}}\right)^{-1}. \quad \text{(C2)}$$

This formula can be proved using the Wronskian relation

$$\frac{\partial}{\partial\rho}\left(f(k,E;\rho)\frac{\partial^2 f}{\partial\rho\partial k}(k,E;\rho) - \frac{\partial f}{\partial\rho}(k,E;\rho)\frac{\partial f}{\partial k}(k,E;\rho)\right) = -2k\left(1+\frac{V_2(\rho)}{2E}\right)[f(k,E;\rho)]^2, \quad \text{(C3)}$$

which is easily derived from the differential equation (1.2). We now set $k=-ip_n$ and integrate from $\rho=0$ to $\rho=+\infty$, taking into account that the normalization condition (A1) implies that

$$f(-ip_n,E_n;\rho) = \frac{\partial}{\partial\rho}f(-ip_n,E_n;\rho)|_{\rho=0}\,\varphi(E_n;\rho). \quad \text{(C4)}$$

The relation so obtained, together with the definition of the bound-state parameters C_n (A2b), is just our thesis (C2).

As a direct consequence of the fact that the function $\tilde{g}(k,E)$ has no poles, we obtain the holomorphy of the functions $\tilde{g}_l(k)$ for $\mathrm{Im}\,k\leq 0$ that allowed us to write down the dispersion relations (2.24). The asymptotic

constants $\tilde{g}_l^{(n)}$, defined through the asymptotic behavior

$$\tilde{g}_l(k) = \sum_{n=0}^{N} \tilde{g}_l^{(n)}(-2ik)^{-n-1} + O(k^{-N-2}), \quad l = 1, 2, \quad \text{Im } k \leq 0, \quad \text{(C5)}$$

are all real as required by the parity relations (2.27); on the other hand, on the basis of the dispersion relations (2.24) we can anticipate some general features of the expressions for $g_l^{(n)}$. From (C5), we obtain for real k

$$\text{Re } \tilde{g}_l(k) = \sum_{n=0}^{N} \tilde{g}_l^{(2n+1)}(-2ik)^{-2n-2} + O(k^{-2N-4}), \quad \text{Im } k = 0, \quad \text{(C6a)}$$

$$\text{Im } \tilde{g}_l(k) = -i \sum_{n=0}^{N} \tilde{g}_l^{(2n)}(-2ik)^{-2n-1} + O(k^{-2N-3}), \quad \text{Im } k = 0. \quad \text{(C6b)}$$

Furthermore, Re $\tilde{g}_l(k)$ and Im $\tilde{g}_l(k)$ are known functionals of the scattering phase shift and bound-state parameters, whose expressions we now write explicitly, using Eqs. (2.24), (2.22), (2.23), and (2.8):

$$\text{Re } \tilde{g}_1(k) = \frac{2}{\pi} P \int_0^\infty dq \, \frac{q^2}{q^2 - k^2} [S_1(q) - 1],$$

$$\text{Re } \tilde{g}_2(k) = \frac{2}{\pi} P \int_0^\infty dq \, \frac{q^2}{q^2 - k^2} S_2(q), \quad \text{Im } k = 0, \quad \text{(C7)}$$

$$\text{Im } \tilde{g}_1(k) = k \Bigg\{ 1 - \prod_{n=1}^{N^+ + N^-} \left(\frac{k^2}{k^2 + p_n^2} \right) \times \exp \left[\frac{2}{\pi} P \int_0^\infty \frac{t \, dt}{t^2 - k^2} (\eta_+(t) + \eta_-(t)) \right] \times \cosh \Bigg[(N^+ - N^-) \ln \left(\frac{E - M}{E + M} \right) + \sum_{n=1}^{N^+ + N^-} \ln \left(\frac{E + E_n}{E - E_n} \right) + \frac{2}{\pi} EP \int_0^\infty \frac{t \, dt}{(t^2 - k^2)E(t)} (\eta_+(t) - \eta_-(t)) \Bigg] \Bigg\}, \quad \text{Im } k = 0, \quad \text{(C8a)}$$

$$\text{Im } \tilde{g}_2(k) = -\frac{k}{E} \Bigg\{ \prod_{n=1}^{N^+ + N^-} \left(\frac{k^2}{k^2 + p_n^2} \right) \times \exp \left[\frac{2}{\pi} P \int_0^\infty \frac{t \, dt}{t^2 - k^2} (\eta_+(t) + \eta_-(t)) \right] \times \sinh \Bigg[(N^+ - N^-) \ln \left(\frac{E - M}{E + M} \right) + \sum_{n=1}^{N^+ + N^-} \ln \left(\frac{E + E_n}{E - E_n} \right) + \frac{2}{\pi} EP \int_0^\infty \frac{t \, dt}{(t^2 - k^2)E(t)} (\eta_+(t) - \eta_-(t)) \Bigg] \Bigg\}, \quad \text{Im } k = 0. \quad \text{(C8b)}$$

From the relations (C7) and (C6a), we expect, for the odd coefficients $\tilde{g}_l^{(2n+1)}$, integral expressions involving the functions $S_l(k)$, $l = 1, 2$, and from (C8) and (C6b), for the even coefficients $\tilde{g}_l^{(2n)}$, integral expressions involving the functions $\eta_l(k)$, $l = 1, 2$, in addition to a contribution depending only on the energies of the bound states. Furthermore, Im $\tilde{g}_l(k) = \text{Im } g_l(k)$, $l = 1, 2$, so that the even coefficients $g_l^{(2n)}$, $l = 1, 2$, do not depend on the bound-state parameters C_n, while the odd coefficients $g_l^{(2n+1)}$ depend on the C_n in the following simple manner:

$$g_1^{(2n+1)} = \tilde{g}_1^{(2n+1)} + 2^{2n+1} \sum_{m=1}^{N^+ + N^-} C_m p_m^{2n},$$

$$g_2^{(2n+1)} = \tilde{g}_2^{(2n+1)} + 2^{2n+1} \sum_{m=1}^{N^+ + N^-} \frac{C_m}{E_m} p_m^{2n}, \quad \text{(C9)}$$

as can be easily derived from the definition (2.21).

We now proceed to apply Roberts method to get the expressions for the odd coefficients $\tilde{g}_l^{(2n+1)}$. We integrate the functions

$$F_l(k) = k^{2n+1} \left[\tilde{g}_l(k) - \sum_{p=0}^{2n+1} \tilde{g}_l^{(p)}(-2ik)^{-p-1} \right], \quad l = 1, 2, \quad \text{(C10)}$$

along a contour composed of a large semicircle in the lower half k plane and of the real axis indented at $k = 0$ and, using the holomorphy and the asymptotic expansions of the functions $\tilde{g}_l(k)$, we get

$$\tilde{g}_l^{(2n+1)} = \frac{(-1)^{n+1}}{\pi} 2^{2n+2} \int_{-\infty}^{+\infty} dk k^{2n+1} \left(\text{Im } \tilde{g}_l(k) - \sum_{m=0}^{n} (-1)^m \tilde{g}_l^{(2m)}(2k)^{-2m-1} \right). \quad \text{(C11)}$$

It is possible to write these equations in a more compact form, eliminating the subtraction terms in the integral by $(2n + 3)$ partial integrations; using the operator

$$D_k = k \frac{d}{dk} k, \quad \text{(C12)}$$

whose properties are discussed in Appendix D of Paper I, to which we refer for details, the expressions (C11) also read

$$\tilde{g}_l^{(2n+1)} = \frac{(-1)^{n+1} 2^{2n+3}}{\pi (2n+3)!} \int_0^\infty dk D_k^{2n+3} [k^{-1} S_l(k)], \quad l = 1, 2. \quad \text{(C13)}$$

Our final result for the odd coefficients is so obtained inserting (C13) in (C9)

$$g_l^{(2n+1)} = \frac{(-1)^{n+1} 2^{2n+3}}{\pi (2n+3)!} \int_0^\infty dk D_k^{2n+3} [k^{-1} S_l(k)] + 2^{2n+1} \sum_{m=1}^{N^+ + N^-} \left(\delta_{l1} + \frac{\delta_{l2}}{E_m} \right) C_m p_m^{2n}, \quad l = 1, 2, \quad \text{(C14)}$$

where δ_{li} is the Kronecker symbol.

Let us proceed now to discuss the even coefficients. From (C6b) and (2.22), we see that we have to study the asymptotic behavior of the functions $S_l(k)$, $l = 1, 2$, for large k:

$$S_l(k) = \delta_{l1} + 2\sum_{n=0}^{N} g_l^{(2n)}(-2ik)^{-2n-2} + O(k^{-2N-4}). \tag{C15}$$

However, the exponential dependence on the phase shifts of the function $S(k, E)$, implied by the definition (2.23) and the dispersion relations (2.8), forces us to derive the functional expression for the even coefficients $g_l^{(2n)}$ by two steps. First, we find the expression of the asymptotic coefficients of the function

$$H(k, E) = \frac{d}{dk}\ln(f(k, E)) = H_1(k) + EH_2(k), \qquad \operatorname{Im} k \leq 0, \tag{C16}$$

as functionals of the phase shifts and bound-state parameters, and then we give the relation between these coefficients and the coefficients $g_l^{(2n)}$. Our starting point is the meromorphy of the functions $H_1(k)$ and $H_2(k)$ for $\operatorname{Im} k \leq 0$, so that we apply the Roberts method to the functions

$$H_1(k) = \frac{1}{2}\frac{d}{dk}\ln[f(k, E)f(k, -E)],$$

$$E^2H_2(k) = \frac{E}{2}\frac{d}{dk}\ln\left[\frac{f(k, E)}{f(k, -E)}\right], \quad \operatorname{Im} k \leq 0. \tag{C17}$$

To this end, let us introduce the asymptotic expressions

$$H_1(k) = \sum_{n=0}^{N} H_1^{(n)}k^{-n-2} + O(k^{-N-3}),$$

$$H_2(k) = \sum_{n=0}^{N} H_2^{(n)}k^{-n-3} + O(k^{-N-4}), \tag{C18}$$

as implied by the asymptotic expansions (2.12) and by Eq. (C16), that can be written, together with the definition (B1),

$$H(k, E) = \int_{\infty}^{0}\frac{d}{dk}g_1(k, \rho)\,d\rho + E\int_{\infty}^{0}\left[\frac{d}{dk}g_2(k, \rho) + \frac{k}{k^2 + M^2}g_2(k, \rho)\right]d\rho. \tag{C19}$$

We may now proceed along the same lines as above for the functions $\tilde{g}_l(k)$; we obtain expressions for the asymptotic coefficients of the real part, for k real, of the functions $H_l(k)$ in terms of the generalized moments of the imaginary part, plus a contribution resulting from the poles lying in the half plane $\operatorname{Im} k \leq 0$, just as shown by the formula (C14). By integrating the function

$$k^{2n}\left[H_1(k) - \sum_{p=0}^{2n-1} H_1^{(p)}k^{-p-2}\right]$$

along the same contour as the function (C10), we obtain the expressions

$$H_1^{(2n+1)} = \frac{n+1}{\pi(2n+1)!}\int_{-\infty}^{+\infty} dk\, D_k^{2n+1}[\eta_+(k) + \eta_-(k)] + (-1)^{n+1}\sum_{m=1}^{N^+ + N^-} p_m^{2n+2}. \tag{C20}$$

Now let us perform the integration along the same contour of the function

$$k^{2n}\left(h(k) - \sum_{p=1}^{2n+1} h_p k^{-p}\right),$$

where $h(k)$ is the function

$$E^2H_2(k) = \frac{E}{2}\frac{d}{dk}\ln\left(\frac{f(k, E)}{f(k, -E)}\right)$$

which behaves asymptotically as implied by (C18):

$$h(k) = \sum_{n=1}^{N} h_n k^{-n} + O(k^{-N-1}). \tag{C21}$$

The odd asymptotic coefficients are then given by

$$h_{2n+1} = \frac{1}{\pi}\int_{-\infty}^{+\infty} dk k^{2n}\left(\frac{E}{2}\frac{d}{dk}[\eta_+(k) - \eta_-(k)] + i\sum_{p=1}^{n} h_{2p}k^{-2p}\right) + (-1)^n\sum_{m=1}^{N^+ + N^-} E_m p_m^{2n}. \tag{C22}$$

We can now write the integral appearing in (C22) in a more compact form integrating by parts and taking into account, in the case $n = 0$, the normalization (B13) of the phase shifts. In fact, in this case the relation (C22) is

$$h_1 = \frac{1}{\pi}\int_0^{\infty} dk\frac{d}{dk}[E(\eta_+(k) - \eta_-(k)) - 2bk] - \frac{1}{2\pi}\int_{-\infty}^{+\infty} dk\left(\frac{k}{E}[\eta_+(k) - \eta_-(k)] - 2b\right) + \sum_{m=1}^{N^+ + N^-} E_m, \tag{C23}$$

where we have used the behavior for large k implied by (B2):

$$\eta_2(k) \xrightarrow[k\to\infty]{} \frac{b}{k} + O(k^{-3}). \tag{C24}$$

The relation between the value of the phase shifts for $k = 0$ and the number of bound states allows us to give the following expression for h_1:

$$h_1 = \frac{1}{2\pi}\int_{-\infty}^{+\infty} dk\, D_k\left(\frac{\eta_+(k) - \eta_-(k)}{E}\right) + \sum_{m=1}^{N^+ + N^-} E_m - M(N^+ - N^-). \quad \text{(C25)}$$

Performing the same steps for the other coefficients, no contribution comes from the $k = 0$ limit of the phase shifts because of the factor k^{2n}, and our final formula reads

$$h_{2n+1} = \frac{2n+1}{2} I_n + nM^2 I_{n-1} + (-1)^n \sum_{m=1}^{N^+ + N^-} E_m p_m^{2n}, \quad n > 0, \quad \text{(C26)}$$

where

$$I_n = \frac{1}{\pi(2n+1)!}\int_{-\infty}^{+\infty} dk\, D_k^{2n+1}\left[\frac{\eta_+(k) - \eta_-(k)}{E}\right]. \quad \text{(C27)}$$

Now it remains to give the relations between the coefficients (C20) and (C26) and the coefficients $g_l^{(2n)}$, using Eqs. (C15) and (C16), where this last equation implies for real k

$$S(k, E)\ \mathrm{Re}\ H(k, E) = -\frac{1}{2}\frac{d}{dk} S(k, E). \quad \text{(C28)}$$

However, the validity of this relation on the two Riemann sheets can be expressed by the two coupled equations

$$\frac{d}{dk} S_1(k) = -2\ \mathrm{Re}\ H_1(k) S_1(k) - 2\ \mathrm{Re}\ h(k) S_2(k),$$

$$\frac{d}{dk} S_2(k) = -\frac{2}{k^2 + M^2}\ \mathrm{Re}\ h(k) S_1(k) - \left(\frac{k}{k^2 + M^2} + 2\ \mathrm{Re}\ H_1(k)\right) S_2(k). \quad \text{(C29)}$$

It is more convenient now to make use of matrix notation, introducing the vector $\mathbf{v}(k)$ and the matrix $\mathbf{A}(k)$,

$$\mathbf{v}(k) \equiv \begin{pmatrix} S_1(k) \\ S_2(k) \end{pmatrix},$$

$$\mathbf{A}(k) \equiv \frac{1}{4i}\begin{pmatrix} -2\ \mathrm{Re}\ H_1(k) & -2\ \mathrm{Re}\ h(k) \\ -\dfrac{2\ \mathrm{Re}\ h(k)}{k^2 + M^2} & -\dfrac{k}{k^2 + M^2} - 2\ \mathrm{Re}\ H_1(k) \end{pmatrix}, \quad \text{(C30)}$$

together with their asymptotic expansions

$$\mathbf{v}(k) = \sum_{n=0}^{N} \mathbf{v}_n(-2ik)^{-2n} + O(k^{-2N-2}), \quad \text{(C31)}$$

$$\mathbf{A}(k) = \sum_{n=0}^{N} \mathbf{A}_n(-2ik)^{-2n-1} + O(k^{-2N-3}). \quad \text{(C32)}$$

Equation (C29), which now reads

$$\frac{d}{dk}\mathbf{v}(k) = 4i\mathbf{A}(k)\mathbf{v}(k), \quad \text{(C33)}$$

implies the following recursion relation for the vectors $\mathbf{v}_n$:

$$n\mathbf{v}_n = \sum_{m=0}^{n} \mathbf{A}_{n-m}\mathbf{v}_m, \quad \text{(C34)}$$

which for $n = 0$ yield the condition

$$\mathbf{A}_0\mathbf{v}_0 = 0. \quad \text{(C35)}$$

To solve the recursion relation (C34), we introduce the matrices

$$\mathbf{B}_n = (n - \mathbf{A}_0)^{-1}, \quad n \geq 1, \quad \text{(C36)}$$

which are assumed to exist for each positive integer (this is proved below). The relations (C34) then become

$$\mathbf{v}_n = \mathbf{B}_n \sum_{m=0}^{n-1} \mathbf{A}_{n-m}\mathbf{v}_m, \quad n \geq 1. \quad \text{(C37)}$$

These recursion relations can be simply solved noting that the only possible expression for $\mathbf{v}_n$ obtained by iteration must be

$$\mathbf{v}_n = \mathbf{B}_n \sum_{l=1}^{n} \sum_{\sigma} \sum_{\tau} [n, l/\sigma_1 \cdots \sigma_l; \tau_1 \cdots \tau_{l-1}] \times \mathbf{A}_{\sigma_1}\mathbf{B}_{\tau_1}\mathbf{A}_{\sigma_2}\mathbf{B}_{\tau_2} \cdots \mathbf{A}_{\sigma_{l-1}}\mathbf{B}_{\tau_{l-1}}\mathbf{A}_{\sigma_l}\mathbf{v}_0, \quad \text{(C38)}$$

where the unknown numerical coefficients $[n, l/\sigma_1, \cdots, \sigma_l; \tau_1, \cdots, \tau_{l-1}]$ are for the moment only required to satisfy the dimensional condition

$$\sum_{j=1}^{l} \sigma_j = n. \quad \text{(C39)}$$

If we now put the expression (C38) into the recursion relation (C37), we find that the coefficients $[n, l/\sigma_1 \cdots \sigma_l; \tau_1 \cdots \tau_{l-1}]$ are different from zero only if

$$\tau_k = n - \sum_{j=1}^{k} \sigma_j, \quad k = 1, 2, \cdots, l-1; \quad \text{(C40)}$$

moreover, these coefficients must satisfy the following recursion relation:

$$[n, l/\sigma_1 \cdots \sigma_l; \tau_1 \cdots \tau_{l-1}] = [n - \sigma_1, l - 1/\sigma_2 \cdots \sigma_l; \tau_2 \cdots \tau_{l-1}], \quad l > 1, \tag{C41}$$

with the starting condition

$$[n, 1/n] = 1 \tag{C42}$$

so that we obtain the simple solution

$$[n, l/\sigma_1 \cdots \sigma_l; \tau_1 \cdots \tau_{l-1}] = 1. \tag{C43}$$

We are now in the position to solve our problem, taking into account that equations (C30), (C31), and (C15) imply

$$\mathbf{v}_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{v}_{n+1} = 2\begin{pmatrix} g_1^{(2n)} \\ g_2^{(2n)} \end{pmatrix}, \quad n \geq 0. \tag{C44}$$

Equations (C32) and (C30), together with the asymptotic conditions (C18) and (C21) and the expressions (C20), (C26), and (C27), completely determine the functional dependence of the matrices $\mathbf{A}_n$ from the phase shifts and the bound-state parameters

$$\mathbf{A}_n = (2i)^{2n}\begin{pmatrix} H_1^{(2n-1)} & h_{2n+1} \\ -\sum_{m=0}^{n-1}(iM)^{2n-2m-2}h_{2m+1} & \tfrac{1}{2}(iM)^{2n} + H_1^{(2n-1)} \end{pmatrix}, \quad n \geq 1. \tag{C45}$$

It remains to check the condition (C35) and to construct the matrices $\mathbf{B}_n$. The matrix $\mathbf{A}_0$ is given by

$$\mathbf{A}_0 = \begin{pmatrix} 0 & h_1 \\ 0 & \tfrac{1}{2} \end{pmatrix}, \tag{C46}$$

so that the condition (C35) is verified because of (C44). On the other hand, the nonvanishing eigenvalue of the matrix $\mathbf{A}_0$ is $\frac{1}{2}$, so that the matrix $\mathbf{B}_n$ exists for any positive integer n:

$$\mathbf{B}_n = \begin{pmatrix} \dfrac{1}{n} & \dfrac{2h_1}{n(2n-1)} \\ 0 & \dfrac{2}{2n-1} \end{pmatrix}, \quad n \geq 1. \tag{C47}$$

With the notation

$$\mathbf{x}_1 \equiv \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{x}_2 \equiv \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{C48}$$

the functional expression of the even coefficient $g_m^{(2n)}$, $m = 1, 2$, is finally given by the matrix element

$$g_m^{(2n)} = \tfrac{1}{2}\langle \mathbf{x}_m, \mathbf{v}_{n+1}\rangle = \tfrac{1}{2}\sum_{l=1}^{n+1}\sum_{\sigma}\langle \mathbf{x}_m, \mathbf{B}_{n+1}\mathbf{A}_{\sigma_1}\mathbf{B}_{\tau_1}\mathbf{A}_{\sigma_2}\cdots\mathbf{A}_{\sigma_{l-1}}\mathbf{B}_{\tau_{l-1}}\mathbf{A}_{\sigma_l}\mathbf{x}_1\rangle, \tag{C49}$$

where the sum $\sum_\sigma$ runs over all the possible combinations of l positive integers σ_j satisfying the condition

$$\sum_{j=1}^{l}\sigma_j = n + 1,$$

while the $(l-1)$ positive integers τ_k are equal to

$$n + 1 - \sum_{j=1}^{k}\sigma_j.$$

We can also state, just by looking at the recursion relation (C37), that the total sum $\sum_{l=1}^{n+1}\sum_\sigma$ is composed of 2^n terms. As an example, we write down the explicit dependence of the coefficients $g_m^{(2n)}$, for $n \leq 2$, on the phase shifts and bound-state parameters. First, it is convenient to introduce the quantities

$$J_1^{(2n)} = (-1)^{n+1}\frac{2^{2n+1}}{n}H_1^{(2n-1)},$$

$$J_2^{(2n+1)} = (-1)^{n+1}\frac{2^{2n+1}}{n}h_{2n+1}, \quad n \geq 1, \; J_2^{(1)} = -2h_1. \tag{C50}$$

If we now insert in (C49) the matrices

$$\mathbf{A}_n = -\frac{1}{2}\begin{pmatrix} nJ_1^{(2n)} & nJ_2^{(2n+1)} \\ 4(2M)^{2n-2}J_2^{(1)} + 4\sum_{m=1}^{n-1}m(2M)^{2n-2m-2}J_2^{(2m+1)} & nJ_1^{(2n)} - (2M)^{2n} \end{pmatrix}, \tag{C51}$$

we obtain

$$g_1^{(0)} = -\tfrac{1}{4}J_1^{(2)} + [J_2^{(1)}]^2, \tag{C52a}$$

$$g_2^{(0)} = -2J_2^{(1)}, \tag{C52b}$$

$$g_1^{(2)} = -\tfrac{1}{4}J_1^{(4)} + \tfrac{1}{16}[J_1^{(2)}]^2 - \tfrac{1}{2}J_1^{(2)}[J_2^{(1)}]^2 + \tfrac{2}{3}J_2^{(3)}J_2^{(1)} + \tfrac{4}{3}M^2[J_2^{(1)}]^2 + \tfrac{1}{3}[J_2^{(1)}]^4, \tag{C52c}$$

$$g_2^{(2)} = -\tfrac{4}{3}J_2^{(3)} - \tfrac{8}{3}[J_2^{(1)}]^3 - \tfrac{32}{3}M^2J_2^{(1)} + 2J_2^{(1)}J_1^{(2)}, \tag{C52d}$$

$$\begin{aligned} g_1^{(4)} = {} & -\tfrac{1}{4}J_1^{(6)} + \tfrac{1}{8}J_1^{(4)}J_1^{(2)} - \tfrac{1}{96}[J_1^{(2)}]^3 - \tfrac{1}{2}J_1^{(4)}[J_2^{(1)}]^2 - \tfrac{1}{8}J_1^{(2)}J_2^{(1)}J_2^{(3)} + \tfrac{4}{5}J_2^{(5)}J_2^{(1)} + \tfrac{1}{9}[J_2^{(3)}]^2 \\ & + \tfrac{4}{9}J_2^{(3)}[J_2^{(1)}]^3 + \tfrac{2}{45}[J_2^{(1)}]^6 + \tfrac{56}{45}M^2J_2^{(1)}J_2^{(3)} - \tfrac{2}{3}M^2J_1^{(2)}[J_2^{(1)}]^2 \\ & + \tfrac{8}{9}M^2[J_2^{(1)}]^4 + \tfrac{128}{45}M^4[J_2^{(1)}]^2 - \tfrac{1}{6}J_1^{(2)}[J_2^{(1)}]^4, \end{aligned} \tag{C52e}$$

$$\begin{aligned} g_2^{(4)} = {} & -\tfrac{4}{5}J_2^{(5)} + J_1^{(4)}J_2^{(1)} + \tfrac{1}{8}J_2^{(3)}J_1^{(2)} - \tfrac{4}{3}J_2^{(3)}[J_2^{(1)}]^2 - \tfrac{1}{4}J_2^{(1)}[J_1^{(2)}]^2 \\ & + \tfrac{2}{3}J_1^{(2)}[J_2^{(1)}]^3 - \tfrac{4}{15}[J_2^{(1)}]^5 - \tfrac{32}{15}M^2J_2^{(3)} + \tfrac{8}{3}M^2J_1^{(2)}J_2^{(1)} - \tfrac{16}{3}[J_2^{(1)}]^3 - \tfrac{256}{15}M^4J_2^{(1)}. \end{aligned} \tag{C52f}$$

APPENDIX D

In this appendix we discuss the properties of the functions $F_l^{(p)}$, $l = 1, 2$, $p = 0, 1, \cdots$, defined by Eq. (2.15): namely, the structure of the recursion relations (2.13) for the functions $g_l^{(n)}(\rho)$, $l = 1, 2$, $n = 0, 1, 2, \cdots$. We start writing down the recursion relations (2.13) in a slightly different form with the substitution $\epsilon = M^2$:

$$\begin{aligned} \frac{d}{d\rho}g_1^{(n)}(\rho) = {} & -g_1^{(n+1)}(\rho) - \sum_{m=0}^{n-1} g_1^{(n-1-m)}(\rho)g_1^{(m)}(\rho) \\ & - \epsilon\sum_{m=0}^{n-1} g_2^{(n-1-m)}(\rho)g_2^{(m)}(\rho) \\ & + \tfrac{1}{4}\sum_{m=0}^{n+1} g_2^{(n+1-m)}(\rho)g_2^{(m)}(\rho), \quad n \geq 0, \end{aligned} \tag{D1a}$$

$$\frac{d}{d\rho}g_2^{(n)}(\rho) = -g_2^{(n+1)}(\rho) - 2\sum_{m=0}^{n-1} g_1^{(n-1-m)}(\rho)g_2^{(m)}(\rho), \quad n \geq 0. \tag{D1b}$$

Here and in the following, a sum is understood to vanish if the upper value of the summation index is smaller than the lower value. The first step is to obtain all the derivatives of the two functions $g_1^{(0)}(\rho)$ and $g_2^{(0)}(\rho)$ in terms of the functions $g_l^{(n)}(\rho)$. It is easy to prove by induction that the nth derivative of the functions $g_l^{(0)}(\rho)$ has an expression of the following kind (we omit the ρ dependence):

$$\begin{aligned} \frac{d^n}{d\rho^n}g_1^{(0)} = {} & (-1)^n g_1^{(n)} + A_1^{(n)}[g_1^{(0)}, g_1^{(1)}, \cdots, g_1^{(n-2)}; \\ & g_2^{(0)}, g_2^{(1)}, \cdots, g_2^{(n)}; \epsilon], \quad n \geq 1, \end{aligned} \tag{D2a}$$

$$\begin{aligned} \frac{d^n}{d\rho^n}g_2^{(0)} = {} & (-1)^n g_2^{(n)} + A_2^{(n)}[g_1^{(0)}, g_1^{(1)}, \cdots, g_1^{(n-2)}; \\ & g_2^{(0)}, g_2^{(1)}, \cdots, g_2^{(n-2)}; \epsilon], \quad n \geq 1. \end{aligned} \tag{D2b}$$

The functions $A_l^{(n)}$ are sums of products of the functions $g_l^{(n)}(\rho)$ and of the nonnegative integer powers of ϵ in such a way that each term of this sum has the correct dimensions. However, not all the possible combinations of the functions $g_l^{(n)}(\rho)$ and the powers ϵ^m appear as terms of the sums $A_l^{(n)}$. In fact, the proliferation of the terms, as the order of the derivative of the $g_l^{(0)}(\rho)$'s increases, must satisfy certain rules because of the expressions (D1) of the first derivative of the functions $g_l^{(n)}(\rho)$. The following is proved immediately by induction:

(i) Each term of the sum $A_1^{(n)}[A_2^{(n)}]$ contains an even or null [odd] number of functions $g_2^{(l)}(\rho)$;

(ii) the powers ϵ^m enter only in those terms of the sum $A_l^{(n)}$ containing at least m pairs of functions $g_2^{(l)}(\rho)$;

(iii) the sum $A_2^{(n)}$ does not contain any term that is a product of functions $g_1^{(n)}(\rho)$ only;

(iv) the sum $A_l^{(2p)}$ does not contain the terms $[g_1^{(0)}(\rho)]^{2p+1+\delta_{l1}}$.

Now we introduce a compact notation for the general term entering in the sum $A_l^{(n)}$, showing explicitly the dimensions

$$P_q[n, l; s_1 \cdots s_l] \equiv \prod_{j=1}^{l} g_q^{(s_j)}(\rho),$$

$$q = 1, 2, \quad n \geq 0, \quad 1 \leq l \leq l_{\max}, \tag{D3}$$

where the index n means that the following conditions on the nonnegative integers s_j hold:

$$\begin{aligned} & \sum_{j=1}^{l} s_j + 2l = n + 2, \quad \text{if } q = 1, \\ & \sum_{j=1}^{l} s_j + l = n + 1, \quad \text{if } q = 2, \end{aligned} \tag{D4}$$

while the maximum value of l for fixed n is given by

$$\begin{aligned} & l_{\max} = \frac{n}{2} + 1, \quad \text{for even } n, \\ & l_{\max} = \frac{n+1}{2}, \quad \text{for odd } n, \quad q = 1, \end{aligned} \tag{D5a}$$

$$l_{\max} = n + 1, \quad q = 2. \tag{D5b}$$

The expressions of the sums $A_i^{(n)}$, therefore, read

$$A_1^{(n)}[g_1^{(0)}, \cdots, g_1^{(n-2)}; g_2^{(0)}, \cdots, g_2^{(n)}; \epsilon]$$
$$= \sum B_1(n, l; s_1 \cdots s_l) P_1[n, l; s_1 \cdots s_l]$$
$$+ \sum B_2(n, m, l; s_1 \cdots s_{2(l+m+\delta_{m0}-1)}) \epsilon^m P_2[n + 1 - 2m, 2(l + m + \delta_{m0} - 1); s_1 \cdots s_{2(l+m+\delta_{m0}-1)}]$$
$$+ \sum B_{12}(n, m, p, l_1, l_2; s_1 \cdots s_{l_1}; t_1 \cdots t_{2(l_2+m+\delta_{m0}-1)}) \epsilon^m P_1[p, l_1; s_1 \cdots s_{l_1}]$$
$$\times P_2[n - 2m - p - 1, 2(l_2 + m + \delta_{m0} - 1); t_1 \cdots t_{2(l_2+m+\delta_{m0}-1)}], \tag{D6}$$

$$A_2^{(n)}[g_1^{(0)}, \cdots, g_1^{(n-2)}; g_2^{(0)}, \cdots, g_2^{(n-2)}; \epsilon]$$
$$= \sum C_2(n, m, l; s_1 \cdots s_{2(l+m)-1}) \epsilon^m P_2[n - 2m, 2(l + m) - 1; s_1 \cdots s_{2(l+m)-1}]$$
$$+ \sum C_{12}(n, m, p, l_1, l_2; s_1 \cdots s_{l_1}; t_1 \cdots t_{2(l+m)-1}) \epsilon^m P_1[p, l_1; s_1 \cdots s_{l_1}]$$
$$\times P_2[n - 2m - p - 2, 2(l_2 + m) - 1, t_1 \cdots t_{2(l_2+m)-1}]. \tag{D7}$$

All the coefficients B and C multiplying the products P_q [Eq. (D3)] are numerical ρ-independent quantities which do not depend on the starting functions $g_1^{(0)}(\rho)$ and $g_2^{(0)}(\rho)$, that is, on the potentials, and could be computed once and for all. The first sum of the expression (D6), i.e., that containing the coefficients B_1, runs over the values of the indices $l, s_1, \cdots, s_l$, satisfying the conditions (D4) and (D5a) [except the value $l = 1$ that corresponds to the explicitly written term in Eq. (D2a)]. The same holds for the second sum containing the coefficients B_2, for any fixed value of m, while the summation over m goes from $m = 0$ to a maximum value $m_{\max}$ fixed by the inequality

$$4m_{\max} - 2 \leq n < 4m_{\max} + 2. \tag{D8}$$

The third sum containing the coefficients B_{12} runs over the indices $l_1, l_2, s_1, \cdots, s_{l_1}, t_1, \cdots, t_{2(l+m+\delta_{m0}-1)}$, for fixed values of m and p, following the same rules

TABLE II. The coefficients B_2 and B_{12} of Eq. (D6) and C_2 and C_{12} of Eq. (D7) for n up to 5. The coefficients $B_1(n, l; s_1 \cdots s_l)$ are not printed here because they are equal to the coefficients $\{n, l; s_1 \cdots s_l\}$ given in Table II of I. All these coefficients have been checked by hand. We note that by definition [see Eqs. (D3), (D6), and (D7)] all the coefficients are completely symmetric in the indices s_i and t_i separately.

$B_2(1, 0, 1; 0\,1) = \frac{1}{2}$,	$B_2(2, 0, 1; 0\,2) = -1$,	$B_2(2, 0, 1; 1\,1) = -\frac{3}{4}$,	$B_2(2, 1, 1; 0\,0) = 1$
$B_{12}(2, 0, 0, 1, 1; 0; 0\,0) = -1$,	$B_2(3, 0, 1; 0\,3) = \frac{3}{2}$,	$B_2(3, 0, 1; 1\,2) = 3$	
$B_2(3, 0, 2; 0\,0\,0\,1) = -\frac{1}{2}$,	$B_2(3, 1, 1; 0\,1) = -4$,	$B_{12}(3, 0, 0, 1, 1; 0; 0\,1) = 8$	
$B_{12}(3, 0, 1, 1, 1; 1; 0\,0) = 3$,	$B_2(4, 0, 1; 0\,4) = -2$,	$B_2(4, 0, 1; 1\,3) = -5$	
$B_2(4, 0, 1; 2\,2) = -\frac{13}{4}$,	$B_2(4, 0, 2; 0\,0\,0\,2) = 2$,	$B_2(4, 0, 2; 0\,0\,1\,1) = \frac{25}{4}$	
$B_2(4, 1, 1; 0\,2) = 6$,	$B_2(4, 1, 1; 1\,1) = 5$,	$B_2(4, 1, 2; 0\,0\,0\,0) = -3$	
$B_{12}(4, 0, 0, 1, 1; 0; 0\,2) = -19$,	$B_{12}(4, 0, 0, 1, 1; 0; 1\,1) = -15$		
$B_{12}(4, 0, 0, 1, 2; 0; 0\,0\,0\,0) = 1$,	$B_{12}(4, 0, 1, 1, 1; 1; 0\,1) = -25$		
$B_{12}(4, 1, 0, 1, 1; 0; 0\,0) = 12$,	$B_{12}(4, 0, 2, 1, 1; 2; 0\,0) = -6$		
$B_{12}(4, 0, 2, 2, 1; 0\,0; 0\,0) = -19$,	$B_2(5, 0, 1; 0\,5) = \frac{5}{2}$,	$B_2(5, 0, 1; 1\,4) = \frac{15}{2}$	
$B_2(5, 0, 1; 2\,3) = 12$,	$B_2(5, 0, 2; 0\,0\,0\,3) = -5$,	$B_2(5, 0, 2; 0\,0\,1\,2) = -\frac{87}{2}$	
$B_2(5, 0, 2; 0\,1\,1\,1) = -\frac{105}{4}$,	$B_2(5, 0, 3; 0\,0\,0\,0\,0\,1) = \frac{1}{2}$,	$B_2(5, 1, 1; 0\,3) = -8$	
$B_2(5, 1, 1; 1\,2) = -18$,	$B_2(5, 1, 2; 0\,0\,0\,1) = 55$,	$B_{12}(5, 0, 0, 1, 1; 0; 0\,3) = 36$	
$B_{12}(5, 0, 0, 1, 1; 0; 12) = 75$,	$B_{12}(5, 0, 0, 1, 2; 0; 0\,0\,0\,1) = -52$		
$B_{12}(5, 0, 1, 1, 1; 1; 0\,2) = 66$,	$B_{12}(5, 0, 1, 1, 1; 1; 1\,1) = \frac{105}{2}$		
$B_{12}(5, 0, 1, 1, 2; 1; 0\,0\,0\,0) = -5$,	$B_{12}(5, 0, 2, 1, 1; 2; 0\,1) = 54$		
$B_{12}(5, 0, 2, 2, 1; 0\,0; 0\,1) = 167$,	$B_{12}(5, 0, 3, 1, 1; 3; 0\,0) = 10$		
$B_{12}(5, 0, 3, 2, 1; 0\,1; 0\,0) = 138$,	$B_{12}(5, 1, 0, 1, 1; 0; 0\,1) = -68$		
$B_{12}(5, 1, 1, 1, 1; 1; 0\,0) = -34$			
$C_{12}(2, 0, 0, 1, 1; 0; 0) = 2$,	$C_2(3, 0, 2; 0\,0\,1) = 1$,	$C_{12}(3, 0, 0, 1, 1; 0; 1) = -4$	
$C_{12}(3, 0, 1, 1, 1; 1; 0) = -4$,	$C_2(4, 0, 2; 0\,0\,2) = -3$,	$C_2(4, 0, 2; 0\,1\,1) = -5$	
$C_2(4, 1, 1; 0\,0\,0) = 4$,	$C_{12}(4, 0, 0, 1, 1; 0; 2) = 6$		
$C_{12}(4, 0, 0, 1, 2; 0; 0\,0\,0) = -2$,	$C_{12}(4, 0, 1, 1, 1; 1; 1) = 10$		
$C_{12}(4, 0, 2, 1, 1; 2; 0) = 6$,	$C_{12}(4, 0, 2, 2, 1; 0\,0; 0) = 12$		
$C_2(5, 0, 2; 0\,0\,3) = 6$,	$C_2(5, 0, 2; 0\,1\,2) = 27$,	$C_2(5, 0, 2; 1\,1\,1) = \frac{15}{2}$	
$C_2(5, 0, 3; 0\,0\,0\,0\,1) = -1$,	$C_2(5, 1, 2; 0\,0\,1) = -34$		
$C_{12}(5, 0, 0. 1, 1; 0; 3) = -8$,	$C_{12}(5, 0, 0, 1, 2; 0; 0\,0\,1) = 44$		
$C_{12}(5, 0, 1, 1, 1; 1; 2) = -18$,	$C_{12}(5, 0, 1, 1, 2; 1; 0\,0\,0) = 8$		
$C_{12}(5, 0, 2, 1, 1; 2; 1) = -18$,	$C_{12}(5, 0, 2, 2, 1; 0\,0; 1) = -34$		
$C_{12}(5, 0, 3, 1, 1; 3; 0) = -8$,	$C_{12}(5, 0, 3, 2, 1; 0\,1; 0) = -68$		

(D4) and (D5), and, for every fixed value of m between $m = 0$ and $m = m_{\max}$, p goes from $p = 0$ to $p = n - 2m - 2$, where $m_{\max}$ in this case is fixed by the following inequality:

$$4m_{\max} \leq n < 4m_{\max} + 4. \tag{D9}$$

On the other hand, the first sum in the expression (D7), namely that containing the coefficients C_2, runs over the indices l, $s_1, \cdots, s_{2(l+m)-1}$ with the usual conditions (D4) and (D5b) [except the value $l = 1$ that corresponds to the explicitly written term in Eq. (D2b)], for any value of m between $m = 0$ and $m = m_{\max}$, where $m_{\max}$ is given by the condition (D9); the same holds for the indices l_1, l_2, $s_1, \cdots, s_{l_1}$, $t_1, \cdots, t_{2(l+m)-1}$ in the second sum for any fixed value of m and p, while, for m between $m = 0$ and $m = m_{\max}$, p goes from $p = 0$ to $p = n - 2m - 2$, where $m_{\max}$ is given by the inequality

$$4m_{\max} + 2 \leq n < 4m_{\max} + 6. \tag{D10}$$

We did not try to find recursion relations for the coefficients B and C, as we did in the simpler nonrelativistic case (see I, Appendix F), but we limit ourselves to the demonstration of the general structure of the solution, with the aim of giving a useful basis for performing an algebraic computation of the coefficients B and C in FORMAC language by computer. We have obtained all these coefficients by computer for n up to 5 (see Table II). However, we note that the coefficients $B_1(n, l; s_1 \cdots s_l)$ are known from Paper I. In fact, making the choice $g_2^{(0)}(\rho) = 0$, we find that all the $g_2^{(n)}(\rho)$ vanish and the recursion relation (D1a) reduces to the nonrelativistic relation. In the notation of Paper I, we have

$$B_1(n, l; s_1 \cdots s_l) = \{n, l; s_1 \cdots s_l\}$$

(see Table II of I).

We finally obtain the expressions of the functions $F_l^{(n)}$ defined by Eqs. (2.15), using the relationships between the functions $g_l^{(0)}(\rho)$ and the potentials (2.14) and performing the nth derivative of the potentials

$$\frac{d^n}{d\rho^n} V_1(\rho) = \frac{1}{4} \sum_{i=0}^{n} \binom{n}{i} \frac{d^{n-i}}{d\rho^{n-i}} g_2^{(0)}(\rho) \frac{d^i}{d\rho^i} g_2^{(0)}(\rho) - \frac{d^n}{d\rho^n} g_1^{(0)}(\rho), \tag{D11a}$$

$$\frac{d^n}{d\rho^n} V_1(\rho) = -\frac{d^n}{d\rho^n} g_2^{(0)}(\rho). \tag{D11b}$$

We get the final result by substituting in the above expression the derivatives of the function $g_l^{(0)}(\rho)$, as given by Eq. (D2), and setting $\epsilon = M^2$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Orthogonality of a Set of Polynomials Encountered in Neutron-Transport and Radiative-Transfer Theories*

ERDAL INÖNÜ†
Center for Theoretical Studies, University of Miami, Coral Gables, Florida

(Received 13 January 1969)

The properties of the polynomials $h_l(\nu)$ which appear in the spherical-harmonics expansion of the eigensolution $\phi_\nu(\mu) = \sum_{l=0}^{\infty} \frac{1}{2}(2l+1)P_l(\mu)h_l(\nu)$ in plane-symmetric one-speed transport problems with anisotropic scattering are reviewed and further investigated. These polynomials are shown to be orthogonal in the Stieltjes sense with a weight distribution which contains a continuous as well as a discrete portion. Some further properties of the $h_l(\nu)$ are listed, taken from the Tchebycheff theory of orthogonal polynomials.

1. INTRODUCTION

In the treatment of the case of anisotropic scattering in neutron-transport as well as radiative-transfer theories, it is customary to express the solution as an expansion in spherical harmonics. One is then led to a special set of polynomials which appear as the expansion coefficients. These polynomials contain the Legendre moments of the scattering function as parameters and reduce to the Legendre polynomials if there is no scattering. They have been used by all the authors (among others, Chandrasekhar,[1] Kuščer,[2] Mika,[3] Lathrop,[4] Mulliken,[5] Inönü and Usseli,[6] McCormick and Kuščer,[7] Boffi and Trombetti[8]) who have considered linear or more general anisotropic scattering. In addition, they appear also in the spherical harmonics treatment of the isotropic scattering (as may be seen, e.g., in the writings of Mark,[9] Davison,[10] and Kofink[11]) as a linear combination of two kinds of Legendre polynomials. In spite of this wide use, however, the orthogonality properties of these polynomials and some of the immediate consequences which follow do not seem to have been pointed out in the literature. This is the aim of the present paper.

In the first part of the article we describe briefly the polynomials in question and obtain the orthogonality relation, thus showing that they fall within the general class of orthogonal polynomials. In the second part, we present some of their properties which may be of interest for neutron-transport as well as radiation-transfer problems. These properties follow directly from the general Tchebycheff theory and are listed here either without proof or with only a sketch of a proof. We note further that although these polynomials have sometimes been defined in the literature[5,7] for the more general case of azimuthal dependence, our considerations here are restricted, for simplicity, to the azimuth-independent case. Throughout the article, the notation of neutron-transport theory is used.

2. POLYNOMIALS $h_l(\nu)$ AND THEIR ORTHOGONALITY RELATION

A. Description of the Polynomials $h_l(\nu)$

We consider the polynomials $h_l(\nu)$ defined by the following recurrence relation:

$$(l+1)h_{l+1}(\nu) + lh_{l-1}(\nu) - (2l+1)(1-cf_l)\nu h_l(\nu) = 0, \quad \text{for} \quad l \geq 0, \quad (1)$$

and the initial condition

$$h_0 = 1. \quad (2)$$

Here c and f_l are real parameters which satisfy the conditions

$$0 < c < 1, \quad (3)$$

$$f_0 = 1, \quad (4)$$

$$|f_l| < 1, \quad \text{for} \quad l \geq 1, \quad (5)$$

* This research is supported by the National Science Foundation, NSF GU 2151 Sub 39.

† On leave from the Middle East Technical University, Ankara, Turkey.

[1] S. Chandrasekhar, *Radiative Transfer* (Oxford University Press, London, 1950).
[2] I. Kuščer, J. Math. & Phys. **34**, 256 (1956).
[3] J. R. Mika, Nucl. Sci. Eng. **11**, 415 (1961).
[4] K. D. Lathrop, LADC-5903, Los Alamos preprint, 1963.
[5] T. W. Mullikin, Astrophys. J. **139**, 379 (1964).
[6] E. Inönü and A. I. Usseli, Nucl. Sci. Eng. **23**, 251 (1965).
[7] N. J. McCormick and I. Kuščer, J. Math. Phys. **7**, 2036 (1966).
[8] V. C. Boffi and T. Trombetti, Nuovo Cimento **47B**, (1967).
[9] J. C. Mark, National Research Council, Canada, Publication No. 1561 (MT-26), 1943.
[10] B. Davison, *Neutron Transport Theory* (Oxford University Press, London, 1957).
[11] W. Kofink, Nuovo Cimento Suppl. **9**, 487 (1958).

and

$$\sum_{l=0}^{\infty}(2l+1)\,|f_l| < \infty. \tag{6}$$

Clearly, $h_l(\nu)$ is a polynomial of degree l in ν which contains only even or odd powers of ν as l is even or odd. The first few polynomials are

$$\begin{aligned} h_1(\nu) &= (1-c)\nu, \\ h_2(\nu) &= \tfrac{1}{2}[3(1-c)(1-cf_1)\nu^2 - 1], \\ h_3(\nu) &= (1/3!)\{3\times 5(1-c)(1-cf_1)(1-cf_2)\nu^3 \\ &\quad - [4(1-c) + 5(1-cf_2)]\nu\}. \end{aligned} \tag{7}$$

We have the symmetry property

$$h_l(-\nu) = (-1)^l h_l(\nu). \tag{8}$$

Note that, for $c = 0$, the $h_l(\nu)$ reduce to the Legendre polynomials $P_l(\nu)$. We also see from relation (1) that

$$h_l(0) = P_l(0). \tag{9}$$

The lth polynomial can be expressed as the following determinant:

$$h_l(\nu) = \frac{1}{l!}\begin{vmatrix} \xi_0\nu & 1 & 0 & 0 & \cdots & \\ 1 & \xi_1\nu & 2 & 0 & \cdots & \\ 0 & 2 & \xi_2\nu & 3 & \cdots & \\ 0 & 0 & 3 & \xi_3\nu & \cdots & \\ \cdot & \cdot & \cdot & \cdot & \cdots & \\ \cdot & \cdot & \cdot & \cdot & \cdots & l-1 \\ \cdot & \cdot & \cdot & \cdot & l-1 & \xi_{l-1}\nu \end{vmatrix}, \quad \text{for } l \geq 1, \tag{10}$$

where the abbreviation

$$\xi_l = (2l+1)(1-cf_l) \tag{11}$$

is used. Expanding the determinant in (10) with respect to the last row or column, one easily sees that it satisfies the recurrence relation (1).

The coefficient of the highest power in $h_l(\nu)$ may be written down from (10) as

$$a_{l,l} = \frac{1}{l!}\left(\prod_{i=0}^{l-1}\xi_i\right) = \frac{(2l-1)!!}{l!}\prod_{i=0}^{l-1}(1-cf_i), \tag{12}$$

while the other terms can be obtained from the following formula [Eqs. (2) and (6)]:

$$\begin{aligned} h_l(\nu) = \frac{1}{l!}\left(\prod_{i=0}^{l-1}\xi_i\right) & \\ \times \nu^l - \nu^{l-2}\sum_{j_1=0}^{l-2} w_{j_1}^2 + \nu^{l-4}\sum_{j_1=0}^{l-4}\sum_{j_2=j_1+2}^{l-2} w_{j_1}^2 w_{j_2}^2 + \cdots + (-1)^m\nu^{l-2m}\sum_{j_1=0}^{l-2m}\sum_{j_2=j_1+2}^{l-2m+2}\cdots\sum_{j_m=j_{m-1}+2}^{l-2} w_{j_1}^2 w_{j_2}^2\cdots w_{j_m}^2 & \\ + \cdots + \begin{cases} (-1)^{\frac{1}{2}l} w_0^2 w_2^2\cdots w_{l-2}^2, & \text{if } l \text{ is even}, \\ (-1)^{\frac{1}{2}(l-1)}\nu\displaystyle\sum_{j_1=0}^{1}\sum_{j_2=j_1+2}^{3}\cdots\sum_{\substack{j_n=j_n-1 \\ n=\frac{1}{2}(l-1)}}^{l-2} w_{j_1}^2 w_{j_2}^2\cdots w_{j_n}^2, & \text{if } l \text{ is odd}, \end{cases} \end{aligned} \tag{13}$$

where

$$w_j^2 = (j+1)^2/\xi_j\xi_{j+1}. \tag{14}$$

B. Connection with Neutron-Transport Theory

The polynomials $h_l(\nu)$ occur in neutron-transport theory in the following way. We consider the one-speed, stationary, plane-symmetric, homogeneous Boltzmann equation for neutron transport in an infinite medium with an azimuth-independent scattering law expressed as an infinite series of Legendre polynomials. We have then, using the notation of Case and Zweifel,[12]

$$\mu\frac{\partial\psi(x,\mu)}{\partial x} + \psi(x,\mu) = \tfrac{1}{2}c\sum_{l=0}^{\infty}(2l+1)f_lP_l(\mu)\int_{-1}^{+1}\psi(x,\mu')P_l(\mu')\,d\mu', \tag{15}$$

where x is the distance measured in mean free paths, μ is the cosine of the angle between the x axis and the neutron velocity, $\psi(x,\mu)$ is the neutron distribution in (x,μ), the f_l are the Legendre coefficients for a scattering law [given as the probability distribution

$$f(\boldsymbol{\Omega}\cdot\boldsymbol{\Omega}') = \sum_{l=0}^{\infty}\frac{2l+1}{4\pi}f_lP_l(\boldsymbol{\Omega}\cdot\boldsymbol{\Omega}') \tag{16}$$

for scattering from the direction $\boldsymbol{\Omega}$ into $\boldsymbol{\Omega}'$ with normalization $f_0 = 1$], and c is the ratio of the cross section for scattering to that for scattering plus absorption. We assume no multiplication, so that $c < 1$.

Making the substitution $\psi(x,\mu) = \phi_\nu(\mu)e^{-x/\nu}$ and defining

$$\phi_{\nu l} = \int_{-1}^{+1}\phi_\nu(\mu)P_l(\mu)\,d\mu, \tag{17}$$

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

we obtain the well-known eigenvalue equation for $\phi_\nu(\mu)$:

$$(\nu - \mu)\phi_\nu(\mu) = \tfrac{1}{2}c\nu \sum_{l=0}^{\infty}(2l+1)f_l P_l(\mu)\phi_{\nu l}. \quad (18)$$

Multiplying Eq. (18) with $P_k(\mu)$, integrating over μ from -1 to $+1$, and using the recurrence relation for the Legendre polynomials, we obtain a recurrence relation for the functions $\phi_{\nu l}$ which is identical to the relation (1). With the normalization $\phi_{\nu 0} = 1$, these functions become identical with the polynomials $h_l(\nu)$.

Thus the polynomials $h_l(\nu)$ are the coefficients in the Legendre expansion of the solution $\phi_\nu(\mu)$ of the eigenvalue equation (18):

$$\phi_\nu(\mu) = \sum_{l=0}^{\infty}\tfrac{1}{2}(2l+1)P_l(\mu)h_l(\nu). \quad (19)$$

If the scattering law is expressed as a finite sum containing Legendre polynomials up to $P_N(\mu)$, one has $f_l = 0$ for $l \geq N+1$ and the relation (1) becomes identical with the Legendre recurrence relation for $l \geq N+1$. In particular, for isotropic scattering, we have $f_l = 0$ for $l \geq 1$ and $h_l(\nu)$ can be expressed as a linear combination of the two sets of polynomials $P_l(\nu)$ and $W_{l-1}(\nu)$ which satisfy the Legendre recurrence relation. We have, in this case,

$$h_l(\nu) = P_l(\nu) - c\nu W_{l-1}(\nu), \quad (20)$$

where

$$W_{l-1}(\nu) = P_l(\nu)Q_0(\nu) - Q_l(\nu) = \tfrac{1}{2}\int_{-1}^{+1}\frac{P_l(\nu) - P_l(\mu)}{\nu - \mu}\,d\mu. \quad (21)$$

Thus, here the $h_l(\nu)$ reduce to the polynomials $(-1)^l G_l(\nu)$ used by Davison[10] and $G_l(\nu)$ used by Kofink.[11]

For a finite anisotropy of order N, it is still convenient to express $h_l(\nu)$ for $l \geq N$ as a linear combination of $P_l(\nu)$ and $W_{l-1}(\nu)$. One can write, following Kofink,[11]

$$h_l(\nu) = A_N(\nu)P_l(\nu) - c\nu B_N(\nu)W_{l-1}(\nu), \quad \text{for} \quad l \geq N, \quad (22)$$

where $A_N(\nu)$ and $B_N(\nu)$ are polynomials in ν^2 of degree N which are determined from the requirement that the expression represent $h_N(\nu)$ and $h_{N+1}(\nu)$ identically. One therefore finds,

$$A_N(\nu) = (N+1)[h_N(\nu)W_N(\nu) - h_{N+1}(\nu)W_{N-1}(\nu)],$$

$$B_N(\nu) = \frac{N+1}{c\nu}[h_N(\nu)P_{N+1}(\nu) - h_{N+1}(\nu)P_N(\nu)]. \quad (23)$$

To establish the orthogonality of the $h_l(\nu)$, we need to recall the essential properties of the eigensolutions $\phi_\nu(\mu)$. As is well known, the eigenvalue spectrum contains a continuous portion extending over the interval $(-1, +1)$, while the discrete eigenvalues are given by the roots of the equation

$$\Lambda(\nu) \equiv 1 - \tfrac{1}{2}c\nu\int_{-1}^{+1}\frac{M(\mu,\nu)}{\nu-\mu}\,d\mu = 1 - \tfrac{1}{2}c\nu\int_{-1}^{+1}\frac{M(\mu,\mu)}{\nu-\mu}\,d\mu = 0, \quad (24)$$

where

$$M(\mu,\nu) = \sum_{l=0}^{\infty}(2l+1)f_l P_l(\mu)h_l(\nu). \quad (25)$$

Both $P_l(\mu)$ and $h_l(\mu)$ are bounded in $-1 \leq \mu \leq +1$; it is clear that the series for $M(\mu,\nu)$ and $M(\mu,\mu)$ will be uniformly convergent if the condition (6) is satisfied.

Since $\Lambda(\nu) = \Lambda(-\nu)$, the roots occur in pairs $\mp\nu_j$. Kuščer[7] has shown that, for $c < 1$, all the roots are real. We shall assume that the roots all lie outside the interval $(-1, +1)$. This means that we are excluding the exceptional case of $\lambda(\nu)$ [which is defined in (28)] and $M(\nu,\nu)$ having common roots in the interval $(-1, +1)$. The total number of roots will depend on the values of c and f_l; it will remain finite even when the f_l form an infinite sequence (provided the convergence conditions are satisfied), as we show in Sec. 3F. We shall enumerate them in decreasing order of magnitude, ν_1 being the largest.

The solutions belonging to the discrete spectrum are given by

$$\phi_{j\pm}(\mu) = \pm\tfrac{1}{2}c\nu_j\frac{M(\mu, \pm\nu_j)}{\pm\nu_j - \mu}, \quad (26)$$

while the continuous spectrum solution is expressed as

$$\phi_\nu(\mu) = \tfrac{1}{2}c\nu\,P\frac{M(\mu,\nu)}{\nu-\mu} + \lambda(\nu)\delta(\mu-\nu) \quad (27)$$

with

$$\lambda(\nu) = 1 - \tfrac{1}{2}c\nu P\int_{-1}^{+1}\frac{M(\mu,\nu)}{\nu-\mu}\,d\mu, \quad (28)$$

where P denotes the Cauchy principal value of the integral.

It is seen from (24) that $\Lambda(\nu)$ is an analytic function of ν over the whole ν plane with only a cut on the real axis between -1 and $+1$. At infinity, $\Lambda(\nu)$ is bounded and equal to

$$\lim_{\nu\to\infty}\Lambda(\nu) = 1 - \tfrac{1}{2}c\int_{-1}^{+1}M(\mu,\mu)\,d\mu = \prod_{i=0}^{\infty}(1 - cf_i). \quad (29)$$

Denoting by $\Lambda^+(\nu)$ and $\Lambda^-(\nu)$ the boundary values of $\Lambda(\nu)$ above and below the cut $(-1, +1)$, respectively, one has the relations

$$\Lambda^\pm(\nu) = \lambda(\nu) \pm \tfrac{1}{2}i\pi c\nu M(\nu,\nu)$$

or

$$\Lambda^+(\nu) + \Lambda^-(\nu) = 2\lambda(\nu),$$
$$\Lambda^+(\nu) - \Lambda^-(\nu) = i\pi c\nu M(\nu, \nu). \tag{30}$$

The normalization integrals for the discrete and continuous spectrum eigenfunctions are defined by

$$N_{j\pm} = \int_{-1}^{+1} \mu \phi_{j\pm}^2(\mu)\, d\mu \tag{31}$$

and

$$N(\nu)\psi(\nu) = \int_{-1}^{+1} \mu \phi_\nu(\mu) \int_{-1}^{+1} \psi(\eta)\phi_\eta(\mu)\, d\eta\, d\mu, \tag{32}$$

where $\psi(\nu)$ is a function which may be expanded in terms of the eigenfunctions $\phi_\nu(\mu)$ and $\phi_{j\pm}(\mu)$. Convenient expressions for these integrals are

$$N_{j\pm} = \pm\tfrac{1}{2}c\nu_j^2 M(\nu_j, \nu_j) \frac{\partial \Lambda(\nu)}{\partial \nu}\bigg|_{\nu_j} \tag{33}$$

and

$$N(\nu) = \nu\Lambda^+(\nu)\Lambda^-(\nu). \tag{34}$$

C. Orthogonality Relations for $h_l(\nu)$

Mika[3] has proved that, for an anisotropy of finite order N, any function $\psi(\mu)$ satisfying the Hölder condition in the interval $(-1, +1)$ can be expanded there in terms of the eigenfunctions (26) and (27) of the whole spectrum as

$$\psi(\mu) = \int_{-1}^{+1} A(\nu)\phi_\nu(\mu)\, d\nu + \sum_{j=1}^{M} [\alpha_{j^+}\phi_{j^+}(\mu) + \alpha_{j^-}\phi_{j^-}(\mu)], \tag{35}$$

where

$$A(\nu) = \frac{1}{N(\nu)} \int_{-1}^{+1} \mu\phi_\nu(\mu)\psi(\mu)\, d\mu,$$
$$\alpha_{j^\pm} = \frac{1}{N_{j^\pm}} \int_{-1}^{+1} \mu\phi_{j^\pm}(\mu)\psi(\mu)\, d\mu, \tag{36}$$

and $2M$ is the total number of the discrete eigenvalues.

Consider the expansion of $P_k(\mu)/(1 - cf_k)$, where $k < N$. One obtains from (36), using the expansion (19), the recurrence and orthogonality relations of the $P_l(\mu)$, and the recurrence relation (1) for $h_k(\nu)$,

$$A(\nu) = [\nu/N(\nu)]h_k(\nu)$$

and

$$\alpha_{j^\mp} = (\mp\nu_j/N_{j^\mp})h_k(\nu_j). \tag{37}$$

Substituting in (35), multiplying by $P_l(\mu)$, and integrating over μ from -1 to $+1$, one is led to the expression

$$\frac{2\delta_{kl}}{(2l+1)(1-cf_l)} = \int_{-1}^{+1} P_l(\mu)\, d\mu \int_{-1}^{+1} \frac{\nu}{N(\nu)} h_k(\nu)\phi_\nu(\mu)\, d\nu + \int_{-1}^{+1} P_l(\mu)\, d\mu \sum_{j=1}^{M} \left(\frac{\nu_j}{N_{j^+}} h_k(\nu_j)\phi_{j^+}(\mu) + \frac{-\nu_j}{N_{j^-}} h_k(-\nu_j)\phi_{j^-}(\mu) \right),$$

or, since the inversion of the order of integrations can be justified here (the first term on the right-hand side, which is the critical one, contains only one principal-value integral[13]), to the orthogonality relation

$$\frac{2\delta_{kl}}{(2l+1)(1-cf_l)} = \int_{-1}^{+1} h_k(\nu)h_l(\nu) \frac{\nu}{N(\nu)}\, d\nu + \sum_{j=1}^{M} \left[h_k(\nu_j)h_l(\nu_j) \frac{\nu_j}{N_{j^+}} + h_k(-\nu_j)h_l(-\nu_j)\left(\frac{-\nu_j}{N_{j^-}}\right) \right]. \tag{38}$$

Thus, for finite N, the polynomials $h_l(\nu)$ form a set which is orthogonal in the Stieltjes sense. The orthogonalization interval covers both the continuous and discrete spectra.

The relation (38) is established for any finite N. However, it may immediately be generalized to the limiting case of infinite N by the following argument. Consider (38) for a given

$$M_N(\mu, \nu) = \sum_{n=0}^{N} (2n+1) f_n P_n(\mu) h_n(\nu)$$

and $k, l < N$. The polynomials $h_k(\nu)$ which have been determined from M_N for $k < N$ do not change when new terms $\sum_{n=N+1}^{N+m} (2n+1) f_n P_n(\mu) h_n(\nu)$ are added to M_N. Only the eigenvalues and the expressions for $N(\nu)$, $N_j(\nu_j)$ are changed. In the limit for $N \to \infty$, $N(\nu)$ and $N_j(\nu_j)$ will exist, since we have assumed the uniform convergence of the series for $M(\mu, \nu)$ so that $\Lambda(\nu)$, $M(\nu, \nu)$, and $\lambda(\nu)$ will all converge for $N \to \infty$. [It will be seen in Sec. 3F that the total number of discrete eigenvalues remains finite even for $N \to \infty$.] It is thus clear that the relation (38) will continue to be valid in the limit for $N \to \infty$.

Another argument for the validity of (38) for $N \to \infty$ may be obtained by considering the relevant moment problem; i.e., the problem of determining a function $\omega(\nu)$ nondecreasing in a given interval (a, b) which will satisfy the equations

$$\int_a^b \nu^k\, dw(\nu) = m_k, \quad \text{for} \quad k = 0, 1, 2, \cdots, \tag{39}$$

where the moments m_k form a given infinite sequence of real numbers. This problem has only one solution if the interval (a, b) is finite. Now, in our case, the interval (a, b) is given by $(-\nu_1, \nu_1)$ and remains finite even for $N \to \infty$. [An expression for ν_1, valid in the case of $N \to \infty$, is given in Ref. 6.] Further, the relations (38) imply that the moments $m_0, m_2, \cdots, m_{2k}$ determined from $M_k(\mu, \nu)$ do not change when

[13] See, e.g., N. I. Muskhelishvili, *Singular Integral Equations* (P. Noordhoff Ltd., Groningen, The Netherlands, 1953), p. 59.

new terms are added to M_k. Thus the moments m_k, obtained from the relations (38) for all k, determine uniquely a function $\omega(\nu)$ which may be expressed as the continuous distribution

$$\omega(\nu) = \int_0^{\nu} \frac{d\nu}{\Lambda^+(\nu)\Lambda^-(\nu)} = \int_0^{\nu} \frac{d\nu}{\lambda^2(\nu) + \frac{1}{4}\pi^2 c^2 \nu^2 M^2(\nu, \nu)}, \quad \text{for} \quad |\nu| \leq 1, \tag{40}$$

and the sequence of horizontal lines

$$\omega(\mp\nu) = \omega(\mp 1) \mp \sum_{i=m}^{M} \frac{\nu_i}{N_i}, \quad \text{for} \quad \nu_m < \nu < \nu_{m-1}, \tag{41}$$

with

$$\omega(\mp\nu) = \omega(\mp 1) \mp \sum_{i=1}^{M} \frac{\nu_i}{N_i}, \quad \text{for} \quad \nu_1 < \nu.$$

The set of orthogonal polynomials constructed by means of this weight function $\omega(\nu)$ over the interval $(-\nu_1, \nu_1)$, according to the Tchebycheff theory, are identical with the polynomials $h_l(\nu)$, ensuring again the validity of the relations (38) for $N \to \infty$. We write them concisely as

$$\int_{-1}^{+1} h_k(\nu) h_l(\nu)\, d\omega(\nu) \equiv (h_k, h_l) = n_l \delta_{kl}, \tag{42}$$

where

$$n_l = 2/(2l + 1)(1 - cf_l) = 2/\xi_l. \tag{43}$$

We have seen that the orthogonality of the $h_k(\nu)$, with $\omega(\nu)$ as the weight function, follows from the completeness of the eigenfunctions $\phi_\nu(\mu)$. It is clear that, conversely, from the completeness of the $P_l(\mu)$ and the orthogonality of the $h_l(\nu)$, the completeness of the $\phi_\nu(\mu)$ will follow. The expansion of $P_l(\mu)$ in terms of $\phi_\nu(\mu)$ may be immediately derived from the orthogonality of $h_l(\nu)$ by multiplying (38) with $P_k(\mu)$ and summing over k. [Here, we are assuming that the infinite sum (19) converges, in the distribution sense, to $\phi_\nu(\mu)$. A proof is given in Appendix A.] In this way the "full-range" completeness theorem of the $\phi_\nu(\mu)$ is extended to cover the case of anisotropic scattering of infinite order.

We should also note here that the orthogonality relations for $\phi_\nu(\mu)$ lead to the closure relations for the normalized $h_l(\nu)$. We have

$$\frac{N(\nu)}{\nu}\delta(\nu - \nu') = \frac{1}{\nu}\int_{-1}^{+1} \phi_\nu(\mu)\phi_{\nu'}(\mu)\mu\, d\mu = \sum_{l=0}^{\infty} \frac{1}{n_l} h_l(\nu)h_l(\nu'), \tag{44}$$

for $|\nu| < 1$, and

$$\frac{N_j}{\nu_j}\delta_{jk} = \frac{1}{\nu_j}\int_{-1}^{+1} \phi_j(\mu)\phi_k(\mu)\mu\, d\mu = \sum_{l=0}^{\infty} \frac{1}{n_l} h_l(\nu_j)h_l(\nu_k), \tag{45}$$

for the discrete spectrum.

Remark: Our restriction that there should be no discrete roots for $|\nu| < 1$ is not an essential one for the proof of orthogonality of $h_l(\nu)$. If $\lambda(\nu)$ and $M(\nu, \nu)$ do have common roots $\mp\nu_i$ in the interval $(-1, +1)$, these ν_i will be discrete eigenvalues embedded in the continuous spectrum, and the orthogonality relation (38) will only be modified then by the addition of the terms

$$\sum_{i=1}^{M'} \left(h_k(\nu_i)h_l(\nu_i)\frac{\nu_i}{N_{i+}} + h_k(-\nu_i)h_l(-\nu_i)\frac{-\nu_i}{N_{i-}} \right),$$

where $2M'$ is the number of all such roots ν_i.

3. SOME PROPERTIES OF $h_l(\nu)$ CONSIDERED AS ORTHOGONAL POLYNOMIALS

A. Various Standardizations

From the Tchebycheff theory of the orthogonal polynomials,[14,15] one can obtain many properties of the $h_l(\nu)$. We list here some of them which seem relevant for the solution of transport problems. We note first that $h_l(\nu)$ form symmetrical polynomials since the interval extends from $-\nu_1$ to $+\nu_1$ and the weight function is odd.

The weight function and the interval of integration determine the orthogonal polynomials up to multiplicative constants. As in neutron-transport theory, the recursion relation (1) and the initial condition (2) arise naturally; one obtains for the coefficient of the highest power in $h_l(\nu)$ the expression (12). In the literature, two other definitions are commonly used:

(i) The normalized polynomials $h_l^{(\mathrm{norm})}(\nu)$, specified by

$$h_l^{(\mathrm{norm})}(\nu) = h_l(\nu)/(n_l)^{\frac{1}{2}},$$

satisfy the recurrence relation

$$\nu h_l^{(\mathrm{norm})}(\nu) = \omega_l h_{l+1}^{(\mathrm{norm})}(\nu) + \omega_{l-1} h_{l-1}^{(\mathrm{norm})}(\nu), \tag{46}$$

where ω_l is given by (14).

(ii) The polynomials which are standardized by having the coefficient of the highest power set to unity,

[14] M. J. Shohat, Théorie générale des polynomes orthogonaux de Tchebichef, Mémorial des Sciences Mathématiques, Fasc. LXVI (1934).

[15] N. I. Akhiezer, *The Classical Moment Problem* (Hafner Publ. Co., New York, 1965).

denoted by $h_l^{(st)}(\nu)$, are

$$h_l^{(st)}(\nu) = h_l(\nu)/\alpha_{l,l}. \tag{47}$$

The recurrence relation satisfied by $h_l^{(st)}(\nu)$ is

$$\nu h_l^{(st)}(\nu) = h_{l+1}(\nu) + w_{l-1}^2 h_{l-1}^{(st)}(\nu). \tag{48}$$

B. Polynomials of the Second Kind and Relations Between Consecutive Polynomials

The finite-difference equation (1) has two linearly independent polynomial-type solutions. We have considered, so far, the solution which corresponds to the initial conditions $h_0 = 1$ and $h_1 = (1 - c)\nu$. The second solution may be defined by $h_0 = 0$, $h_1 = 1$ or, introducing $k_{l-1}(\nu) = h_l(\nu)$, by

$$k_{-1} = 0 \quad \text{and} \quad k_0 = 1. \tag{49}$$

We have, then,

$$k_1(\nu) = \tfrac{3}{2}(1 - cf_1)\nu,$$
$$k_2(\nu) = (1/3!)[3 \times 5(1 - cf_1)(1 - cf_2)\nu^2 - 4]; \tag{50}$$

$k_l(\nu)$ is a polynomial of lth degree in ν. We have again $k_l(-\nu) = (-1)^l k_l(\nu)$ and the coefficient of the highest power is given by

$$\alpha_{l,l}(k) = \frac{1}{(l+1)!}\left(\prod_{i=1}^{l} \xi i\right) = \frac{1}{\xi_0}\,\alpha_{l+1,l+1}(h). \tag{51}$$

It is easy to see that, in general,

$$k_{l-1}(\nu) = \tfrac{1}{2}\int_{-\nu_1}^{+\nu_1} \frac{h_l(\nu) - h_l(\nu')}{\nu - \nu'}\, dw(\nu'). \tag{52}$$

The $k_l(\nu)$ may be called polynomials of the second kind associated with the recurrence relation (1). For $c = 0$, the $k_l(\nu)$ reduce to the $w_l(\nu)$ associated with the Legendre equation. Also for $f_0 = 1$, $f_l = 0$ $(l > 1)$, i.e., isotropic scattering, one has again $k_l(\nu) = w_l(\nu)$.

Using the difference equation (1) and the initial conditions, one easily establishes the analog of the Liouville–Ostrogradskii formula

$$h_l(\nu)k_l(\nu) - h_{l+1}(\nu)k_{l-1}(\nu) = 1/(l+1). \tag{53}$$

In general, if $u_k(\lambda)$ and $v_k(\mu)$ are two solutions of Eq. (1), one has the analog of Green's formula,

$$n[u_{n-1}(\lambda)v_n(\mu) - u_n(\lambda)v_{n-1}(\mu)] - m[u_{m-1}(\lambda)v_m(\mu) - u_m(\lambda)v_{m-1}(\mu)] = (\mu - \lambda)\sum_{k=m}^{n-1} \xi_k u_k(\lambda)v_k(\mu). \tag{54}$$

In particular, taking $m = 1$, $u_n(\lambda) = h_n(\lambda)$, $v_n(\mu) = h_n(\mu)$, one obtains one of the Christoffel–Darboux formulas

$$\frac{n[h_{n-1}(\lambda)h_n(\mu) - h_n(\lambda)h_{n-1}(\mu)]}{2(\mu - \lambda)} = \sum_{k=0}^{n-1} \frac{h_k(\lambda)h_k(\mu)}{n_k} = \sum_{k=0}^{n-1} h_k^{(\text{norm})}(\lambda)h_k^{(\text{norm})}(\mu). \tag{55}$$

The sum in (55) is usually called the kernel polynomial $S_{n-1}(\lambda, \mu)$ belonging to the system $h_k(\lambda)$. The formula can be transformed, using the relation (1), into

$$\frac{(n+1)}{2(\mu - \lambda)}[\lambda h_n(\lambda)h_{n+1}(\mu) - \mu h_{n+1}(\lambda)h_n(\mu)] = \sum_{k=0}^{n-1}(k+1)h_k(\lambda)h_{k+1}(\mu). \tag{56}$$

In the limit for $\mu \to \lambda$, one obtains, from (55) and (56),

$$n[h_{n-1}(\lambda)h_n'(\lambda) - h_n(\lambda)h_{n-1}'(\lambda)] = \sum_{k=0}^{n-1} \xi_k h_k^2(\lambda); \tag{57}$$

$$\tfrac{1}{2}(n+1)\{\lambda[h_n(\lambda)h_{n+1}'(\lambda) - h_{n+1}(\lambda)h_n'(\lambda)] - h_n(\lambda)h_{n+1}(\lambda)\} = \sum_{k=0}^{n-1}(k+1)h_k(\lambda)h_{k+1}(\lambda). \tag{58}$$

It is of interest for applications to transport problems to consider relations which involve the $h_n(\nu)$ and the Legendre functions. Out of a large number of such relations, which can be constructed using the respective recurrence relations, we give a few involving P_l, h_l and Q_l, h_l:

$$(\lambda - \mu)\sum_{l=0}^{n}(2l+1)P_l(\lambda)h_l(\mu) = (n+1)[P_{n+1}(\lambda)h_n(\mu) - P_n(\lambda)h_{n+1}(\mu)] - c\mu\sum_{l=0}^{n}(2l+1)f_lP_l(\lambda)h_l(\mu), \tag{59}$$

$$(\lambda - \mu)\sum_{l=0}^{n}(2l+1)Q_l(\lambda)h_l(\mu) = (n+1)[Q_{n+1}(\lambda)h_n(\mu) - Q_n(\lambda)h_{n+1}(\mu)] - c\mu\sum_{l=0}^{n}(2l+1)f_lQ_l(\lambda)h_l(\mu) + 1, \tag{60}$$

$$\sum_{l=0}^{n}(2l+1)f_lP_l(\lambda)h_l(\lambda) = \left(\frac{n+1}{c\lambda}\right)[P_{n+1}(\lambda)h_n(\lambda) - P_n(\lambda)h_{n+1}(\lambda)], \tag{61}$$

$$\sum_{l=0}^{n}(2l+1)f_lQ_l(\lambda)h_l(\lambda) = \frac{1}{c\lambda} + \left(\frac{n+1}{c\lambda}\right)[Q_{n+1}(\lambda)h_n(\lambda) - Q_n(\lambda)h_{n+1}(\lambda)]. \tag{62}$$

One may define a Tchebycheff function of the second kind $K_l(\nu)$ by the integral

$$K_l(\nu) = \tfrac{1}{2}\int_{-\nu_1}^{+\nu_1} \frac{h_l(\nu')}{\nu - \nu'}\, dw(\nu'). \tag{63}$$

The functions $K_l(\nu)$ satisfy the recurrence relation (1). In analogy to Legendre functions, $K_l(\nu)$ may be expressed by means of $K_0(\nu)$ and the two sets of polynomials as

$$K_l(\nu) = K_0(\nu)h_l(\nu) - k_{l-1}(\nu). \tag{64}$$

C. Expression in Terms of the Moments

The moments of $w(\nu)$ were defined in (39) with $a = -\nu_1$, $b = +\nu_1$. From symmetry, we have

$$\begin{aligned} m_k &= 0, && \text{for odd } k, \\ m_k &= 2\int_0^{+\nu_1} \nu^k\, dw(\nu), && \text{for even } k. \end{aligned} \tag{65}$$

Consider the determinants

$$\Delta_l \equiv \begin{vmatrix} m_0 & m_1 & \cdots & m_{l-1} \\ m_1 & m_2 & \cdots & m_l \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ m_{l-1} & m_l & \cdots & m_{2l-2} \end{vmatrix}, \quad \text{for } l \geq 1, \quad \Delta_0 = 1. \tag{66}$$

Because Δ_l is the discriminant of the positive-definite quadratic form

$$\sum_{i,j=0}^{l-1} m_{i+j} g_i g_j = \int_{-\nu_1}^{+\nu_1} \left(\sum_{i=0}^{l-1} g_i \nu^i\right)^2 dw(\nu),$$

all the Δ_l are positive. The normalized polynomials $h_l^{(\mathrm{norm})}(\nu)$ can be expressed in terms of the moments as

$$h_l^{(\mathrm{norm})}(\nu) = (\Delta_l \Delta_{l+1})^{-\frac{1}{2}} \begin{vmatrix} m_0 & m_1 & \cdots & m_l \\ m_1 & m_2 & \cdots & m_{l+1} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ m_{l-1} & m_l & \cdots & m_{2l-1} \\ 1 & \nu & \cdots & \nu^l \end{vmatrix}. \tag{67}$$

The connection between Δ_l and a_l is given by

$$\Delta_l/\Delta_{l+1} = a_{l,l}^2/n_l$$

or

$$\Delta_l = \left(\prod_{i=0}^{l-1} n_i\right) \Big/ \left(\prod_{i=0}^{l-1} a_{i,i}\right)^2 = 2^l \Big/ l!\, a_{l,l} \left(\prod_{i=0}^{l-1} a_{i,i}\right)^2 \tag{68}$$

with the convention $a_{0,0} = 1$. These relations may be used to express all the moments successively in terms of c and f_l. One has for the first few moments

$$\begin{aligned} m_0 &= 2/(1-c), \\ m_2 &= 2/3(1-c)^2(1-cf_1), \\ m_4 &= \frac{2}{9(1-c)^2(1-cf_1)^2}\left(\frac{1}{1-c} + \frac{4}{5(1-cf_2)}\right). \end{aligned} \tag{69}$$

We note that the moment m_{2l} involves the coefficients f_k only up to $k = l$.

D. Expression by Means of a Continued Fraction

Consider the following infinite continued fraction:

$$F(\nu) = \cfrac{1}{\nu - \cfrac{w_0^2}{\nu - \cfrac{w_1^2}{\nu - \cfrac{w_2^2}{\nu - \cdots}}}} \equiv \frac{1}{|\nu} - \frac{w_0^2|}{|\nu} - \frac{w_1^2|}{|\nu} - \cdots, \tag{70}$$

where w^2 is defined in (14). Following Akhiezer,[15] we denote by

$$N_l(\nu) = \nu^{l-1} + \alpha_1 \nu^{l-3} + \cdots + \alpha_{l-1}$$

and

$$D_l(\nu) = \nu^l + \beta_1 \nu^{l-2} + \cdots + \beta_l,$$

the numerators and denominators of the successive finite approximations to this continued fraction. We thus have

$$\begin{aligned} F_1 &= N_1(\nu)/D_1(\nu) = 1/\nu, \\ F_2 &= N_2(\nu)/D_2(\nu) = 1/(\nu - w_0^2/\nu) = \nu/(\nu^2 - w_0^2), \end{aligned}$$

etc. We specify further that

$$N_0(\nu) = 0, \quad N_1(\nu) = 1 \quad \text{and} \quad D_0(\nu) = 1, \quad D_1(\nu) = \nu.$$

It can be easily checked that both of the polynomials $N_l(\nu)$ and $D_l(\nu)$ satisfy the recurrence relation (48) for the $h_l^{(\mathrm{st})}(\nu)$. Hence, for the infinite continued fraction $(1/\xi_0)F(\nu)$, the lth approximation will be just equal to

$$(1/\xi_0)F_l(\nu) = k_{l-1}(\nu)/h_l(\nu). \tag{71}$$

E. Asymptotic Expressions of $h_{l+1}(\nu)/h_l(\nu)$ and $h_l(\nu)$ for $l \to \infty$ and $|\nu| > 1$

The asymptotic value of $h_{l+1}(\nu)/h_l(\nu)$ for $l \to \infty$ and $|\nu| > 1$ can be easily obtained from an application of Poincaré's theorem.[16] Writing the recurrence relation (1) as

$$h(l+2, \nu) + P_1(l)h(l+1, \nu) + P_0(l)h(l, \nu) = 0,$$

[16] A. O. Guelfond, *Calcul des différences finies* (Dunod Cie, Paris, 1963).

where

$$P_1(l) = -[(2l+3)/(l+2)](1 - cf_{l+1})\nu,$$
$$P_0(l) = (l+1)/(l+2),$$

so that

$$\lim_{l\to\infty} P_1(l) = -2\nu,$$
$$\lim_{l\to\infty} P_0(l) = 1,$$

one sees that the equation

$$\lambda^2 - 2\nu\lambda + 1 = 0$$

must be considered. The two roots are

$$\lambda_{1,2} = \nu \pm (\nu^2 - 1)^{\frac{1}{2}}.$$

According to Poincaré's theorem, the limit of $h(l+1, \nu)/h(l, \nu)$ for $l \to \infty$ will be equal to one of these roots, provided they are different in modulus, which is the case for $|\nu| > 1$, but not for $|\nu| < 1$. Since, for $c = 0$, $h_l(\nu)$ reduces to $P_l(\nu)$ and for the Legendre polynomials we have the asymptotic relation[17]

$$\lim_{l\to\infty} \frac{P_{l+1}(\nu)}{P_l(\nu)} = \lambda_1,$$

it follows that

$$\lim_{l\to\infty} \frac{h_{l+1}(\nu)}{h_l(\nu)} = \lim_{l\to\infty} \frac{P_{l+1}(\nu)}{P_l(\nu)} = \lambda_1 = \nu + (\nu^2 - 1)^{\frac{1}{2}}. \quad (72)$$

Consequently, we also have, for $|\nu| > 1$, the limiting relation

$$\lim_{l\to\infty} \frac{h_l(\nu)}{P_l(\nu)} = \text{a function of } \nu, \text{ independent of } l.$$

To find this function, we consider the definition of $\Lambda(\nu)$ in (24). Integrating over μ, after changing the order of summation and integration, one obtains

$$\Lambda(\nu) = 1 - c\nu \sum_{l=0}^{\infty} (2l+1) f_l Q_l(\nu) h_l(\nu) \quad (73)$$

or, using (62),

$$\Lambda(\nu) = \lim_{l\to\infty} (l+1)[Q_l(\nu)h_{l+1}(\nu) - Q_{l+1}(\nu)h(\nu)], \quad (74)$$

which, with the relation (72), may be transformed into

$$\Lambda(\nu) = \lim_{l\to\infty} \frac{h_l(\nu)}{P_l(\nu)} \quad \text{for} \quad |\nu| > 1. \quad (75)$$

It may also be of interest to consider the remainder term in the relation (72). Defining $\epsilon_n(\nu)$ by

$$h_{l+1}(\nu)/h_l(\nu) = P_{l+1}(\nu)/P_l(\nu) - \epsilon_l(\nu), \quad (76)$$

where $\epsilon_l \to 0$ for $l \to \infty$, we obtain, from (25) and (61),

$$M(\nu, \nu) = \lim_{l\to\infty} \frac{l+1}{c\nu} [P_{l+1}(\nu)h_l(\nu) - P_l(\nu)h_{l+1}(\nu)] \quad (77)$$

$$= \lim_{l\to\infty} \frac{l+1}{c\nu} P_l(\nu)h_l(\nu)\epsilon_l(\nu). \quad (78)$$

F. Zeros of the Polynomials $h_l(\nu)$ and $k_l(\nu)$

We quote from Refs. 14 and 15 some of the well-known theorems about the zeros of the orthogonal polynomials which relate here to h_l and k_l.

(i) All the zeros of $h_l(\nu)$ are real, simple, and situated between $-\nu_1$ and $+\nu_1$.[15]

(ii) Any two zeros of the polynomial $h_l(\nu)$ are separated by a zero of $h_{l-1}(\nu)$ and vice versa.[15]

(iii) Two consecutive zeros of $h_l(\nu)$ are separated by at least one zero of $h_{l'+l}(\nu)$, where $l' \geq 1$.[14]

(iv) The zeros of the polynomial $k_l(\nu)$ are real, simple, and contained in the interval $(-\nu_1, +\nu_1)$.

(v) Any two zeros of $k_{l-1}(\nu)$ are separated by a zero of $h_l(\nu)$ and vice versa.[15]

(vi) The two zeros of $h_l(\nu)$ which have the largest absolute value approach $\mp\nu_1$, in the limit for $l \to \infty$.

(vii) In each partial interval (α, β) within $(-1, +1)$ one can find, choosing l sufficiently large, as many roots of $h_l(\nu)$ as one wishes. The general theorem requires for this result the condition

$$\int_{\alpha^1}^{\beta^1} dw(\nu) \neq 0, \quad \text{for every } \alpha^1, \beta^1,$$

where $\alpha \leq \alpha^1 < \beta^1 \leq \beta$, which is clearly satisfied in this case as $dw > 0$ in $(-1, +1)$.

Outside the interval $(-1, +1)$, $h_l(\nu)$ can have at most one root between any two consecutive roots ν_j, ν_{j+1} (or $-\nu_j$, $-\nu_{j+1}$) of $\Lambda(\nu)$.[14]

Considering the whole sequence of polynomials $h_l(\nu)$ for $l \to \infty$, one sees that the roots cover the interior of the interval $(-1, +1)$ everywhere densely, while outside $(-1, +1)$, the $h_l(\nu)$ have a limited number of zeros.[14]

By means of relation (75), one can obtain from these theorems very useful information about the zeros of $\Lambda(\nu)$ for $|\nu_j| > 1$. We see that the total number of zeros of $\Lambda(\nu)$ outside the interval $(-1, +1)$ is finite, all these zeros are real and simple, [the non-degeneracy of the zeros of $\Lambda(\nu)$ for $0 < c < 1$ previously proved by Zweifel[18]] and they will be reached as the limits of the zeros of the polynomials $h_l(\nu)$ for $l \to \infty$ and that the approach to the limiting

[17] G. Szegö, Am. Math. Soc. 23, 192 (1959).

[18] P. F. Zweifel, private communication, 1968.

values is always from the left for $\nu > 1$ and from the right for $\nu < -1$.

G. Quadrature Formula

Consider an arbitrary polynomial of degree $2l - 1$, say $R_{2l-1}(\nu)$. One can write, by division with $h_l(\nu)$,

$$R_{2l-1}(\nu) = h_l(\nu)S_{l-1}(\nu) + T_{l-1}(\nu), \tag{79}$$

where $S_{l-1}(\nu)$ and $T_{l-1}(\nu)$ are some polynomials of degrees $n < l - 1$. Using Lagrange's interpolation formula, $T_{l-1}(\nu)$ may be expressed as

$$\begin{aligned} T_{l-1}(\nu) &= h_l(\nu)\sum_{k=1}^{l} \frac{T_{l-1}(\nu_{lk})}{h_l'(\nu_{lk})(\nu - \nu_{lk})} \\ &= h_l(\nu)\sum_{k=1}^{l} \frac{R_{2l-1}(\nu_{lk})}{h_l'(\nu_{lk})(\nu - \nu_{lk})}, \end{aligned} \tag{80}$$

where ν_{lk} are the l zeros of $h_l(\nu)$ which are all situated between $-\nu_1$ and $+\nu_1$.

On the other hand, using the orthogonality property of $h_l(\nu)$, one obtains

$$\begin{aligned} \int_{-\nu_1}^{+\nu_1} R_{2l-1}(\nu)\, dw(\nu) &= \int_{-\nu_1}^{+\nu_1} R_{l-1}(\nu)\, dw(\nu) \\ &= \sum_{k=1}^{l} \frac{2R_{2l-1}(\nu_{lk})}{h_l'(\nu_{lk})} k_{l-1}(\nu_{lk}). \end{aligned} \tag{81}$$

Thus the following quadrature formula, valid for an arbitrary polynomial $R(\nu)$ of degree $n \leq 2l - 1$, is established:

$$\int_{-\nu_1}^{+\nu_1} R(\nu)\, dw(\nu) = \sum_{k=1}^{l} \mu_{lk} R(\nu_{lk}), \tag{82}$$

where

$$\mu_{lk} = 2[k_{l-1}(\nu_{lk})/h_l'(\nu_{lk})]. \tag{83}$$

Two other representations for μ_{lk} may be of use. One can write

$$\mu_{lk} = 2\,\frac{k_{l-1}(\nu_{lk})h_{l-1}(\nu_{lk}) - k_{l-2}(\nu_{lk})h_l(\nu_{lk})}{h_l'(\nu_{lk})h_{l-1}(\nu_{lk}) - h_l(\nu_{lk})h_{l-1}'(\nu_{lk})}$$

which, using relations (53) and (57), reduces to

$$\mu_{lk} = \left(\sum_{i=0}^{l-1} \tfrac{1}{2}\xi_i h_i^2(\nu_{lk})\right)^{-1} \tag{84}$$

On the other hand, consider the expression

$$R(\nu) = [h_l(\nu)/h_l'(\nu_{lk})(\nu - \nu_{lk})]^2,$$

representing a polynomial of degree $2l - 2$ which vanishes at all the zeros of $h_l(\nu)$ except at $\nu = \nu_{lk}$ where it equals unity. Inserting $R(\nu)$ into the quadrature formula, one obtains immediately

$$\mu_{lk} = \int_{-\nu_1}^{+\nu_1} \left(\frac{h_l(\nu)}{h_l'(\nu_{lk})(\nu - \nu_{lk})}\right)^2 d\omega(\nu). \tag{85}$$

Both (84) and (85) show the positive nature of the weights μ_{lk}.

We note, further, the relation obtained from (82) by letting $R(\nu) = 1$. We have

$$\sum_{k=1}^{l} \mu_{lk} = \int_{-\nu_1}^{+\nu_1} dw(\nu) = m_0 = 2/(1 - c) \tag{86}$$

or, in general,

$$\sum_{k=1}^{l} \mu_{lk}\nu_{lk}^n = \int_{-\nu_1}^{+\nu_1} \nu^n\, dw(\nu) = m_n, \quad \text{for} \quad n \leq 2l - 1. \tag{87}$$

A quadrature formula may also be obtained for the quasiorthogonal polynomial

$$h_l(\nu, \tau) = h_l(\nu) - \tau h_{l-1}(\nu),$$

where τ is a real parameter (cf. Ref. 15).

As an application of these formulas, we point out the expansion of the ratio $2k_{l-1}(\nu)/h_l(\nu)$ in inverse powers of ν.

Using Lagrange's interpolation formula, one may write

$$\frac{2k_{l-1}(\nu)}{h_l(\nu)} = \sum_{k=1}^{l} \frac{2k_{l-1}(\nu_{lk})}{h_l'(\nu_{lk})(\nu - \nu_{lk})},$$

which, using (83) and expanding in powers of $1/\nu$, leads, through (87), to

$$\frac{m_0}{\nu} + \frac{m_2}{\nu^3} + \cdots + \frac{m_{2l-2}}{\nu^{2l-1}} + O\left(\frac{1}{\nu^{2l+1}}\right).$$

Comparing with (71), we see the connection between the continued fraction (70) and the moment expansion. We have

$$\frac{2}{\xi_0} F_l(\nu) = \frac{2k_{l-1}(\nu)}{h_l(\nu)} = \sum_{i=0}^{2l-2} \frac{m_i}{\nu^{i+1}} + O\left(\frac{1}{\nu^{2l+1}}\right), \tag{88}$$

which gives, in the limit for $l \to \infty$,

$$\frac{2}{\xi_0} F(\nu) = \lim_{l\to\infty} \frac{2k_{l-1}(\nu)}{h_l(\nu)} = \sum_{i=0}^{\infty} \frac{m_i}{\nu^{i+1}}. \tag{89}$$

This expansion is also related to $K_0(\nu)$, since one has for $|\nu| > \nu_1$,

$$K_0(\nu) = \tfrac{1}{2} \int_{-\nu_1}^{+\nu_1} \frac{dw(\nu')}{\nu - \nu'} = \tfrac{1}{2} \sum_{i=0}^{\infty} \frac{m_i}{\nu^{i+1}}. \tag{90}$$

ACKNOWLEDGMENTS

This work was started at the Physics Department of Columbia University and completed at the Center for Theoretical Studies of Miami University. It is a pleasure to thank Professor A. Sachs, Chairman of the Physics Department, and the members of the Theoretical group, in particular, Professor T. D. Lee

and Professor R. Serber at Columbia, and Professor B. Kurşunoğlu, Director of the Center for Theoretical Studies at Miami, for the kind hospitality accorded to the author at these institutions. Stimulating and enlightening discussions with Professor T. D. Lee, Professor P. F. Zweifel, Professor A. Cakmak, and Dr. W. Holman, III are gratefully acknowledged.

APPENDIX: THE SUM OF THE SPHERICAL HARMONICS EXPANSION (19)

We wish to indicate here that the infinite sum (19) converges, in the distribution sense, to the eigenfunctions (26) for $\nu = \nu_j$ and to (27) for $|\nu| < 1$. For isotropic and linearly anisotropic scattering, the proofs have been given by Nonnenmacher[19] and Mehner,[20] respectively. Their method can be applied immediately to the present case of anisotropy of infinite order.

(i) $\nu = \nu_j$: We have, from Eqs. (19), (59), and (75), that

$$\phi_{\nu_j}(\mu) = \lim_{n\to\infty} \sum_{l=0}^{n} \frac{2l+1}{2} P_l(\mu) h_l(\nu_j) \tag{A1}$$

$$= \frac{c\nu_j}{2(\nu_j - \mu)} \lim_{n\to\infty} \sum_{l=0}^{n} (2l+1) f_l P_l(\mu) h_l(\nu_j) - \frac{1}{2(\nu_j - \mu)} \lim_{n\to\infty} (n+1) \times [P_{n+1}(\mu) h_n(\nu_j) - P_n(\mu) h_{n+1}(\nu_j)] \tag{A2}$$

$$= \frac{c\nu_j}{2(\nu_j - \mu)} M(\mu, \nu_j). \tag{A3}$$

(ii) $|\nu| < 1$: We have, similarly, from (19) and (59),

$$\phi_\nu(\mu) = [c\nu/2(\nu - \mu)] M(\mu, \nu) + \lim_{n\to\infty} \Delta_n(\mu, \nu), \tag{A4}$$

where

$$\Delta_n(\mu, \nu) = [(n+1)/2(\nu - \mu)] \times [P_n(\mu) h_{n+1}(\nu) - P_{n+1}(\mu) h_n(\nu)]. \tag{A5}$$

Both terms of (A5) are distributions. The first one defines a distribution $\langle 1/(\nu - \mu), \phi(\mu)\rangle$ for every test function $\phi(\mu)$ which vanishes outside $(-1, +1)$ by the principal-value integral

$$\left\langle \frac{1}{\nu - \mu}, \phi(\mu) \right\rangle = P \int_{-1}^{+1} \frac{\phi(\mu)}{\nu - \mu} \, d\mu. \tag{A6}$$

[19] T. Nonnenmacher, Atomkernenergie, 12, 183 (1967).
[20] J. Mehner, Nukleonik 11, 182 (1968).

To show that the second term $\Delta_n(\mu, \nu)$ converges for $n \to \infty$ to the distribution $\Delta(\mu, \nu) = \lambda(\nu)\delta(\mu - \nu)$, we must show that

$$\lim_{n\to\infty} \langle \Delta_n(\mu, \nu), \phi(\mu)\rangle = \langle \Delta(\mu, \nu), \phi(\mu)\rangle = \langle \lambda(\nu)\delta(\mu - \nu), \phi(\mu)\rangle \tag{A7}$$

for every such test function. Representing a test function in a uniformly convergent series of Legendre polynomials,

$$\phi(\mu) = \alpha(\mu) \sum_{m=0}^{\infty} c_m P_m(\mu), \tag{A8}$$

where

$$\alpha(\mu) = 1, \quad \text{over a neighborhood of support } (-1, +1),$$
$$= 0, \quad \text{otherwise},$$

we have

$$\lim_{n\to\infty} \left\langle \Delta_n(\mu, \nu), \alpha(\mu) \sum_{m=0}^{\infty} c_m P_m(\mu) \right\rangle = \lim_{n\to\infty} (n+1) \left(h_{n+1}(\nu) \left\langle \frac{P_n(\mu)}{2(\nu - \mu)}, \alpha(\mu) \sum_{m=0}^{\infty} c_m P_m(\mu) \right\rangle - h_n(\nu) \left\langle \frac{P_{n+1}(\nu)}{2(\nu - \mu)}, \alpha(\mu) \sum_{m=0}^{\infty} c_m P_m(\mu) \right\rangle \right),$$

which, using the relation

$$\left\langle \frac{P_l(\mu)}{2(\nu - \mu)}, P_m(\mu) \right\rangle = P_m(\nu) P Q_l(\nu), \quad \text{for} \quad m \le l, \tag{A9}$$

gives

$$\lim_{n\to\infty} \langle \Delta_n(\mu, \nu), \phi(\mu)\rangle = \lim_{n\to\infty} (n+1)[h_{n+1}(\nu) P Q_n(\nu) - h_n(\nu) P Q_{n+1}(\nu)] \times \alpha(\nu) \sum_{m=0}^{\infty} c_m P_m(\nu), \tag{A10}$$

where $PQ_l(\nu)$ means the principal value of $Q_l(\nu)$ for $|\nu| < 1$. Finally, from the equivalent of relation (74) for $|\nu| < 1$, we see that

$$\lim_{n\to\infty} \langle \Delta_n(\mu, \nu), \phi(\mu)\rangle = \lambda(\nu)\phi(\nu) = \langle \lambda(\nu)\delta(\mu - \nu), \phi(\mu)\rangle.$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Gravitational Radiation: Cutting the Tail*

WILLIAM H. HALLIDY† AND ALLEN I. JANIS
University of Pittsburgh, Pittsburgh, Pennsylvania

(Received 4 March 1969)

Using the Newman–Penrose spin-coefficient approach to gravitational radiation, we consider necessary conditions for stationary–radiative–stationary transitions to occur between states in a Riemannian space–time which satisfies the empty-space Einstein field equations everywhere outside of a spatially bounded timelike cylinder. It is found that for axially symmetric radiation of finite multipole expansion such transitions cannot occur; however, a model has been constructed for nonaxially symmetric radiation for which these transitions can occur to a certain asymptotic order in a relevant parameter while satisfying the Newman–Penrose conservation laws.

1. INTRODUCTION

In discussions involving gravitational radiation the question has been considered as to whether it is possible for a gravitational system to reach a stationary state in a finite time after it has emitted a burst of radiation from an initially stationary state.[1–7] The problem of the existence of a final stationary state is closely associated with the problem of the existence of radiation tails. Radiation tails are defined as either incoming radiation or nonradiative motions. The incoming radiation with which we are here concerned is created by backscattering of the outgoing radiation, caused by interaction of the outgoing radiation with itself or with the curvature of space–time. The interaction with space–time curvature may be thought of as scattering by a dissipative medium.[6] Nonradiative motions correspond in the linear theory, and in the theory of electricity and magnetism, to imploding (exploding) sources which are not radiating.[1,4,6] An example of this would be a dipole with linear time dependence. If we think of a dipole in electromagnetic theory as being made up of equal positive and negative charges separated by a distance, then a motion having linear time dependence would correspond to the charges moving toward (away from) each other with uniform velocity.

Kundt and Newman[6] investigated the problem of backscattering for scalar and for electromagnetic waves in a flat background space–time and in a Schwarzschild background space–time. Their results showed that, although in the flat space–time background there was no backscattering, in the Schwarzschild background there was always backscattering. Couch *et al.*[4] (referred to as CTJN) have investigated the problem for general relativity by developing the multipole structure of the field through a perturbative approach, as was done to first order (the linear theory of gravitation) in an earlier paper by Janis and Newman,[8] who identified the multipoles for the linear theory by analogy with the theory of electricity and magnetism and then extended this definition to the full theory. CTJN found that an outgoing first-order quadrupole wave would interact with the first-order Schwarzschild mass to give a second-order incoming quadrupole wave. It has been conjectured[5b] that such incoming waves "will (in general) *die off* with time so the system should become asymptotically stationary." However, in addition, CTJN found that to second order the quadrupole–quadrupole self-interaction gave rise to nonradiative motions for which the time dependence did not die off asymptotically. For this sort of situation, one could not even expect the system to become asymptotically stationary. The sort of solutions whose existence we shall here investigate will be such that they have neither incoming radiation nor nonradiative motions present after the burst of radiation has passed.

Newman and Penrose[5] have discussed ten exactly conserved quantities in general relativity. They have shown that these quantities may be used to show that certain transitions between two stationary states are forbidden. For a stationary state the conserved quantities are equal to an expression $D^2 - MQ$, where D, M, and Q are related to the dipole moment, the mass (or monopole moment), and the quadrupole moment, respectively. Therefore, if for two different stationary

* Supported in part by the Aerospace Research Laboratories, Office of Aerospace Research. This work incorporates some of the results of the first author's doctoral dissertation at the University of Pittsburgh.

† Present address: California State College, California, Pennsylvania 15419.

[1] H. Bondi, M. Van der Burg, and A. Metzner, Proc. Roy. Soc. (London) **A269**, 21 (1962).

[2] E. Newman and T. Unti, J. Math. Phys. **6**, 1806 (1965).

[3] (a) R. Torrence, Ph.D. thesis, University of Pittsburgh, 1965; (b) R. Torrence and A. Janis, J. Math. Phys. **8**, 1355 (1967).

[4] E. Couch, R. Torrence, A. Janis, and E. Newman, J. Math. Phys. **9**, 484 (1968).

[5] (a) E. Newman and R. Penrose, Phys. Rev. Letters **15**, 231 (1965); (b) E. Newman and R. Penrose, Proc. Roy. Soc. (London) **A305**, 175 (1968).

[6] W. Kundt and E. Newman, J. Math. Phys. **9**, 2193 (1968).

[7] E. Couch, Ph.D. thesis, University of Pittsbugrh, 1966.

[8] A. Janis and E. Newman, J. Math. Phys. **6**, 902 (1965).

states, corresponding to different configurations of the sources, we have different values for this expression, then these two states may not be connected by a radiative transition. For a given initially stationary state which emits a burst of radiation, the expression $D^2 - MQ$ must have the same value after radiation as it did before in order to allow the possibility that the final state may be stationary. However, satisfaction of this condition is not sufficient to ensure that the final state is stationary.

In their approach CTJN assume that the burst of radiation is axially symmetric. Their calculations for quadrupole radiation then lead in second order to the conclusion that there necessarily exist tails which prohibit the system from returning to a stationary state. Using a somewhat different approach, Bonner and Rotenberg[9] obtained the result that if the system were Schwarzschild before radiation, it could not be stationary after radiation; however, their approach also assumed that the radiation was axially symmetric. Unt[10] has shown, and it will be established in this paper, that a space–time corresponding to a particular gravitational system cannot pass from a stationary initial state to a stationary final state by emitting a burst of axially symmetric radiation of finite multipole expansion. In fact, solutions which appear to represent nonradiative motions have to be present after radiation, and these solutions are of such a type that the system cannot even become asymptotically stationary. Incoming radiation may also be present; however, it does not appear possible definitely to establish its presence using the approach in this paper.

In this paper we assume that the empty-space field equations hold everywhere outside of a spatially bounded timelike cylinder, that a stationary solution to the field equations exists in an empty space–time region to the past of some initial null hypersurface, and that a pulse of outgoing gravitational radiation occurs in a space–time region to the future of this null hypersurface. We then integrate the field equations for the time development of our solution from the initial hypersurface to a null hypersurface for wichh the radiation occurred entirely in the past, and we examine the final solution thus obtained to see whether there exists a nontrivial choice for the radiation pulse for which the final state is stationary.

The formal method used in this investigation, due to Newman–Penrose,[11,12] Newman–Unti,[13] and Janis–Newman,[8] is outlined in Sec. 2. In Sec. 3 we discuss the case for which the radiation pulse has axial symmetry, and we find that in this case the system cannot have a stationary final state. In Sec. 4 we discuss the case for which the radiation pulse has no assumed axial symmetry, and we find that in this case the system may have a final state which is at least stationary to a certain asymptotic order in a relevant parameter while satisfying the Newman–Penrose conservation laws.[5] In Sec. 5 we summarize and interpret the results of Secs. 3 and 4. Finally, there follow two appendices introducing some relevant definitions and containing details of calculations.

In the following, Greek letters used as subscripts or superscripts have the range 0, 1, 2, 3 and lower case Latin letters used in the same manner have the range 2, 3. The Einstein summation convention is employed throughout for Greek and lower case Latin letters.

2. SUMMARY OF THE TETRAD FORMALISM

We introduce a family of null hypersurfaces in our space–time and associate a complex null tetrad with each point of each hypersurface. The tetrad is chosen to have the orthonormality properties

$$l^\mu n_\mu = -m^\mu \bar{m}_\mu = 1,$$

with all other scalar multiplications zero. The vectors l^μ and n^μ are real, and m^μ and its complex conjugate $\bar{m}^\mu$ may be defined as

$$m^\mu = (1/\sqrt{2})(a^\mu + ib^\mu),$$
$$\bar{m}^\mu = (1/\sqrt{2})(a^\mu - ib^\mu),$$

where a^μ and b^μ are appropriate real spacelike vectors.

We choose our coordinates and associate the tetrad with the coordinate system in the following way: Let $u = x^0$ label the null hypersurfaces and choose[14] $l_\mu = u_{,\mu}$ so that l_μ is hypersurface orthogonal and equal to a gradient. Next let $r = x^1$ be the affine parameter along null geodesics in each of these hypersurfaces and choose $l^\mu = dx^\mu/dr$. Choose $\theta = x^2$ and $\varphi = x^3$ to label a particular null geodesic in each hypersurface, and choose n^μ and m^μ to propagate parallelly along these null geodesics. Our metric now reduces to[11]

$$g^{\mu\nu} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & g^{11} & & g^{1k} \\ 0 & & & \\ & g^{1k} & & g^{kl} \\ 0 & & & \end{pmatrix},$$

[9] W. Bonnor and M. Rotenberg, Proc. Roy. Soc. (London) **A289**, 247 (1965).
[10] V. Unt, private communication.
[11] E. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[12] E. Newman and R. Penrose, J. Math. Phys. **7**, 863 (1966).
[13] E. Newman and T. Unti, J. Math. Phys. **3**, 891 (1962).

[14] A comma denotes partial differentiation; a semicolon denotes covariant differentiation.

where

$$g^{11} = 2(U - \omega\bar{\omega}), \quad g^{1k} = X^k - (\xi^k\bar{\omega} + \bar{\xi}^k\omega),$$
$$g^{kl} = -(\xi^k\bar{\xi}^l + \bar{\xi}^k\xi^l),$$

with

$$l^\mu = \delta^\mu_1, \quad m^\mu = \omega\delta^\mu_1 + \xi^k\delta^\mu_k,$$
$$n^\mu = \delta^\mu_0 + U\delta^\mu_1 + X^k\delta^\mu_k.$$

The functions ω, U, X^k, and ξ^k are defined by their appearance in these expressions for the tetrad vectors.

If we form scalars by contraction of the empty-space Riemann tensor with all possible combinations of the tetrad vectors, we obtain five independent complex quantities:

$$\begin{aligned}
\psi_0 &= -C_{\mu\nu\rho\sigma}l^\mu m^\nu l^\rho m^\sigma,\\
\psi_1 &= -C_{\mu\nu\rho\sigma}l^\mu n^\nu l^\rho m^\sigma,\\
\psi_2 &= -C_{\mu\nu\rho\sigma}\bar{m}^\mu n^\nu l^\rho m^\sigma,\\
\psi_3 &= -C_{\mu\nu\rho\sigma}\bar{m}^\mu n^\nu l^\rho n^\sigma,\\
\psi_4 &= -C_{\mu\nu\rho\sigma}\bar{m}^\mu n^\nu \bar{m}^\rho n^\sigma,
\end{aligned} \tag{2.1}$$

where $C_{\mu\nu\rho\sigma}$ is the empty-space Riemann tensor (the Weyl tensor). We may also use this tetrad to obtain a set of scalar equations (the Newman–Penrose equations) that are equivalent to the empty-space Einstein field equations.[11]

In this paper we shall be interested in space–times which have all gravitational sources contained in a finite timelike world tube: As the distance from this tube becomes large, the space–time becomes Minkowskian. In order to assure that this flatness will hold, if the distance is measured in terms of the affine parameter along null geodesics, it is sufficient to put the following conditions[11,15] on ψ_0:

$$\begin{aligned}
&\psi_0 = O(r^{-5}),\\
&\frac{\partial}{\partial r}\psi_0 = O(r^{-6}),\\
&\frac{\partial\psi_0}{\partial x^i} = O(r^{-5}), \cdots, \frac{\partial^4\psi_0}{\partial x^i\partial x^j\partial x^k\partial x^l} = O(r^{-5}),\\
&\frac{\partial}{\partial x^i}\frac{\partial\psi_0}{\partial r} = O(r^{-6}), \cdots, \frac{\partial^3}{\partial x^i\partial x^j\partial x^k}\frac{\partial\psi_0}{\partial r} = O(r^{-6}).
\end{aligned} \tag{2.2}$$

With these conditions one can then prove that[11]

$$\psi_1 = O(r^{-4}), \quad \psi_2 = O(r^{-3}), \quad \psi_3 = O(r^{-2}),$$
$$\psi_4 = O(r^{-1}).$$

These asymptotic results, which are known as the peeling-off theorem, show manifestly that the space–time becomes asymptotically Minkowskian. In this paper we adopt the stronger condition[16]

$$\psi_0 = \frac{\psi_0^0}{r^5} + \frac{\psi_1^0}{r^6} + \frac{\psi_0^2}{r^7} + O(r^{-8}),$$

with appropriate derivative conditions analogous to Eqs. (2.2). If we assume that the intersections of the null hypersurfaces with a timelike cylinder at infinity are spacelike spheres at infinity, and if we make θ and φ the usual polar coordinates, then the equations for the Weyl tensor and its time development become

$$\psi_4^0 = -\bar{\sigma}^0_{,00}, \tag{2.3a}$$
$$\psi_3^0 = \eth\bar{\sigma}^0_{,0}, \tag{2.3b}$$
$$\psi_2^0 - \bar{\psi}_2^0 = -\eth^2\bar{\sigma}^0 + \bar{\eth}^2\bar{\sigma}^0 + \bar{\sigma}^0\sigma^0_{,0} - \sigma^0\bar{\sigma}^0_{,0}, \tag{2.3c}$$
$$(\psi_2^0 + \bar{\psi}_2^0)_{,0} = -\eth\psi_3^0 - \bar{\eth}\bar{\psi}_3^0 + \sigma^0\psi_4^0 + \bar{\sigma}^0\bar{\psi}_4^0, \tag{2.3d}$$
$$\psi^0_{1,0} = -\eth\psi_2^0 + 2\sigma^0\psi_3^0, \tag{2.3e}$$
$$\psi^0_{0,0} = -\eth\psi_1^0 + 3\sigma^0\psi_2^0, \tag{2.3f}$$
$$\psi^1_{0,0} = \bar{\eth}(-\eth\psi_0^0 + 4\sigma^0\psi_1^0), \tag{2.3g}$$
$$\begin{aligned}
2\psi^2_{0,0} = &-5\psi_0^1 - \eth\bar{\eth}\psi_0^1\\
&+ 10(\psi_1^0\psi_1^0 - \psi_0^0\psi_2^0 - \tfrac{1}{2}\psi_0^0\bar{\psi}_2^0) + E,
\end{aligned} \tag{2.3h}$$

with $E = 0$ whenever $\sigma^0 = 0$, and where

$$\begin{aligned}
\psi_1 &= (\psi_1^0/r^4) + O(r^{-5}),\\
\psi_2 &= (\psi_2^0/r^3) + O(r^{-4}),\\
\psi_3 &= (\psi_3^0/r^2) + O(r^{-3}),\\
\psi_4 &= (\psi_4^0/r) + O(r^{-2}),
\end{aligned}$$

and

$$\sigma = l_{\mu;\nu}m^\mu m^\nu = (\sigma^0/r^2) + O(r^{-4})$$

is the complex shear of the null geodesics. The symbol $\eth$ stands for a differential operator which is explained in Appendix A.

To specify completely a solution to the Newman–Penrose equations we must give the following data: ψ_0 on an initial null hypersurface, σ^0 on a timelike cylinder at infinity, ψ_1^0 and $\psi_2^0 + \bar{\psi}_2^0$ on the intersection of the cylinder with the hypersurface, and (as we have already done) the geometry of their intersection. The quantity ψ_0 gives information about the quadrupole and higher moments of the sources, as well as information about the incoming gravitational radiation; σ^0 gives information about the outgoing gravitational radiation; ψ_1^0 gives information about the dipole

[15] The symbol O which appears means the following: We say that $f(r, u, x^i)$ is of order $g(r)$, which is written

$$f(r, \mu, x^i) = O\,g(r),$$

if and only if we have

$$|f(r, u, x^i)| < g(r)F(u, x^i)$$

for some $F(u, x^i)$ independent of r, and for all $r > R_0$, where R_0 is some fixed value of the affine parameter.

[16] The numerators in these expressions (and in similar expressions that follow) are independent of r.

and higher moments; and $\psi_2^0 + \bar{\psi}_2^0$ gives information about the monopole (mass) and higher moments.

In addition to the coordinate conditions that we have explicitly mentioned, certain additional conditions have been imposed in order to simplify certain equations.[13] Not all of the coordinate freedom has been used in this manner, however. The remaining transformation freedom may be studied asymptotically as has been done by Bondi *et al.*,[1] Sachs,[17] and Newman and Penrose.[12] In their studies classifying the asymptotic group they have noted that, apart from certain inversions, it may be thought of as being composed of two subgroups: the Lorentz transformations and the supertranslations. For the Lorentz transformation the leading terms of the transformations can be written as

$$\begin{aligned} u' &= K(\theta, \varphi)u + O(r^{-1}), \\ r' &= K^{-1}(\theta, \varphi)r + O(1), \\ \theta' &= f(\theta, \varphi) + O(r^{-1}), \\ \varphi' &= g(\theta, \varphi) + O(r^{-1}), \end{aligned} \tag{2.4}$$

where the leading terms on the (θ, φ) part of the transformations are just the conformal transformations of a sphere into itself with K as the conformal factor; i.e.,

$$d\theta'^2 + \sin^2\theta'\, d\varphi'^2 = K^2(d\theta^2 + \sin^2\theta\, d\varphi^2).$$

For the supertranslations the leading terms of the transformations can be written as

$$\begin{aligned} u' &= u - \alpha(\theta, \varphi) + O(r^{-1}), \\ r' &= r + O(1), \\ \theta' &= \theta + O(r^{-1}), \\ \varphi' &= \varphi + O(r^{-1}), \end{aligned} \tag{2.5}$$

where α may be written in terms of spherical harmonics as

$$\alpha = \sum_{l,m} a_{lm} Y_{lm}(\theta, \varphi).$$

The supertranslations are thus an infinite-parameter group of transformations.

Solutions to Eqs. (2.3) will be stationary if and only if there exists a coordinate system, consistent with all of the coordinate conditions, such that the partial derivative with respect to x^0, $\partial/\partial u$, of all of the ψ's, Eqs. (2.1), vanishes and such that $\sigma^0 = 0$ in this coordinate system.[5b] Therefore, given a solution to the Newman–Penrose equations which is not manifestly stationary, we need only examine the behavior of the ψ's and σ^0 under the Lorentz transformations and the supertranslations to decide whether or not this solution is, in fact, stationary.

We have now reviewed all the formalism necessary for the rest of the paper. In the next section we shall look at the problem of radiative transitions between stationary states for the case when the radiation is axially symmetric.

3. THE AXIALLY SYMMETRIC RADIATION PULSE

In Sec. 2 we mentioned that the outgoing gravitational radiation field is described by the function σ^0. If we want the radiation to be axially symmetric, then σ^0 must have the form

$$\sigma^0(u, \theta) = \sum_{l \geq 2} b_l(u)\, {}_2Y_{l0}(\theta),$$

where ${}_2Y_{l0}(\theta)$ is a spin-weight-2 spherical harmonic (defined in Appendix A) and the b_l are functions of u such that $b_{l,00}$ may be only piecewise continuous. In addition we shall impose the constraint that σ^0 have a finite multiple structure; i.e., we may write

$$\sigma^0(u, \theta) = \sum_{l=2}^{L} b_l(u)\, {}_2Y_{l0}(\theta), \tag{3.1}$$

where $b_L \not\equiv 0$. It seems likely that this restriction is sufficient (but not necessary) to insure that one can determine a unique solution of the radiation problem by giving data as described in Sec. 2. Friedlander[18] has considered the problem for scalar radiation and has found that the quantity analogous to σ^0 must satisfy coherence conditions, which are automatically satisfied by finite multipole expansions.

In Sec. 2 we found that the u dependence in ψ_2^0, when outgoing radiation is present, is determined by Eqs. (2.3a)–(2.3d) to be

$$\psi_{2,0}^0 = -\delta\psi_3^0 + \sigma^0\psi_4^0. \tag{3.2}$$

Let us first assume that σ^0 has the form of a radiation pulse that satisfies the conditions

$$\begin{aligned} \sigma^0 &\equiv 0, \quad u \leq u_1, \\ \sigma^0 &\equiv 0, \quad u \geq u_2. \end{aligned} \tag{3.3}$$

Later we shall consider the case of a radiation pulse that doesn't satisfy (3.3). We can now integrate Eq. (3.2) from u_1 to u_2. With the aid of Eqs. (2.3a) and (2.3b), we obtain the following result:

$$\psi_2^0(u_2, \theta, \varphi) = \psi_2^0(u_1, \theta, \varphi) + \int_{u1}^{u2} |\sigma_{,0}^0|^2\, du. \tag{3.4}$$

[17] R. Sachs, J. Math. Phys. **3**, 908 (1962).

[18] F. Friedlander, Proc. Roy. Soc. (London) **A269**, 53 (1962); **A279**, 386 (1964); **A299**, 264 (1967).

From Eqs. (3.2) and (2.3b) it is clear that we have $\psi^0_{2,0} \equiv 0$ if $\sigma^0 \equiv 0$, so that before and after radiation ψ^0_2 satisfies the requirement of Sec. 2 for the state to be stationary. However, we must examine all of the scalar components of the Weyl tensor to see whether they all satisfy the requirement. Of course, we *choose* the initial state to be stationary, but we must integrate the Newman–Penrose equations to find out whether the final state is also stationary. The u dependence of ψ^0_1 is given by Eq. (2.3e) as

$$\psi^0_{1,0} = -\eth\psi^0_2 + 2\sigma^0\psi^0_3.$$

If we are to have $\psi^0_{1,0} \equiv 0$ when $\sigma^0 \equiv 0$, then ψ^0_2 must have the form

$$\psi^0_2 = a_{0\,0}Y_{00}, \tag{3.5}$$

where a_0 is a constant. We choose ψ^0_2 to have this form initially, so that we have

$$\psi^0_{1,0}(u, \theta) \equiv 0, \quad u \leq u_1.$$

From Eq. (3.4) the change in ψ^0_2 is given by the expression

$$\Delta\psi^0_2 = \int_{u_1}^{u_2} |\sigma^0_{,0}|^2\, du.$$

With the initial choice (3.5) of ψ^0_2 we see that when we express $\Delta\psi^0_2$ in terms of spin-weighted spherical harmonics, all the coefficients must be zero except that of ${}_0Y_{00}$ if the final state is also to be of the form (3.5). These coefficients may be written in the form

$$Q_j \equiv \int {}_0\bar{Y}_{j0}(\Delta\psi^0_2)\, d\Omega = \int {}_0\bar{Y}_{j0}\left(\int_{u_1}^{u_2} |\sigma^0_{,0}|^2\, du\right) d\Omega,$$

where $d\Omega = \sin\theta\, d\theta\, d\varphi$ and the integration is over the spacelike sphere at infinity. Now expressing the product of spin-weighted spherical harmonics involved in $|\sigma^0_{,0}|^2$ by means of Eq. (A5) of Appendix A, we have

$$\begin{aligned} Q_j = \sum_{l,k=2}^{L} \left(\int_{u_1}^{u_2} b_{l,0}\bar{b}_{k,0}\, du\right) \left(\frac{(2l+1)(2k+1)}{4\pi(2j+1)}\right)^{\frac{1}{2}} \\ \times (k0l0 \mid klj0)(k2l-2 \mid klj0), \quad j \leq 2L, \\ Q_j = 0, \quad j > 2L. \end{aligned} \tag{3.6}$$

We now look at the case $j = 2L$. For this case the sum reduces to a single term and we have

$$\begin{aligned} Q_{2L} = \frac{2L+1}{[4\pi(4L+1)]^{\frac{1}{2}}} \left(\int_{u_1}^{u_2} |b_{L,0}|^2\, du\right) \\ \times (L0L0 \mid LL(2L)0)(L2L-2 \mid LL(2L)0). \end{aligned}$$

Since we have assumed that $b_L \not\equiv 0$, we have

$$\int_{u_1}^{u_2} |b_{L,0}|^2\, du > 0,$$

so that it is easily shown that

$$Q_{2L} > 0. \tag{3.7}$$

It therefore appears that ψ^0_1 is time dependent after radiation, but, in order to be sure that the solution is not stationary, we must examine the effects of the Lorentz transformations and the supertranslations on ψ^0_2 and σ^0 to see whether the time dependence may be removed by these transformations.

The changes in $\psi^0_2 + \bar{\psi}^0_2$ and σ^0 for the supertranslation (2.5) may be shown to be given by

$$\begin{aligned} (\psi^0_2 + \bar{\psi}^0_2)' = (\psi^0_2 + \bar{\psi}^0_2) - 2(\eth\alpha)(\eth\bar{\sigma}^0_{,0}) - 2(\bar{\eth}\alpha)(\bar{\eth}\sigma^0_{,0}) \\ - (\eth\alpha)^2\bar{\sigma}^0_{,00} - (\bar{\eth}\alpha)^2\sigma^0_{,00}, \\ \sigma^{0'} = \sigma^0 - \eth^2\alpha, \end{aligned} \tag{3.8}$$

where the left-hand sides are in terms of the primed variables. It can also easily be shown that if $\sigma^0 \equiv 0$, so that from Eq. (2.3c) $\psi^0_2 - \bar{\psi}^0_2 = 0$, then $\psi^{0'}_2 - \bar{\psi}^{0'}_2 = 0$ also. We are interested in finding a supertranslation which maps us to a coordinate system for which $\psi^0_2 = a_{0\,0}Y_{00}$ and $\sigma^0 \equiv 0$. We can see from Eq. (3.8) and the above argument that if $\sigma^0 \equiv 0$, then $\psi^{0'}_2 = \psi^0_2$, so that it is not possible to remove the term Q_{2L} from $\psi^{0'}_2$ by a supertranslation.

Under the Lorentz transformations (2.4), σ^0 transforms as

$$\sigma^{0'} = K^{-1}e^{2i\psi}\sigma^0,$$

where ψ is the angle between θ and θ'. From this we see that if $\sigma^0 \equiv 0$, then we have $\sigma^{0'} \equiv 0$ under a Lorentz transformation. If $\sigma^0 \equiv 0$, the effect of the Lorentz transformation on ψ^0_2 is

$$\psi^{0'}_2 = K^{-3}\psi^0_2. \tag{3.9}$$

The conformal factor K may be written as

$$\begin{aligned} K^{-1} = (\pi)^{\frac{1}{2}}(a\bar{a} + b\bar{b} + c\bar{c} + d\bar{d})\, {}_0Y_{00} \\ + (\tfrac{1}{3}\pi)^{\frac{1}{2}}(a\bar{a} - b\bar{b} + c\bar{c} - d\bar{d})_0Y_{10} \\ - (\tfrac{2}{3}\pi)^{\frac{1}{2}}(a\bar{b} + c\bar{d})_0Y_{11} \\ + (\tfrac{2}{3}\pi)^{\frac{1}{2}}(\bar{a}b + \bar{c}d)_0Y_{1,-1}, \end{aligned} \tag{3.10}$$

where a, b, c, and d are complex constants satisfying the restriction $ad - bc = 1$. However, we must express the right-hand side of Eq. (3.9) in the new (primed) coordinate system to obtain an interpretable expresion. To do this we make use of the well-known expression connecting the orthogonal functions for two different coordinate systems on the unit sphere:

$$Y_{lm}(\theta, \varphi) = \sum_{n=-l}^{l} {}_ma_{ln}Y_{ln}(\theta', \varphi').$$

It is clear from this that in the new coordinate system we have

$$\psi_2^{0\prime}(u', \theta', \varphi') = \sum_{\substack{l=0 \\ -l\le m\le l}}^{3} a'_{lm} Y_{lm}(\theta', \varphi').$$

Since the Lorentz transformations have unique inverses, if there is no Lorentz transformation taking us from a state $\psi_2^0 = a_{0\,0}Y_{00}$, $\sigma^0 \equiv 0$ to a state

$$\psi_2^{0\prime} = \sum_l a'_{l\,0} Y_{l0}, \quad \sigma^{0\prime} \equiv 0,$$

then there is no Lorentz transformation taking us from the state $\psi_2^{0\prime}$, $\sigma^{0\prime}$ to the state ψ_2^0, σ^0. We see from Eqs. (3.9) and (3.10) that if $\psi_2^0 = a_{0\,0}Y_{00}$, the Lorentz transformations cannot give it an angular dependence larger than $l = 3$; i.e.,

$$\psi_2^{0\prime} = a'_{0\,0}Y_{00} + \sum_m a'_{1m\,0}Y_{1m} + \sum_m a'_{2m\,0}Y_{2m} + \sum_m a'_{3m\,0}Y_{3m}. \quad (3.11)$$

However, the term Q_{2L} is the coefficient of ${}_0Y_{(2L)0}$ in the expansion of $\Delta\psi_2^0$ in terms of spin-weighted spherical harmonics. The smallest value for L is 2, for which value Eq. (3.1) becomes $\sigma^0 = b_{2\,2}Y_{20}$. This implies, from Eq. (3.7), that the coefficient of ${}_0Y_{40}$ is greater than zero. This term cannot be removed by a Lorentz transformation as can be seen from Eq. (3.11). Therefore the coefficient of ${}_0Y_{(2L)0}$ is unaffected by a Lorentz transformation for any value of L.

Since the effect of Eq. (3.7) cannot be removed by coordinate transformations, it must be regarded as real. We may therefore say that a gravitational system which was initially stationary will not be even asymptotically (in u) stationary after the emission of an axially symmetric gravitational radiation pulse that satisfies the conditions (3.1) and (3.3). In the linear theory, a solution of the type we have found would describe a nonradiative motion.[4]

We must now consider the case of a radiation pulse that doesn't satisfy the restrictions (3.3). Since we assume that the initial state is stationary, we can always choose our coordinate system such that the first of these equations holds. In such a coordinate system, however, the most general σ^0 after the pulse would satisfy $\sigma^0_{,0} = 0$. Thus we now adopt the condition

$$\sigma^0(u) \equiv 0, \quad u \le u_1,$$
$$\sigma^0_{,0}(u) \equiv 0, \quad u \ge u_2, \quad (3.12)$$

which, together with the condition that the system be stationary initially, leads to the following equation for the change in ψ_2^0:

$$\Delta\psi_2^0 = \int_{u_1}^{u_2} |\sigma^0_{,0}|^2\, du - \eth^2\bar{\sigma}^0(u_2). \quad (3.13)$$

The positive definiteness of Q_{2L}, defined by the relation

$$Q_{2L} \equiv \int {}_0\bar{Y}_{(2L)0}(\Delta\psi_2^0)\, d\Omega,$$

is not altered by the new term appearing in Eq. (3.13), as it only contributes to terms up to and including order L.

Again we must consider the possibility that Q_{2L} may be removed by a coordinate transformation; however, the situation is slightly different from the preceding because σ^0 may be initially different from zero. Arguing as before by inverse transformations from an initially stationary state, it can easily be seen that we cannot remove the Q_{2L} term from $\Delta\psi_2^0$.

We thus see that removal of the restrictions (3.3) does not alter our results.

In the above argument we have assumed that the radiation was symmetric with respect to a given axis. In the next section we completely remove this restriction on σ^0 to find out what effect this has on the problem.

4. THE NONAXIALLY SYMMETRIC RADIATION PULSE

For nonaxially symmetric outgoing radiation, σ^0 has the form

$$\sigma^0(u, \theta, \varphi) = \sum_{\substack{l=2 \\ -l\le m\le l}}^{L} b_{lm}(u)\, {}_2Y_{lm}(\theta, \varphi),$$

where the b_{lm} are functions of u such that $b_{lm,00}$ may be only piecewise continuous. As in Sec. 3, we have imposed the constraint that σ^0 have finite multipole structure.

Also as in Sec. 3, the change in ψ_2^0 is given by

$$\psi_2^0(u_2, \theta, \varphi) = \psi_2^0(u_1, \theta, \varphi) + \int_{u_1}^{u_2} |\sigma^0_{,0}|^2\, du,$$

where we have again assumed that the radiation is of the form of a pulse between u_1 and u_2 that satisfies the conditions (3.3).

The coefficient of the ${}_0Y_{jm}$ term of $\Delta\psi_2^0$ for this case may be given by

$$Q_{jm} = \int {}_0\bar{Y}_{jm}\left(\int_{u_1}^{u_2} |\sigma^0_{,0}|^2\, du\right) d\Omega. \quad (4.1)$$

Again, if $\psi^0_{1,0}$ is to vanish after radiation, ψ_2^0 must be of the form $\psi_2^0 = a_{00\,0}Y_{00}$ after radiation so that we must have

$$Q_{jm} = 0 \quad \text{unless} \quad j = m = 0.$$

We may rewrite Eq. (4.1) for Q_{jm} in terms of the Clebsch–Gordan coefficients as follows:

$$Q_{jm} = \sum_{l,k,r=2}^{L} (-1)^n \left(\int_{u_1}^{u_2} b_{lr,0} \bar{b}_{kn,0}\, du \right) \times \left(\frac{(2l+1)(2k+1)}{4\pi(2j+1)} \right)^{\frac{1}{2}} (k - nlr \mid kljm)(k2l-2 \mid klj0), \tag{4.2}$$

with $r - m = n$.

In order to simplify Eq. (4.2), we will assume that σ^0 has the form of a simple pole; i.e.,

$$\sigma^0 = \sum_{-l \le m \le l} b_{lm}(u)\, {}_2Y_{lm}(\theta, \varphi),$$

where l is fixed and $l \geq 2$. For this case Eq. (4.2) becomes

$$Q_{jm} = \sum_r (-1)^n \left(\int_{u_1}^{u_2} b_{lr,0} \bar{b}_{ln,0}\, du \right) \frac{2l+1}{[4\pi(2j+1)]^{\frac{1}{2}}} \times (l - nlr \mid lljm)(l2l-2 \mid llj0). \tag{4.3}$$

We see immediately that we have $Q_{jm} = 0$ for all $j > 2l$ as a consequence of Eq. (A7) of Appendix A. First we look at the equations $Q_{j0} = 0$, for these are just the generalization to the nonaxially symmetric case of the equations $Q_j = 0$ of Sec. 3. A sufficient condition for $Q_{j0} = 0$ for all $j = 1, 2, \cdots, 2l$ will be shown below. First we shall develop a useful identity.

As a consequence of the fact that the eigenvectors $|l_1 m_1 l_2 m_2)$ form a basis on the vector space for addition of angular momentum in quantum mechanics, we have the following elementary identity[19]:

$$|l_1 l_2 jm) = \sum_{m_1} |l_1 m_1 l_2 m_2)(l_1 m_1 l_2 m_2 \mid l_1 l_2 jm),$$

where $m_2 = m - m_1$, $|l_1 - l_2| \leq j \leq l_1 + l_2$, and all vectors are normalized. The coefficients of $|l_1 m_1 l_2 m_2)$ in the expansion are just the Clebsch–Gordan coefficients. The eigenvectors $|l_1 l_2 jm)$ are orthonormal; therefore we have

$$(l_1 l_2 kn \mid l_1 l_2 jm) = \delta_{jk}\delta_{nm}.$$

Thus we must have

$$\sum_{m_1} (l_1 m_1 l_2 m_2 \mid l_1 l_2 kn)(l_1 m_1 l_2 m_2 \mid lljm) = \delta_{jk}\delta_{nm},$$

where we have chosen the Clebsch–Gordan coefficients to be real. A special case of this relation is

$$\sum_m (l - m\, lm \mid llj0)(l - m\, lm \mid ll00) = \delta_{j0}.$$

But $(l - m\, lm \mid ll00)$ is just given by

$$(l - m\, lm \mid ll00) = (-1)^{l+m}(2l+1)^{\frac{1}{2}}.$$

[19] See, for example, E. Merzbacher, *Quantum Mechanics* (John Wiley & Sons, Inc., New York, 1961), pp. 416 *et seq.*

Therefore we have

$$\sum_m (-1)^m (l - m\, lm \mid ll00) = 0, \quad j \neq 0.$$

Now we see that a sufficient condition for $Q_{j0} = 0$ for all $j = 1, \cdots, 2l$ can be written as

$$\int_{u_1}^{u_2} |b_{lm,0}|^2\, du = \int_{u_1}^{u_2} |b_{ln,0}|^2\, du, \quad \text{for all } m \text{ and } n. \tag{4.4}$$

The remaining equations for $Q_{jm} = 0$, $m \neq 0$, are trivially satisfied if we choose

$$\int_{u_1}^{u_2} b_{lm,0} \bar{b}_{ln,0}\, du = 0, \quad m \neq n. \tag{4.5}$$

We shall now construct a model for the b_{lm} which satisfies Eqs. (4.4) and (4.5).

Consider any complex function $f(u)$ such that $f(u) \in C'$ and $f(u)$ has compact support contained in $[a, b]$ with $0 < a < b$. We then define our b_{lm} to be

$$b_{lm}(u) = f[u - (l+m)b].$$

Letting $u_1 = 0$ and $u_2 = (2l+2)b$, we see by inspection that these b_{lm} satisfy Eqs. (4.4) and (4.5). Therefore we have $\psi_2^0 = a'_{00\,0} Y_{00}$ after radiation.

Satisfying the condition that $\psi_2^0 = a'_{00\,0} Y_{00}$ after radiation is not enough to ensure that the state is stationary. We must examine all the scalar components of the Weyl tensor to see whether or not they are all independent of u afterward while satisfying the Newman–Penrose conservation laws.[5] If we choose the initial state of our system to be that of the Schwarzschild solution, i.e.,

$$\psi_0 = \psi_1^0 = \psi_3^0 = \psi_4^0 = 0$$

and

$$\psi_2^0 = a_{00\,0} Y_{00} \quad \text{at} \quad u = u_1,$$

and demand that our final state as far as we can examine it have this form also, i.e.,

$$\psi_4^0 = \psi_3^0 = \psi_1^0 = \psi_0^0 = \psi_0^1 = 0$$

and

$$\psi_2^0 = a'_{00\,0} Y_{00} \quad \text{at} \quad u = u_2, \tag{4.6}$$

then we will have trivially satisfied the Newman–Penrose conservation laws and have assured that, to the highest asymptotic order for which we have been able to examine the solutions, they appear stationary (provided, of course, we can find a model for the b_{lm} satisfying the resulting conditions). We note from Eqs. (2.3) that these conditions make $\psi_{0,0}^2 = 0$; this is as far as we have been able to carry out the calulations in our asymptotic expansion.

A sufficient set of conditions on the b_{lm} which will satisfy the situation outlined above are written out in Appendix B and are classified there into three distinct types: overlap, profile likeness, and symmetry. If we allow α_{lm}, β_{lm}, and γ_{lm} to represent, generically, any

b_{lm} or its derivatives which occur in the equations, the first two types of conditions may be written in the form

$$\int_{u_1}^{u_2}\cdots\int_{u_1}^{u'}\alpha_{lm}\beta_{ln}\,du\cdots du''=0,\quad m\neq n,$$

for the overlap condition and, for the profile likeness condition,

$$\int_{u_1}^{u_2}\cdots\int_{u_1}^{u'}\gamma_{lm}\,du\cdots du''=\int_{u_1}^{u_2}\cdots\int_{u_1}^{u'}\gamma_{ln}\,du\cdots du''$$

for all m, n.

The choice of b_{lm} made above to satisfy $\psi_2^0=a'_{00}\,{}_0Y_{00}$ after radiation also automatically satisfies these conditions. The remaining conditions to be satisfied, the symmetry conditions, make statements about each b_{lm} individually. These are the most difficult conditions to satisfy; in order to do so we shall need to place additional restrictions on σ^0. The conditions already imposed on σ^0 will show that if we can satisfy the symmetry conditions for a particular b_{lm}, then they are automatically satisfied for all the other b_{ln} occurring in a given σ^0. If we find a function b_{lm} satisfying Eq. (B3f) of Appendix B for any $l\geq 2$, this function satisfies Eq. (B3g) of Appendix B as a special case if $l=2$; i.e., the integrals on both the right-hand side and the left-hand side are zero. Equation (B3d) only needs to be satisfied if $l>2$; however, it may also be satisfied for $l=2$. Therefore, in the following we shall always satisfy Eqs. (B3d) and (B3f) for $l\geq 2$. In order to find a function satisfying the conditions implied by Eqs. (B3), we shall first apply another condition to b_{lm} which will allow us to simplify the form of these equations somewhat. Let us write $b_{lm}=R_{lm}\exp(i\theta_{lm})$, where R_{lm} and θ_{lm} are real. Our new condition will be that $\theta_{lm,0}=0$. Now we can rewrite Eqs. (B3); and since we refer to a particular, fixed, though arbitrary, value for l and m, we drop the subscripts. Then, after performing certain of the indicated integrations, Eqs. (B3) become

$$\int_{u_1}^{u_2}R\,du=0, \tag{4.7a}$$

$$\int_{u_1}^{u_2}\int_{u_1}^{u}R\,du\,du'=0, \tag{4.7b}$$

$$\int_{u_1}^{u_2}\int_{u_1}^{u}\int_{u_1}^{u'}R\,du\,du'\,du''=0, \tag{4.7c}$$

$$\int_{u_1}^{u_2}R\int_{u_1}^{u}(R_{,0})^2\,du\,du'=0, \tag{4.7d}$$

$$\int_{u_1}^{u_2}\int_{u_1}^{u}R\int_{u_1}^{u'}(R_{,0})^2\,du\,du'\,du''=0, \tag{4.7e}$$

$$\int_{u_1}^{u_2}R\int_{u_1}^{u}\int_{u_1}^{u'}(R_{,0})^2\,du\,du'\,du''=0, \tag{4.7f}$$

$$\int_{u_1}^{u_2}R^3\,du=0. \tag{4.7g}$$

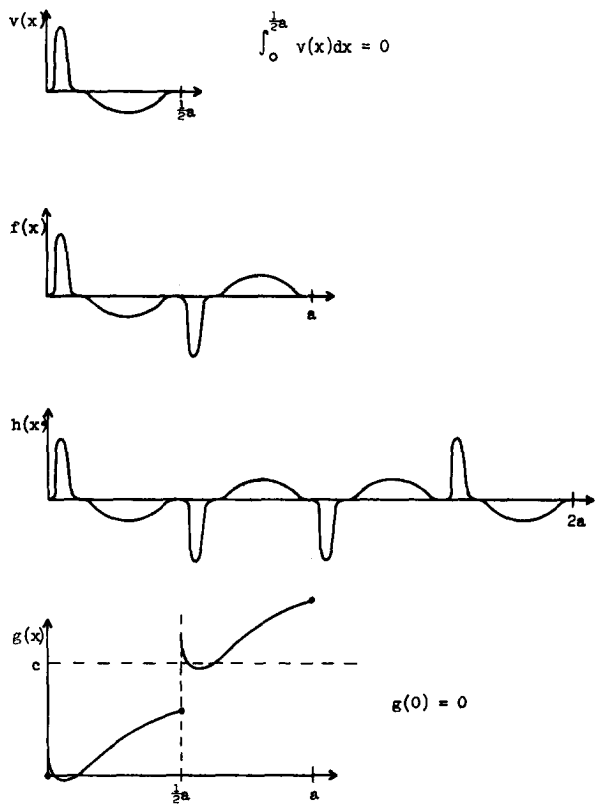


FIG. 1. Examples of functions $v(x)$, $f(x)$, $h(x)$, and $g(x)$.

We shall now construct a function R which satisfies these conditions. (See Fig. 1 for an example of the function to be constructed.) Consider a real-valued differentiable function $v(x)$ belonging to class C', having support contained in the interval $[0, \frac{1}{2}a]$, and such that the function and its first derivative approach zero as their arguments approach 0 or $\frac{1}{2}a$. Also, assume that the integral of this function over $[0, \frac{1}{2}a]$ vanishes. Now choose a real-valued function $g(x)$ such that

$$g(0)=0,$$

$$g(x+\tfrac{1}{2}a)=g(x)+c,\quad 0<x\leq\tfrac{1}{2}a,$$

with c a constant. Define a new function $f(x)$ such that

$$f(x)=v(x),\qquad 0\leq x\leq\tfrac{1}{2}a,$$

$$f(x)=-v(x-\tfrac{1}{2}a),\quad \tfrac{1}{2}a\leq x\leq a,$$

so that the support of $f(x)$ is contained in the interval $[0, a]$. The function $f(x)$ then satisfies the following conditions:

$$\int_0^a f(x)\,dx=0,$$

$$\int_0^a\int_0^x f(x')\,dx\,dx'=0,$$

$$\int_0^a[f(x)]^3\,dx=0,$$

$$\int_0^a f(x)g(x)\,dx=0.$$

We now construct a new function $h(x)$ such that

$$h(x) = f(x), \qquad 0 \leq x \leq a,$$
$$h(x) = -f(x-a), \quad a \leq x \leq 2a,$$

so that $h(x)$ has support contained in the interval $[0, 2a]$. It is immediate that $h(x)$ satisfies the following conditions:

$$\int_0^{2a} h(x)\,dx = 0,$$
$$\int_0^{2a}\int_0^{x} h(x')\,dx\,dx' = 0, \tag{4.8}$$
$$\int_0^{2a} h(x)^3\,dx = 0.$$

We also assert that we have

$$\int_0^{2a}\int_0^{x}\int_0^{x'} h(x'')\,dx\,dx'\,dx'' = 0,$$
$$\int_0^{2a}\int_0^{x} h(x')g(\tfrac{1}{2}x')\,dx\,dx' = 0, \tag{4.9}$$
$$\int_0^{2a} h(x)g(\tfrac{1}{2}x)\,dx = 0.$$

In addition, our function $h(x)$ satisfies the identity

$$\int_0^{2a} h(x)\int_0^{x} g(\tfrac{1}{2}x')\,dx\,dx' = 0. \tag{4.10}$$

It can easily be seen that the choice of $g(\frac{1}{2}x)$ as

$$g(\tfrac{1}{2}x) = \int_0^{x}\left(\frac{dh}{dx}\bigg|_{x=x'}\right)^2 dx'$$

is compatible with the definition of $g(x)$. Therefore, if we now choose

$$h(x) = R(x),$$
$$g(\tfrac{1}{2}x) = \int_0^{x}\left(\frac{dR(x')}{dx}\right)^2 dx',$$

we see that Eqs. (4.8), (4.9), and (4.10) are equivalent to Eqs. (4.7) with $[u_1, u_2]$ appropriately chosen to include $[0, 2a]$ such that all of the b_{lm} may be included with zero overlap in the interval (u_1, u_2). It follows that we can satisfy Eqs. (4.6) with the choice of σ^0 described above. While we cannot conclude from this that the state of the system is stationary after radiation, it is at least not incompatible with the Newman–Penrose equations as far as they have been written asymptotically to say that a gravitational system might reach a stationary state from an initially stationary state after the emission of a pulse of radiation.

5. CONCLUSION

In Sec. 3 we proved that a gravitational system cannot make a transition between stationary states by emission of a burst of axially symmetric radiation of finite multipole expansion. There must always be a tail after such a radiation pulse, and in fact the solution afterwards is not even asymptotically stationary. Part of this solution may in fact represent a nonradiative motion corresponding to an explosion or implosion of the source after radiation. The proof of this theorem hinged on the fact that ψ_2^0 had to be spherically symmetric (i.e., $\psi_2^0 = a_{00\,0}Y_{00}$) for a state to be stationary. It was found that, for σ^0 axially symmetric, if ψ_2^0 was spherically symmetric before radiation, it could not be made spherically symmetric afterwards.

In Sec. 4 the restriction that σ^0 be axially symmetric was dropped. It was then found that if σ^0 had the form of a simple pole, a model of σ^0 could be found such that, if ψ_2^0 was spherically symmetric before radiation, it could be made spherically symmetric afterwards. It was also shown, using the asymptotic approach, that a model could be found for σ^0 for which ψ_1^0, ψ_0^0, and ψ_0^1, assumed zero before radiation, could be made zero afterwards. The conserved quantities of Newman–Penrose were trivially satisfied by this model. The final state solution was therefore time independent through order $1/r^7$ in the tetrad-formalism variables.

It therefore seems that, although it is not possible for a gravitational system to return to a stationary state after emitting a burst of axially symmetric gravitational radiation of finite multipole expansion, the question of whether or not it can return to such a state is still open if the radiation is not assumed to have axial symmetry or if it has an infinite multipole expansion. Should the asymptotic model we have constructed for return to a stationary state prove to be extendable to an exact stationary solution, it would be the first time (to our knowledge) that dropping the restriction of axial symmetry leads to significantly different physical results in general relativity.

Note Added in Proof: Results similar to some of the results of this paper have been independently obtained by Papapetrov, Ann. Inst. H. Poincaré **11A**, 57 (1969).

ACKNOWLEDGMENTS

The authors would like to thank Professor E. T. Newman for several valuable suggestions and Dr. D. Pajerski for many helpful discussions. We also wish to express our gratitude to Dr. W. Kinnersley for pointing out errors in our original presentation.

APPENDIX A

In Sec. 2 we introduced a differential operator $\eth$ which we shall define below. Consider a spherical surface having a pair of orthonormal tangent vectors defined on it at all points except the poles. We denote these vectors at a given point as a^μ and b^μ. We then make two complex vectors out of these tangent vectors by requiring that

$$m^\mu = (1/\sqrt{2})(a^\mu + ib^\mu)$$

and

$$\bar{m}^\mu = (1/\sqrt{2})(a^\mu - ib^\mu).$$

We assume that the vector field defined by picking a^μ and b^μ at each point is sufficiently differentiable.[20] If we now choose a second vector field, i.e., if we pick new orthonormal tangent vectors at each point $\tilde{a}^\mu$ and $\tilde{b}^\mu$, then the new complex vectors will be related to the old ones by a rotation

$$\tilde{m}^\mu = m^\mu e^{if(\theta,\varphi)},$$

where $f(\theta, \varphi)$ is the angle between the vectors a^μ and $\tilde{a}^\mu$ at the point (θ, φ).

We may in general define functions on the sphere which transform according to the rule

$$\tilde{\eta} = \eta e^{isf(\theta,\varphi)}, \quad s \text{ integral}, \tag{A1}$$

under the above transformation on m^μ. For example, such a function might be given by the expression

$$T_{\mu_1\cdots\mu_s} m^{\mu_1}\cdots^{\mu_s},$$

where $T_{\mu_1\cdots\mu_s}$ is any constant convariant tensor of rank s. Functions which transform according to this prescription will be said to have spin weight s. It follows from Eq. (A1) that if η has spin weight s, then $\tilde{\eta}$ has spin weight $-s$.

Our sphere will be taken to be the sphere at infinity as mentioned in Sec. 2. Although m^μ and $\bar{m}^\mu$ as defined in our null tetrad are not asymptotically tangent to this sphere, we may write all our equations in terms of $\eth$ if we assign to the following tetrad-formalism variables the spin weights they would have if m^μ and $\bar{m}^\mu$ were tangent to the sphere at infinity:

Variables	Spin weight
ψ_4^0	-2
ψ_3^0	-1
ψ_2^0	0
ψ_1^0	1
$\sigma^0, \psi_0^0, \psi_0^1, \psi_0^2$	2

Other tetrad-formalism variables may have spin weight assigned to them, but the above variables are the ones we are concerned with here.

Now we are ready to define $\eth$. Suppose η is a function defined on a sphere and having spin weight s according to Eq. (A1). We consider $\eth\eta$ to be a function on the sphere having spin weight $s + 1$ and $\bar{\eth}\eta$ to be a function on the sphere having spin weight $s - 1$. We choose polar coordinates (θ, φ) on the sphere and choose a^μ tangent to the curves $\varphi = \text{const}$ and b^μ tangent to the curves $\theta = \text{const}$, where

$$m^\mu = (1/\sqrt{2})(a^\mu + ib^\mu),$$

and define

$$\eth\eta \equiv -\frac{1}{\sqrt{2}}(\sin\theta)^s\left(\frac{\partial}{\partial\theta} + \frac{i}{\sin\theta}\frac{\partial}{\partial\varphi}\right)[(\sin\theta)^{-s}\eta] \tag{A2}$$

and

$$\bar{\eth}\eta = -\frac{1}{\sqrt{2}}(\sin\theta)^{-s}\left(\frac{\partial}{\partial\theta} - \frac{i}{\sin\theta}\frac{\partial}{\partial\varphi}\right)[(\sin\theta)^{s}\eta].$$

We note that the commutation properties of $\eth$ and $\bar{\eth}$. are $(\bar{\eth}\eth - \eth\bar{\eth})\eta = s\eta$ if η has spin weight s. This operator can now be used to define spin-weighted spherical harmonics in terms of ordinary spherical harmonics[12,21] as

$$\begin{aligned}
{}_sY_{lm}(\theta,\varphi) &= (2)^{\frac{1}{2}s}\left[\frac{(l-s)!}{(l+s)!}\right]^{\frac{1}{2}}\eth^s Y_{lm}(\theta,\varphi), \quad 0 \le s \le l,\\
&= (-1)^{-s}(2)^{-\frac{1}{2}s}\left[\frac{(l+s)!}{(l-s)!}\right]^{\frac{1}{2}}\eth^{-s}Y_{lm}(\theta,\varphi), \quad -l \le s \le 0,\\
{}_sY_{lm}(\theta,\varphi) &= 0, \quad |s| > l,\\
{}_0Y_{lm} &\equiv Y_{lm}.
\end{aligned} \tag{A3}$$

These spin-weight-s spherical harmonics have the additional properties

$$\begin{aligned}
{}_s\bar{Y}_{lm} &= (-1)^{m+s}{}_{-s}Y_{l,-m},\\
\bar{\eth}\,{}_sY_{lm} &= \tfrac{1}{2}(\sqrt{2})[(l-s)(l+s+1)]^{\frac{1}{2}}{}_{s+1}Y_{lm},\\
\bar{\eth}\,{}_sY_{lm} &= -\tfrac{1}{2}(\sqrt{2})[(l+s)(l-s+1)]^{\frac{1}{2}}{}_{s-1}Y_{1m},\\
\bar{\eth}\eth\,{}_sY_{lm} &= -\tfrac{1}{2}(l-s)(l+s+1)\,{}_sY_{lm},\\
\int {}_s\bar{Y}_{lm}\,{}_sY_{kn}\,d\Omega &= \delta_{lk}\delta_{mn},
\end{aligned} \tag{A4}$$

[20] See, for example, L. Eisenhart, *Riemannian Geometry* (Princeton University Press, Princeton, N.J., 1960).

[21] J. Goldberg, A. J. Macfarlane, E. T. Newman, F. Rohrlich, and E. C. G. Sudarshan, J. Math. Phys. 8, 2155 (1967).

where $d\Omega = \sin\theta\, d\theta\, d\varphi$ and the integral is taken over the entire surface of the sphere.

A function $\eta(\theta, \varphi)$, having spin weight s, may be written in terms of the spin-weight s spherical harmonics which form a complete orthonormal set for each value of s as

$$\eta(\theta, \varphi) = \sum_{\substack{l \geq |s| \\ m \leq |l|}} \eta_{lm}\, {}_sY_{lm}(\theta, \varphi).$$

The expression of the product of two spherical harmonics of spin weights s and t was given by Exton[22] as

$$\begin{aligned} {}_sY_{lm}\, {}_tY_{kn} = \sum_{j=|l-k|}^{l+k} & \left[\frac{(2l+1)(2k+1)}{4\pi(2j+1)}\right]^{\frac{1}{2}} \\ & \times (knlm \mid k\, lj, m+n) \\ & \times (k-t\, l-s \mid k\, lj, -s-t)\, {}_{s+t}Y_{j,(m+n)}, \end{aligned} \tag{A5}$$

where $(k\beta l\alpha \mid k\, lj, \alpha+\beta)$ are the real Clebsch–Gordan coefficients given by[23]

$$\begin{aligned} & (k\beta l\alpha \mid k\, lj, \alpha+\beta) \\ & \quad + \left[\frac{(l+k-j)!\,(j+k-l)!\,(j+l-k)!\,(2j+1)}{(j+l+k+1)!}\right. \\ & \quad \times (k+\beta)!\,(k-\beta)!\,(l+\alpha)!\,(l-\alpha)! \\ & \quad \left.\times (j+\alpha+\beta)!\,(j-\alpha-\beta)!\right]^{\frac{1}{2}} \\ & \quad \times \sum_p (-)^p [p!\,(k+l-j-p)! \\ & \quad \times (k-\beta-p)!\,(l+\alpha-p)! \\ & \quad \times (j-l+\beta+p)!\,(j-\beta-\alpha+p)!]^{-1}, \end{aligned} \tag{A6}$$

provided that the following set of equations is satisfied:

$$\begin{aligned} l+k-j &\geq 0, \\ l+j-k &\geq 0, \\ k+j-l &\geq 0. \end{aligned} \tag{A7}$$

The Clebsch–Gordan coefficients are automatically zero if they do not satisfy Eqs. (A7). In Eq. (A5) we have summed over those values of p for which the contents of all the factorials are greater than or equal to zero.

APPENDIX B

The conditions given below are sufficient to guarantee that Eqs. (4.6) are satisfied.[24] The conditions have been grouped, for convenience, into three general categories: overlap, profile likeness, and symmetry.

Overlap conditions: These conditions must hold for each $l \geq 2$ and for each m, n, and r such that either $m \neq n$ or m, n, and r are not all equal, whichever applies:

$$\begin{aligned}
&\int_{u_1}^{u_2} b_{lm,0}\bar{b}_{ln,0}\, du = 0, && \int_{u_1}^{u_2}\int_{u_1}^{u}\int_{u_1}^{u'} b_{lm}\bar{b}_{ln,0}\, du\, du'\, du'' = 0, \\
&\int_{u_1}^{u_2} b_{lm}\bar{b}_{ln,0}\, du = 0, && \int_{u_1}^{u_2}\int_{u_1}^{u}\int_{u_1}^{u'}\int_{u_1}^{u''} b_{lm,0}\bar{b}_{ln,0}\, du\, du'\, du''\, du''' = 0, \\
&\int_{u_1}^{u_2}\int_{u_1}^{u} b_{lm,0}\bar{b}_{ln,0}\, du\, du' = 0, && \int_{u_1}^{u_2}\int_{u_1}^{u} b_{lm}\bar{b}_{ln}\, du\, du' = 0, \\
&\int_{u_1}^{u_2}\int_{u_1}^{u} b_{lm}\bar{b}_{ln,0}\, du\, du' = 0, && \int_{u_1}^{u_2} b_{lm}\int_{u_1}^{u} \bar{b}_{ln}\, du\, du' = 0, \\
&\int_{u_1}^{u_2}\int_{u_1}^{u}\int_{u_1}^{u'} b_{lm,0}\bar{b}_{ln,0}\, du\, du'\, du'' = 0, && \int_{u_1}^{u_2}\int_{u_1}^{u} b_{lm}b_{lr}\bar{b}_{ln,0}\, du\, du' = 0, \\
&\int_{u_1}^{u_2} b_{lm}\bar{b}_{ln}\, du = 0, && \int_{u_1}^{u_2}\int_{u_1}^{u} b_{lm}\int_{u_1}^{u'} b_{lr,0}\bar{b}_{ln,0}\, du\, du'\, du'' = 0, \\
&\int_{u_1}^{u_2} b_{lm}\int_{u_1}^{u} b_{lr,0}\bar{b}_{ln,0}\, du\, du' = 0, && \int_{u_1}^{u_2} b_{lr}\int_{u_1}^{u} b_{lm}\bar{b}_{ln,0}\, du\, du' = 0, \\
&\int_{u_1}^{u_2} b_{lm}b_{lr}\bar{b}_{ln,0}\, du = 0, && \int_{u_1}^{u_2} b_{lr}\int_{u_1}^{u}\int_{u_1}^{u'} b_{lm,0}\bar{b}_{ln,0}\, du\, du'\, du'' = 0.
\end{aligned} \tag{B1}$$

[22] A. Exton, Ph.D. thesis, University of Pittsburgh, 1967.

[23] (a) Institute of Atomic Energy, Academia Sinica, Eds., *Tables of the Clebsch–Gordan Coefficients* (Science Press, Peking, 1965); (b) M. Abramowitz and I. Stegun, Eds., *Handbook of Mathematical Functions* (Dover Publications, Inc., New York, 1965).

[24] W. H. Hallidy, Ph.D. thesis, University of Pittsburgh, 1968.

Profile-likeness conditions: These conditions must hold for each $l \geq 2$ and for all m and n:

$$\int_{u_1}^{u_2} |b_{lm,0}|^2\, du = \int_{u_1}^{u_2} |b_{ln,0}|^2\, du,$$

$$\int_{u_1}^{u_2} b_{lm}\bar{b}_{lm,0}\, du = \int_{u_1}^{u_2} b_{ln}\bar{b}_{ln,0}\, du,$$

$$\int_{u_1}^{u_2}\int_{u_1}^{u} |b_{lm,0}|^2\, du\, du' = \int_{u_1}^{u_2}\int_{u_1}^{u} |b_{ln,0}|^2\, du\, du',$$

$$\int_{u_1}^{u_2}\int_{u_1}^{u} b_{lm}\bar{b}_{lm,0}\, du\, du' = \int_{u_1}^{u_2}\int_{u_1}^{u} b_{ln}\bar{b}_{ln,0}\, du\, du',$$

$$\int_{u_1}^{u_2}\int_{u_1}^{u}\int_{u_1}^{u'} |b_{lm,0}|^2\, du\, du'\, du'' = \int_{u_1}^{u_2}\int_{u_1}^{u}\int_{u_1}^{u'} |b_{ln,0}|^2\, du\, du'\, du'',$$

$$\int_{u_1}^{u_2} |b_{lm}|^2\, du = \int_{u_1}^{u_2} |b_{ln}|^2\, du, \tag{B2}$$

$$\int_{u_1}^{u_2}\int_{u_1}^{u}\int_{u_1}^{u'} b_{lm}\bar{b}_{lm,0}\, du\, du'\, du'' = \int_{u_1}^{u_2}\int_{u_1}^{u}\int_{u_1}^{u'} b_{ln}\bar{b}_{ln,0}\, du\, du'\, du'',$$

$$\int_{u_1}^{u_2}\int_{u_1}^{u}\int_{u_1}^{u'}\int_{u_1}^{u''} |b_{lm,0}|^2\, du\, du'\, du''\, du''' = \int_{u_1}^{u_2}\int_{u_1}^{u}\int_{u_1}^{u'}\int_{u_1}^{u''} |b_{ln,0}|^2\, du\, du'\, du''\, du''',$$

$$\int_{u_1}^{u_2}\int_{u_1}^{u} |b_{lm}|^2\, du\, du' = \int_{u_1}^{u_2}\int_{u_1}^{u} |b_{ln}|^2\, du\, du',$$

$$\int_{u_1}^{u_2} b_{lm}\int_{u_1}^{u} \bar{b}_{lm}\, du\, du' = \int_{u_1}^{u_2} b_{ln}\int_{u_1}^{u} \bar{b}_{ln}\, du\, du'.$$

Symmetry conditions: These conditions must hold for each $l \geq 2$ unless explicitly noted otherwise:

$$\int_{u_1}^{u_2} b_{lm}\, du = 0, \tag{B3a}$$

$$\int_{u_1}^{u_2}\int_{u_1}^{u} b_{lm}\, du\, du' = 0, \tag{B3b}$$

$$\int_{u_1}^{u_2} b_{lm}\int_{u_1}^{u} |b_{lm,0}|^2\, du\, du' = \int_{u_1}^{u_2} b_{lm}^2\bar{b}_{lm,0}\, du, \tag{B3c}$$

$$\int_{u_1}^{u_2}\int_{u_1}^{u}\int_{u_1}^{u'} b_{lm}\, du\, du'\, du'' = 0, \quad l > 2, \tag{B3d}$$

$$\int_{u_1}^{u_2}\int_{u_1}^{u} b_{lm}^2\bar{b}_{lm,0}\, du\, du' = \int_{u_1}^{u_2}\int_{u_1}^{u} b_{lm}\int_{u_1}^{u'} |b_{lm,0}|^2\, du\, du'\, du'', \tag{B3e}$$

$$\int_{u_1}^{u_2} b_{lm}\int_{u_1}^{u} b_{lm}\bar{b}_{lm,0}\, du\, du' = 0, \quad l > 2,$$

$$\int_{u_1}^{u_2} b_{lm}\int_{u_1}^{u}\int_{u_1}^{u'} |b_{lm,0}|^2\, du\, du'\, du'' = 0, \quad l > 2, \tag{B3f}$$

$$3\int_{u_1}^{u_2} b_{2m}\int_{u_1}^{u} b_{2m}\bar{b}_{2m,0}\, du\, du' = \int_{u_1}^{u_2} b_{2m}\int_{u_1}^{u}\int_{u_1}^{u'} |b_{2m,0}|^2\, du\, du'\, du''. \tag{B3g}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Alternative Method for Deriving the Fock Currents

R. H. Ott
Environmental Science Services Administration/Institute for Telecommunication Sciences, Boulder, Colorado 80302

(Received 13 May 1969)

Fock's principle is extended to apply to the fields near the shadow boundary on a parabolic cylinder, illuminated by a plane wave propagating at an arbitrary angle but normal to the cylinder axis. By solving the wave equation in parabolic coordinates, with an impedance boundary condition at the surface and a radiation condition at infinity, we find that the fields thus obtained agree very well with those predicted by Fock's principle, as long as the observation point is near where the wave propagation vector is tangent to the cylinder. When the propagation vector is tangent to the cylinder's apex, Fock's principle gives the exact currents everywhere on the cylinder. These results are more general than those of Jones, which are limited to perfectly conducting cylinders.

I. INTRODUCTION

A number of investigators have studied the problem of diffraction of radio waves by cylindrical (parabolic and circular) surfaces. Rice[1] derived the penumbral currents for a plane wave striking a perfectly conducting parabolic cylinder when the direction of the incident wave is tangent to the apex of the cylinder and normal to the axis of the cylinder. Fock[2] derived the penumbral currents for a plane wave illuminating a finitely conducting parabolic cylinder when the plane wave propagates in a direction tangent to the apex of the parabolic cylinder. Jones[3] derived the penumbral currents for a plane wave striking a perfectly conducting parabolic cylinder at an arbitrary angle of incidence. Wait and Conda[4] derived the penumbral currents for a finitely conducting circular cylinder. Ivanov[5] derived the penumbral currents for a plane wave incident on a perfectly conducting cylinder, so that the direction of the incident wave makes an angle ψ with the axis of the parabolic cylinder but is normal to the generatrix of the cylinder.

If we could establish the conditions under which Fock's assumption is valid, that is, that all bodies with a smoothly varying curvature have the same generic current distribution in the penumbra, we could obtain a universal formula for the surface field in terms of the field of a plane wave tangent to the apex. When the wave is not tangent to the apex, we would replace the original wave direction and cylinder geometry with a wave, tangent to the apex of a replacement cylinder, that has a radius of curvature matching that of the original cylinder at the point of tangency.

To this end, we derive the currents induced by a plane wave striking a parabolic cylinder, normal to its axis but at an arbitrary angle of incidence (not necessarily tangent to its apex). We assume that the field on the surface of the cylinder satisfies an impedance boundary condition and that the effects of coupling between E_z and H_z are negligible.[6] The results show that Fock's principle yields extremely accurate penumbral currents, for *all* angles of incidence, provided that the observation point on the cylinder is near where the incident field grazes the cylinder. When the propagation vector is tangent to the cylinder's apex, Fock's principle yields surface currents identical to those derived by solution of the wave equation.

2. THE PARABOLIC CYLINDER

We now consider the field on the surface of a parabolic cylinder produced by a plane wave traveling in a direction making an angle φ_0 with the positive x axis (cf. Fig. 1). We will assume that $0 < \varphi_0 < \pi$. The parabolic coordinates are defined by

$$x = \tfrac{1}{2}(\xi^2 - \eta^2),$$
$$y = \xi\eta, \tag{1}$$

where $\xi = \xi_0$ defines the particular cylinder in Fig. 1.

We seek a solution of the wave equation in parabolic cylinder coordinates

$$\frac{1}{(\xi^2+\eta^2)}\left(\frac{\partial^2\psi}{\partial\xi^2}+\frac{\partial^2\psi}{\partial\eta^2}\right)+k^2\psi=0, \tag{2}$$

together with the boundary condition

$$\frac{\partial\psi}{\partial n}+\frac{z}{(\xi_0^2+\eta^2)^{\frac{1}{2}}}\psi=0, \quad \text{at } \xi=\xi_0, \tag{3}$$

and the radiation condition at infinity.

[1] S. O. Rice, Bell System Tech. J. 33. 417 (1953)
[2] V. A. Fock, *International Series of Monographs on Electromagnetic Waves* (Pergamon Press, New York, 1965), Chaps. 1,2, and 5.
[3] D. S. Jones, *The Theory of Electromagnetism* (The Macmillan Co., New York, 1964), pp. 467–478.
[4] J. R. Wait and A. M. Conda, IRE Trans. Antennas Propagation 6, 348 (1958).
[5] V. I. Ivanov, Radiotekhn. Elektron. 5, 393 (1959).
[6] J. R. Wait, *Electromagnetic Radiation from Cylindrical Structures* (Pergamon Press, New York, 1959).

Following Jones,[3] we obtain for the total field

$$\psi = \frac{1}{2(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{\infty} \frac{d\mu}{\cosh \mu\pi} \frac{(\tan \frac{1}{2}\varphi_0)^{i\mu-\frac{1}{2}}}{\cos \varphi_0} \times \left(D_{i\mu-\frac{1}{2}}(-h\xi) - \frac{z D_{i\mu-\frac{1}{2}}(-h\xi_0) - h D'_{i\mu-\frac{1}{2}}(-h\xi_0)}{z D_{i\mu-\frac{1}{2}}(h\xi_0) + h D'_{i\mu-\frac{1}{2}}(h\xi_0)} D_{i\mu-\frac{1}{2}}(h\xi) \right) D_{-i\mu-\frac{1}{2}}(h\eta), \quad (4)$$

and it is easy to verify that ψ in Eq. (4) satisfies the wave equation (2) and $h = (2ik)^{\frac{1}{2}}$.

The normal derivative of the current on the parabolic cylinder is

$$\frac{1}{(\xi_0^2 + \eta^2)^{\frac{1}{2}}} \frac{\partial \psi}{\partial \xi}\bigg|_{\xi=\xi_0} = \frac{-z\psi}{(\xi_0^2 + \eta^2)^{\frac{1}{2}}}\bigg|_{\xi=\xi_0} = \frac{zh}{2\pi(\xi_0^2 + \eta^2)^{\frac{1}{2}}} \int_{-\infty}^{\infty} \Gamma(i\mu + \tfrac{1}{2}) \frac{(\tan \frac{1}{2}\varphi_0)^{i\mu-\frac{1}{2}}}{\cos \frac{1}{2}\varphi_0} \frac{D_{-i\mu-\frac{1}{2}}(h\eta)}{z D_{i\mu-\frac{1}{2}}(h\xi_0) + h D'_{i\mu-\frac{1}{2}}(h\xi_0)} d\mu, \quad (5)$$

where we have used the Wronskians connecting the solutions as

$$D'_{i\mu-\frac{1}{2}}(z) D_{i\mu-\frac{1}{2}}(-z) + D_{i\mu-\frac{1}{2}}(z) D'_{i\mu-\frac{1}{2}}(-z) = (2/\pi)^{\frac{1}{2}} \Gamma(i\mu + \tfrac{1}{2}) \cosh \mu\pi. \quad (6)$$

As Rice[1] has rather convincingly shown and as we shall indicate later on, integrals of the type in Eq. (5) have a saddle point somewhere near $\mu = \frac{1}{2}k\xi_0^2$, while the path of steepest descent has the form indicated in Fig. 2.

It follows that Eq. (5) can be adequately approximated if the integral can be well approximated near this saddle point. But if $\frac{1}{2}k\xi_0^2$ is large (that is, if the parabola has a minimum radius of curvature large with respect to the wavelength), then we may resort to approximations that are asymptotic to the parameter μ.

Olver[7] has developed such asymptotic results in very complete detail. The result corresponding to our situation is

$$D_{i\mu-\frac{1}{2}}(hz) \sim 2\pi^{\frac{1}{2}}\mu^{\frac{1}{6}} \exp[\tfrac{1}{2}(i\mu - \tfrac{1}{2}) \ln i\mu - \tfrac{1}{2}i\mu + \tfrac{1}{12}i\pi] \times \left(\frac{\xi_1 \mu}{2kz^2 - 4\mu} \right)^{\frac{1}{4}} \mathrm{Ai}(\mu^{\frac{2}{3}} e^{\frac{1}{3}i\pi} \xi_1), \quad (7)$$

where

$$\tfrac{2}{3}\xi_1^{\frac{3}{2}} = 2\int_1^{(kz^2/2\mu)^{\frac{1}{2}}} (s^2 - 1)^{\frac{1}{2}}\, ds. \quad (8)$$

Equation (7) is valid for large $|\mu|$ uniformly with respect to $\arg \mu$ when

$$-\tfrac{5}{2}\pi < \arg \mu < \tfrac{3}{2}\pi, \quad (9)$$

and z is real ($z = \xi_0$ or η) and does not run into the lines shown in Fig. 3. The upper limit of integration in Eq. (8) is confined to the paths shown in Fig. 3; i.e., the locus of t,

$$t = (kz^2/2\mu)^{\frac{1}{2}}, \quad (10)$$

is plotted in Fig. 3 along the original and steepest descent contours.

Using the asymptotic form for the Airy function

$$\mathrm{Ai}(\xi) \sim \frac{\exp(-\frac{2}{3}\xi^{\frac{3}{2}})}{2(\pi)^{\frac{1}{2}}\xi^{\frac{1}{4}}}, \quad |\arg \xi| < \pi, \quad (11)$$

in Eq. (7), we obtain an asymptotic expansion for

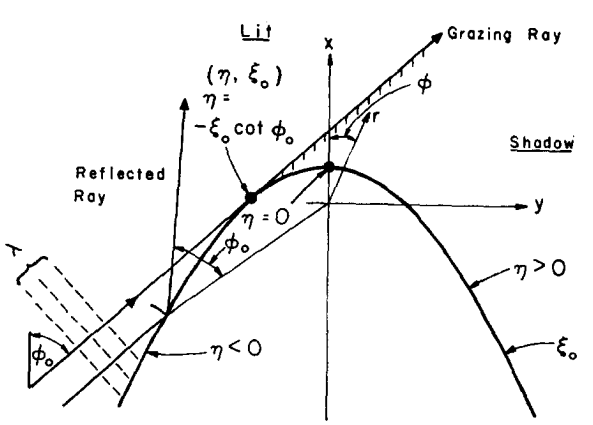


FIG. 1. Geometry of incident wave and parabolic cylinder used in the derivation of the Fock currents. The origin is at the focus of the parabola. From simple geometric considerations, it is not difficult to show that φ_0 remains constant for all points on the cylinder.

[7] F. W. J. Olver, J. Res. Nat. Bur. Std., B63, 131 (1959).

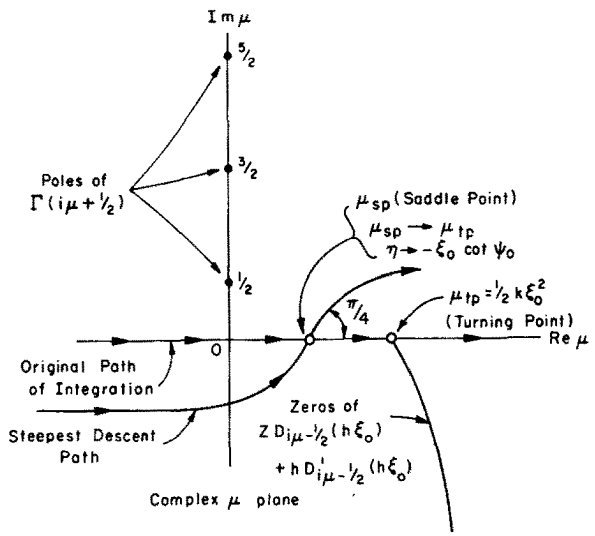


FIG. 2. Paths of integration used in studying the current density on the parabolic cylinder. Also shown are the poles, zeros and saddle points of the integrand.

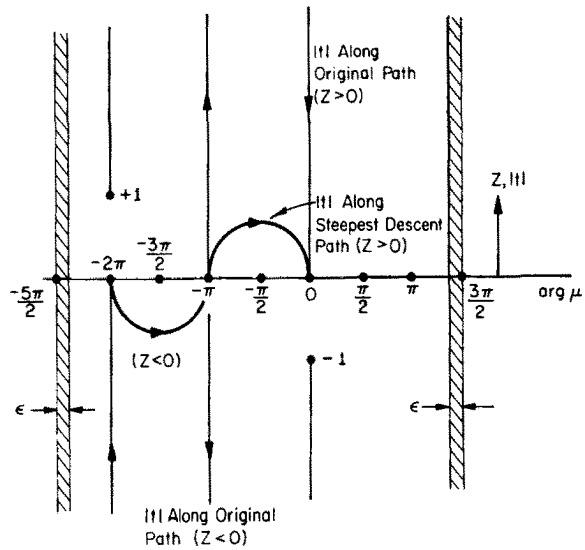


FIG. 3. Region of validity of the asymptotic expansion for $D_{i\mu-\frac{1}{2}}(z)$, together with $|t|$ along original path and steepest descent path.

$D_{-i\mu-\frac{1}{2}}(h\eta)$, i.e.,

$$D_{-i\mu-\frac{1}{2}}(h\eta) \sim (4 + 2k\eta^2/\mu)^{-\frac{1}{4}} \exp\left[-\tfrac{1}{2}(i\mu + \tfrac{1}{2}) \ln i\mu + \tfrac{1}{2}i\mu\right] \times \begin{cases} e^{-i\mu\xi_2}, & \eta > 0, \\ e^{i\mu\xi_2}, & \eta < 0, \end{cases} \quad (12)$$

where

$$\xi_2 = \left(\frac{k\eta^2}{2\mu}\right)^{\frac{1}{2}}\left(1 + \frac{k\eta^2}{2\mu}\right)^{\frac{1}{2}} + \ln\left[\left(\frac{k\eta^2}{2\mu}\right)^{\frac{1}{2}} + \left(1 + \frac{k\eta^2}{2\mu}\right)^{\frac{1}{2}}\right]. \quad (13)$$

We also need an asymptotic expansion for $D'_{i\mu-\frac{1}{2}}$. Differentiating the asymptotic expansion in Eq. (7), we find, considering $h\xi_0$ a parameter,

$$\frac{\partial D_{i\mu-\frac{1}{2}}(h\xi_0)}{\partial(h\xi_0)} \sim 2(\pi)^{\frac{1}{2}} \exp\left[\tfrac{1}{2}(i\mu - \tfrac{1}{2}) \ln i\mu - \tfrac{1}{2}i\mu + \tfrac{1}{12}i\pi\right]\mu^{\frac{1}{6}}\left(\frac{\xi_1}{2k\xi_0^2/\mu - 4}\right)^{\frac{1}{4}} \times e^{\frac{1}{3}i\pi}\mu^{\frac{2}{3}}\,\mathrm{Ai}'(e^{\frac{1}{3}i\pi}\xi_1\mu^{\frac{2}{3}})\frac{\partial\xi_1}{\partial(h\xi_0)}, \quad (14)$$

where the terms coming from the differentiation of the quantity raised to the power $\frac{1}{4}$ are neglected since they result in a higher-order negative power of μ. From Eq. (8) it is not difficult to show that

$$\frac{\partial\xi_1}{\partial(h\xi_0)} = \frac{e^{-\frac{1}{4}i\pi}}{(\mu\xi_1)^{\frac{1}{2}}}\left(\frac{k\xi_0^2}{2\mu} - 1\right)^{\frac{1}{2}}, \quad \text{for all } \xi_1. \quad (15)$$

We have

$$z D_{i\mu-\frac{1}{2}} + h D'_{i\mu-\frac{1}{2}} \sim 2(\pi)^{\frac{1}{2}} \exp\left[\tfrac{1}{2}(i\mu - \tfrac{1}{2}) \ln i\mu - \tfrac{1}{2}i\mu + \tfrac{1}{12}i\pi\right]\mu^{\frac{1}{6}} \times \left[z\left(\frac{\xi_1}{2k\xi_0^2/\mu - 4}\right)^{\frac{1}{4}}\mathrm{Ai}\,(e^{\frac{1}{3}i\pi}\mu^{\frac{2}{3}}\xi_1) + \frac{h\mu^{\frac{1}{6}}}{2}\, e^{\frac{1}{12}i\pi}\left(\frac{2k\xi^2/\mu - 4}{\xi_1}\right)^{\frac{1}{4}}\mathrm{Ai}'(e^{\frac{1}{3}i\pi}\mu^{\frac{2}{3}}\xi_1)\right]. \quad (16)$$

The exponent of the integrand in Eq. (5) is examined for saddle points when the arguments of the Airy functions are supposed large, and Jones[3] shows that the saddle points are solutions of

$$\mu = \tfrac{1}{2}k(\xi_0^2 - \eta^2)\sin^2\varphi_0 - k\xi_0\eta\sin\varphi_0\cos\varphi_0. \quad (17)$$

Jones' textbook contains a typographical error; i.e., Jones shows

$$\mu = \tfrac{1}{2}k(\xi_0^2 - \eta^2)\sin\varphi_0 - k\xi_0\eta\sin\varphi_0\cos\varphi_0.$$

The greatest contribution to the integral in Eq. (5) must come from the point $\eta = -\xi_0\cos\varphi_0$ since it corresponds to $\mu = \frac{1}{2}k\xi_0^2$ (the turning point) and $\xi_1 = 0$, so it is no longer possible to replace the Airy functions in Eq. (7) by their asymptotic expansions.

The expression for the normal derivative of the current on the cylinder becomes

$$\frac{1}{(\xi_0^2+\eta^2)^{\frac{1}{2}}}\frac{\partial\psi}{\partial\xi_0} = \frac{-zh}{2\pi(\xi_0^2+\eta^2)^{\frac{1}{2}}}\frac{e^{-\frac{1}{12}i\pi}}{(\sin\varphi_0)^{\frac{1}{2}}}\int_{-\infty}^{\infty}d\mu e^{-i\mu[\xi_2-\ln\tan(\frac{1}{2}\varphi_0)]} \times\left(4+\frac{2k\eta^2}{\mu}\right)^{-\frac{1}{4}}\mu^{\frac{1}{6}}\left[z\left(\frac{\xi_1}{2k\xi_0^2/\mu-4}\right)^{\frac{1}{4}}\mathrm{Ai}\,(e^{\frac{1}{3}i\pi}\mu^{\frac{2}{3}}\xi_1)+\frac{he^{i\frac{1}{12}\pi}}{2}\mu^{\frac{1}{6}}\left(\frac{2k\xi_0^2/\mu-4}{\xi_1}\right)^{\frac{1}{4}}\mathrm{Ai}'\,(e^{\frac{1}{3}i\pi}\mu^{\frac{2}{3}}\xi_1)\right]^{-1}. \quad (18)$$

The Laplace approximation[8] states that the major contribution to the integral in Eq. (18) will arise from the neighborhoods of the points at which

$$f(\mu,\eta)=\mu[\xi_2-\ln\tan(\tfrac{1}{2}\varphi_0)] \quad (19)$$

attains its supremum. These points will be solutions of

$$\frac{\partial f}{\partial\mu}=0, \quad (20)$$

and substituting for ξ_2 from (13) into (20) yields

$$\ln\left(\frac{\tan(\frac{1}{2}\varphi_0)}{\eta(k/2\mu)^{\frac{1}{2}}+(1+k\eta^2/2\mu)^{\frac{1}{2}}}\right)=0. \quad (21)$$

It is easy to verify that the point

$$\eta=-\xi_0\cot\varphi_0,$$
$$\mu_{\mathrm{SP}}=\tfrac{1}{2}k\xi_0^2 \quad (22)$$

satisfies Eq. (21). Thus, the greatest contribution to the integral comes from the point on the parabolic cylinder near grazing. This is not a surprising result. Continuing with Laplace's method, we define the new variable τ as

$$\mu=\mu_{\mathrm{SP}}+\mu_{\mathrm{SP}}^{\frac{1}{3}}\tau,\quad |\tau|\ll 1, \quad (23)$$

with

$$d\mu=\mu_{\mathrm{SP}}^{\frac{1}{3}}\,d\tau \quad (24)$$

and μ_{SP} as given in Eq. (22). When Eq. (23) is substituted into Eq. (8) and terms up to and including powers of τ to the $\frac{3}{2}$ are retained, we find

$$\xi_1=-\tau\mu_{\mathrm{SP}}^{-\frac{2}{3}}. \quad (25)$$

Since we are interested in the integral near the point where it assumes its greatest contribution, it would appear that a Taylor-series expansion of $f(\mu,\eta)$ about this point would be useful. That is, consider

$$f(\mu,\eta)=f(\mu_{\mathrm{SP}},\eta_0)+\frac{\partial f}{\partial\mu}\Big|_{\mu_{\mathrm{SP}},\eta_0}(\mu-\mu_{\mathrm{SP}})+\frac{\partial f}{\partial\eta}\Big|_{\mu_{\mathrm{SP}},\eta_0}(\eta-\eta_0) +\frac{1}{2!}\left(\frac{\partial^2 f}{\partial\eta^2}\Big|_{\mu_{\mathrm{SP}},\eta_0}(\eta-\eta_0)^2+2\frac{\partial^2 f}{\partial\mu\partial\eta}\Big|_{\mu_{\mathrm{SP}},\eta_0}(\eta-\eta_0)(\mu-\mu_{\mathrm{SP}})+\frac{\partial^2 f}{\partial\mu^2}\Big|_{\mu_{\mathrm{SP}},\eta_0}(\mu-\mu_{\mathrm{SP}})^2\right) +\frac{1}{3!}\left(\frac{\partial^3 f}{\partial\eta^3}\Big|_{\mu_{\mathrm{SP}},\eta_0}(\eta-\eta_0)^3+3\frac{\partial^3 f}{\partial\mu\partial\eta^2}\Big|_{\mu_{\mathrm{SP}},\eta_0}(\eta-\eta_0)^2(\mu-\mu_{\mathrm{SP}})\right.$$
$$\left.+3\frac{\partial^3 f}{\partial\mu^2\partial\eta}\Big|_{\mu_{\mathrm{SP}},\eta_0}(\eta-\eta_0)(\mu-\mu_{\mathrm{SP}})^2+\frac{\partial^3 f}{\partial\mu^3}\Big|_{\mu_{\mathrm{SP}},\eta_0}(\mu-\mu_{\mathrm{SP}})^3\right), \quad (26)$$

where terms through the third derivative have been retained. Since we have assumed that τ in Eq. (23) is a small quantity, we may neglect squares and higher powers of τ in Eq. (26). This does not mean we may neglect higher-order powers of $(\eta-\eta_0)$, since we have not assumed this quantity small. Indeed, for angles near $\varphi_0=\frac{1}{2}\pi$, the terms that contribute are $(\eta-\eta_0)^3$, as we shall discover. We have

$$f(\mu,\eta)=f(\mu_{\mathrm{SP}},\eta_0)+\frac{\partial f}{\partial\eta}\Big|_{\mu_{\mathrm{SP}},\eta_0}(\eta-\eta_0) +\frac{1}{2!}\left(\frac{\partial^2 f}{\partial\eta^2}\Big|_{\mu_{\mathrm{SP}},\eta_0}(\eta-\eta_0)^2+2\frac{\partial^2 f}{\partial\mu\partial\eta}\Big|_{\mu_{\mathrm{SP}},\eta_0}(\eta-\eta_0)(\mu-\mu_{\mathrm{SP}})\right) +\frac{1}{3!}\left(\frac{\partial^3 f}{\partial\eta^3}\Big|_{\mu_{\mathrm{SP}},\eta_0}(\eta-\eta_0)^3+3\frac{\partial^3 f}{\partial\mu\partial\eta^2}\Big|_{\mu_{\mathrm{SP}},\eta_0}(\eta-\eta_0)^2(\mu-\mu_{\mathrm{SP}})\right), \quad (27)$$

[8] E. T. Copson, *Asymptotic Expansions* (Cambridge University Press, Cambridge, England, 1965).

where

$$\left.\frac{\partial f}{\partial \mu}\right|_{\mu_{SP},\eta_0} = 0,$$

since this is how the point (μ_{SP}, η_0) was determined. The various terms in Eq. (27) are given by

$$f(\mu_{SP}, \eta_0) = -\frac{k\xi_0^2}{2}\frac{\cos\varphi_0}{\sin^2\varphi_0}, \tag{28a}$$

$$\left.\frac{\partial f}{\partial \eta}\right|_{\mu_{SP},\eta_0} = \frac{k\xi_0}{\sin\varphi_0}, \tag{28b}$$

$$\left.\frac{\partial^2 f}{\partial \eta^2}\right|_{\mu_{SP},\eta_0} = -k\cos\varphi_0, \tag{28c}$$

$$\left.\frac{\partial^2 f}{\partial \mu \partial \eta}\right|_{\mu_{SP},\eta_0} = \frac{\sin\varphi_0}{\xi_0}, \tag{28d}$$

$$\left.\frac{\partial^3 f}{\partial \eta^3}\right|_{\mu_{SP},\eta_0} = \frac{k\sin^3\varphi_0}{\xi_0}, \tag{28e}$$

$$\left.\frac{\partial^3 f}{\partial \mu \partial \eta^2}\right|_{\mu_{SP},\eta_0} = \frac{\cos\varphi_0 \sin^2\varphi_0}{\xi_0^2}. \tag{28f}$$

Substituting Eqs. (28) into Eq. (27) yields

$$f(\mu,\eta) = \frac{k\xi_0\omega}{\sin^2\varphi_0}\left(1 - \frac{\omega\cos\varphi_0}{2\xi_0}\right) + \frac{k\xi_0^2}{2}\left(-\frac{\cos\varphi_0}{\sin^2\varphi_0} + \frac{\omega^3}{3\xi_0^3}\right) + \left(\frac{k}{2\xi_0}\right)^{\frac{1}{3}}\omega\tau\left(1 + \frac{\omega\cos\varphi_0}{2\xi_0}\right), \tag{29}$$

where

$$\omega = \eta\sin\varphi_0 + \xi_0\cos\varphi_0. \tag{30}$$

As a check on the result in Eq. (29), consider the case

$$\varphi_0 = \tfrac{1}{2}\pi:$$

$$f(\mu,\eta) = \underbrace{k\xi_0\eta}_{\substack{\text{incident}\\ \text{wave}}} + \frac{k\eta^3}{6\xi_0} + \left(\frac{k}{2\xi_0}\right)^{\frac{1}{3}}\eta\tau. \tag{31}$$

Fock[2] defines $f(\mu, \eta)$ in terms of the variable $l = \xi_0\eta$. Substituting for l in Eq. (57) gives

$$f(\mu, l) = \underbrace{kl}_{\substack{\text{incident}\\ \text{wave}}} + \frac{kl^3}{6\xi_0^4} + \left(\frac{k}{2\xi_0}\right)^{\frac{1}{3}}\frac{\tau l}{\xi_0}, \tag{32}$$

which agrees exactly with Fock's result, where the radius of curvature is

$$R_0 = \xi_0^2/\sin^3\varphi_0 = \xi_0^2, \quad \varphi_0 = \tfrac{1}{2}\pi. \tag{33}$$

It is convenient in many applications to separate the phase of the incident wave from the remaining factors in $f(\mu, \eta)$. If we add and subtract the phase of the incident wave

$$-\tfrac{1}{2}k\xi_0(\eta\sin\varphi_0 + \xi_0\cos\varphi_0) + \tfrac{1}{2}k\eta(\eta\cos\varphi_0 - \xi_0\sin\varphi_0), \tag{34}$$

then $f(\mu, \eta)$ becomes

$$f(\mu,\eta) = \underbrace{-kr\cos(\varphi - \varphi_0)}_{\text{incident wave}} + f_2(\Omega, \varphi_0) + f_1(\Omega, \varphi_0)\tau, \tag{35}$$

where we have defined the normalized distance along the parabolic cylinder

$$\Omega = \eta/\xi_0, \tag{36}$$

and where

$$f_1(\Omega, \varphi_0)/(\tfrac{1}{2}k\xi_0^2)^{\frac{1}{3}} = (\Omega\sin\varphi_0 + \cos\varphi_0) \times [1 + \tfrac{1}{2}(\Omega\sin\varphi_0 + \cos\varphi_0)\cos\varphi_0] \tag{37}$$

and

$$f_2(\Omega, \varphi_0)/k\xi_0^2 = \tfrac{1}{6}(\Omega\sin\varphi_0 + \cos\varphi_0)^3. \tag{38}$$

The corresponding functions given by Fock and and Jones[9] are

$$f_1^F(\Omega, \varphi_0)/(\tfrac{1}{2}k\xi_0^2)^{\frac{1}{3}} = \Omega\sin\varphi_0 + \cos\varphi_0 \tag{39}$$

and

$$f_2^F(\Omega, \varphi_0)/k\xi_0^2 = \tfrac{1}{6}(\Omega\sin\varphi_0 + \cos\varphi_0)^3, \tag{40}$$

and we find that Eq. (40) agrees exactly with the result in Eq. (38). Also, when $\Omega \cong -\cot\varphi_0$, Eq. (37) agrees with Eq. (39); however, as Ω moves away from the grazing point (i.e., $-\cot\varphi_0$) the results in Eqs. (37) and (39) will differ. Since in most applications one is interested only in the penumbral currents near grazing, the differences in Eqs. (37) and (39) do not appear to be significant.

The results in (37) and (38) show the range of validity of the Fock currents when the incident wave is not tangent to the apex of the cylinder. This is an extension of the results given in Jones' textbook.

[9] Note that the origin in Fock's equivalent replacement cylinder is at $\Omega = -\cot\varphi_0$ and our distance along the y axis [i.e., $\xi_0^3(\Omega + \cot\varphi_0)$] needs to be projected onto the direction of the impressed field.

3. CONCLUDING REMARKS

The validity of Fock's principle for the local field in the penumbra was investigated for the case of a plane wave striking a finitely conducting, parabolic cylinder at an arbitrary angle of incidence. The results show that Fock's principle is extremely accurate, provided that the observation point on the parabolic cylinder is near the point where the impressed field grazes the cylinder. When the impressed field travels in a direction tangent to the apex of the cylinder, Fock's principle agrees exactly with the results presented in this paper for all observation points.

ACKNOWLEDGMENTS

I wish to thank Dr. G. Hufford and Dr. J. Wait for their comments and suggestions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Stress-Tensor Commutators in Nonlinear Electrodynamics*

S. DESER AND L. K. MORRISON
Brandeis University, Waltham, Massachusetts 02154

(Received 9 May 1969)

Equal-time stress-tensor commutators are shown to have a simple form for general systems of spin zero and one. Electrodynamics with arbitrary self-coupling is treated in detail. Ordering and other problems of a purely quantum nature are not discussed.

I. INTRODUCTION

As is well known, Lorentz invariance of a local theory with spin ≤ 1 is effectively equivalent to the validity of the single local equal-time commutation relation[1,2]:

$$i[T^{00}(\mathbf{r}), T^{00}(\mathbf{r}')] = [T^{0k}(\mathbf{r}) + T^{0k}(\mathbf{r}')]\partial_k \delta^3(\mathbf{r} - \mathbf{r}'). \quad (1)$$

It is also known[1-4] that, for higher spins, additional model-dependent terms necessarily appear on the right.

In this paper we shall explicitly evaluate the full set of stress-tensor equal-time commutators $i[T^{0\mu}(\mathbf{r}), T^{\lambda\sigma}(\mathbf{r}')]$, relevant to the Poincaré algebra for a wide class of theories of spin zero and one.[5] We shall illustrate in terms of the self-coupled Maxwell (or massive vector) field, which is a highly derivative coupled theory, involving, as it does, arbitrary powers of $F_{\mu\nu}F^{\mu\nu}$ and $\epsilon^{\mu\nu\alpha\beta}F_{\mu\nu}F_{\alpha\beta}$ (but not derivatives of $F_{\mu\nu}$). For the best known model, Born–Infeld electrodynamics,[6] Dirac[1] has already shown by direct computation that neither nonlinearity nor the derivative coupling disturbs the simple form of Eq. (1). This is in spite of the fact that T^{00} becomes a complicated function, while T^{0k} retains its simple character as a spatial translation-generator density.

Our technique is an extension of that given in I, which is based on the introduction of an external metric field $g_{\mu\nu}$, the response to which defines $T^{\mu\nu}$ as the gravitational current. The general results we shall obtain are, strictly speaking, purely classical in that the operator-ordering problem is not considered. This is adequate for the Abelian case, where there is effectively no ordering ambiguity. For non-Abelian systems like the Yang–Mills field, it is known[7] that such problems are important and require the introduction of new, nonclassical terms.

We also do not consider the purely quantum problem, discussed in I, of Schwinger terms in the stress-tensor commutators.

II. GENERAL EQUAL-TIME STRESS-TENSOR COMMUTATORS

The action W of a local system in an arbitrary external metric field is obtained by the usual "minimal" substitution $\eta_{\mu\nu} \to g_{\mu\nu}$ in the Lorentz covariant form $W[\eta_{\mu\nu}]$. The (symmetric) contravariant stress-tensor density $\mathfrak{T}^{\mu\nu}[g]$ is the response of the system to a variation of the metric $\delta W = \frac{1}{2}\int \mathfrak{T}^{\mu\nu}(x)\delta g_{\mu\nu}(x)$. It is shown in I, using covariant conservation[8] of $\mathfrak{T}^{\mu\nu}$, that the normal stress tensor $T^{\mu\nu} = \mathfrak{T}^{\mu\nu}\,[g = \eta]$ obeys the following equal-time commutation relations:

$$i[T^{0\mu}(x), T^{\lambda\sigma}(x')]\delta(x^0 - x'^0)$$
$$= [\eta^{\mu\nu}T^{\lambda\sigma}(x) - \eta^{\mu\sigma}T^{\nu\lambda}(x) - \eta^{\mu\lambda}T^{\nu\sigma}(x)]\partial_\nu \delta(x - x') - 2\partial_\nu \frac{\delta \mathfrak{T}^{\nu\mu}(x)}{\delta g_{\lambda\sigma}(x')}, \quad (2)$$

where the variational derivative is also to be evaluated at $g_{\mu\nu} = \eta_{\mu\nu}$.

It would appear at first sight that the choice of basic field variables can alter the metric dependence of the action and, hence, both $\mathfrak{T}^{\mu\nu}$ and $\delta\mathfrak{T}^{\mu\nu}(x)/\delta g_{\alpha\beta}(x')$. We shall see that there is effectively no ambiguity. The choice of appropriate initial field variables in the action will be effectively forced by covariance considerations, while the elimination of redundant variables leading to a canonical set will define the correct dependence for evaluating $\delta\mathfrak{T}^{\mu\nu}/\delta g_{\alpha\beta}$.

Consider now the general form of the action expressed as a function of a set of unconstrained canonical momenta π^A and fields ϕ_A, to be varied independently in obtaining the field equations. (The

* Work supported in part by USAF, OAR, under OSR Grant 368-67.

[1] P. A. M. Dirac, Rev. Mod. Phys. **34**, 592 (1962).

[2] J. Schwinger, Phys. Rev. **130**, 406, 800 (1963). In this paper, we do not consider the mathematical problems of the existence of equal-time commutators as either three-, or limits of four-, dimensional distributions. For a recent discussion of this question for current commutators, see, for example, D. Boulware and S. Deser, Phys. Rev. **175**, 1912 (1968).

[3] D. G. Boulware and S. Deser, J. Math. Phys. **8**, 1468 (1967), referred to as I in text.

[4] An explicit example is given in Ref. 2 for spin $\frac{3}{2}$ and in S. Deser, J. Trubatch, and S. Trubatch [Nuovo Cimento **39**, 1159 (1965)] for spin 2.

[5] Fermions are not treated here, since somewhat different techniques are involved. It is known that they satisfy Eq. (1).

[6] M. Born and L. Infeld, Proc. Roy. Soc. (London), **A144**, 425 (1934).

[7] J. Schwinger, Phys. Rev. **127**, 324 (1962).

[8] This assumes the absence of preferred vectors in the original action, i.e., Lorentz invariance.

second-order formalism will be discussed later.) It is assumed that elimination of redundant ("constraint") variables maintains locality—which is the case if and only if the spin is ≤ 1. We shall see that one may write the action[9] as

$$W = \int \left(\sum_A \pi^A \partial_0 \phi_A + N\theta_0^0(\pi, \phi, g_{ij}) + {}^3g^{ij} N_i \theta_j^0(\pi, \phi, g_{ij})\, d^4x \right), \quad (3)$$

where $N \equiv (-g^{00})^{-\frac{1}{2}} = (N_i N_j^3 g_{ij} - g_{00})^{\frac{1}{2}}$, $N_i \equiv g_{0i}$, and ${}^3g^{ij}$ is the matrix inverse of the spatial components g_{ij}. The dependence on the N, N_i components of the metric is linear; accordingly, their coefficients, namely θ_μ^0, when evaluated at $g_{ij} = \delta_{ij}$, are just the components of the stress tensor T_μ^0. The generator of canonical transformations, $\int \pi^A \partial_0 \phi_A \, d^4x$, is entirely metric independent and so does not contribute at all to $T^{\mu\nu}$. The right-hand side of the canonical commutation relations (CCR), $[\pi^A(\mathbf{r}), \phi_A(\mathbf{r}')]$, will likewise be metric independent, transforming like π^A and ϕ_A at $\mathbf{r}$ and $\mathbf{r}'$, respectively. The $g_{0\mu}$ independence of the generator is, of course, to be expected if the (π, ϕ) variables are appropriate initial data at $t = 0$, since the $g_{0\mu}$ only describe the change of coordinates off this surface. (This question has been discussed, for weak g_{00}, in Ref. 2.) The further lack of g_{ij} dependence means that the matter variables have been correctly chosen to be kinematically independent of the gravitational field's variables.[10]

Variation of W with respect to $g_{0\mu}$ immediately yields

$$\mathfrak{T}^{00} = -N^{-1}\theta_0^0, \quad \mathfrak{T}^{0i} = {}^3g^{ij}(\theta_j^0 + N^{-1}N_j\theta_0^0). \quad (4a)$$

The spatial stresses $\mathfrak{T}^{ij}$ will involve the particular g_{ij} dependence in θ_μ^0:

$$\tfrac{1}{2}\mathfrak{T}^{ij} = N\frac{\delta\theta_0^0}{\delta g_{ij}} - \tfrac{1}{2}(g^{il}N_l\theta_m^0 g^{mj} + g^{jl}N_l\theta_m^0 g^{mi}) + {}^3g^{lk}N_k\frac{\delta\theta_l^0}{\delta g_{ij}} - (2N)^{-1}N^iN^j\theta_0^0. \quad (4b)$$

From these expressions and the reciprocity relation

$$\frac{\delta\mathfrak{T}^{\mu\nu}(x)}{\delta g_{\alpha\beta}(x')} \equiv \frac{\delta^2 W}{\delta g_{\alpha\beta}(x')\delta g_{\mu\nu}(x)} = \frac{\delta\mathfrak{T}^{\alpha\beta}(x')}{\delta g_{\mu\nu}(x)}, \quad (5)$$

we obtain the variational derivatives $\delta\mathfrak{T}/\delta g$, evaluated at $g_{\mu\nu} = \eta_{\mu\nu}$, i.e., $N = 1$, $N_i = 0$, $g_{ij} = \delta_{ij}$:

$$\frac{\delta\mathfrak{T}^{00}}{\delta g'_{00}} = -\tfrac{1}{2}T^{00}\delta(x - x'), \quad (6a)$$

$$\frac{\delta\mathfrak{T}^{00}}{\delta g'_{0i}} = \frac{\delta\mathfrak{T}^{0i'}}{\delta g_{00}} = 0, \quad (6b)$$

$$\frac{\delta\mathfrak{T}^{00}}{\delta g'_{ij}} = \frac{\delta\mathfrak{T}^{ij'}}{\delta g_{00}} = -\frac{\delta\theta_0^0}{\delta g'_{ij}}, \quad (6c)$$

$$\frac{\delta\mathfrak{T}^{0i}}{\delta g'_{0j}} = -\tfrac{1}{2}\delta_{ij}T^{00}\delta(x - x'), \quad (6d)$$

$$\frac{\delta\mathfrak{T}^{0i}}{\delta g'_{kl}} = \frac{\delta\mathfrak{T}^{kl'}}{\delta g_{0i}} = -\tfrac{1}{2}(T^{0l}\delta_{ik} + T^{0k}\delta_{il})\delta(x - x'), \quad (6e)$$

$$\frac{\delta\mathfrak{T}^{ij}}{\delta g'_{lm}} = 2\frac{\delta^2\theta_0^0}{\delta g'_{lm}\delta g_{ij}}. \quad (6f)$$

Now from Eq. (4b), taken at $g_{\mu\nu} = \eta_{\mu\nu}$, we find

$$\frac{\delta\theta_0^0}{\delta g'_{ij}} = \tfrac{1}{2}T^{ij}\delta(x - x').$$

The momentum density will be seen to maintain its simple metric independent (noninteracting) form, so that $\delta\theta_i^0/\delta g'_{kl}$ actually vanishes. This is essential for the time locality of the theory. The remaining derivative, $\delta^2\theta_0^0/\delta g_{ij}\delta g'_{lm}$, which appears in only one place, cannot be further reduced, but it does not affect time locality. Inserting the above results into the basic formula, Eq. (2), finally yields, after some simple manipulations, the following general form[11]:

$$i[T^{00}(\mathbf{x}), T^{00}(\mathbf{x}')] = [T^{0i}(\mathbf{x}) + T^{0i}(\mathbf{x}')]\partial_i\delta(\mathbf{x} - \mathbf{x}'), \quad (7a)$$

$$i[T^{00}(\mathbf{x}), T^{0i}(\mathbf{x}')] = [T^{ij}(\mathbf{x}) + \delta^{ij}T^{00}(\mathbf{x}')]\partial_j\delta(\mathbf{x} - \mathbf{x}'), \quad (7b)$$

$$i[T^{0i}(\mathbf{x}), T^{0j}(\mathbf{x}')] = [T^{0j}(\mathbf{x})\partial_i + T^{0i}(\mathbf{x}')\partial_j]\delta(\mathbf{x} - \mathbf{x}'), \quad (7c)$$

$$i[T^{00}(\mathbf{x}), T^{ij}(\mathbf{x}')] = [T^{0j}(\mathbf{x}')\partial_i + T^{0i}(\mathbf{x}')\partial_j - (\partial^0 T^{ij}(\mathbf{x}))]\delta(\mathbf{x} - \mathbf{x}'), \quad (7d)$$

$$i[T^{0i}(\mathbf{x}), T^{lm}(\mathbf{x}')] = [T^{lm}(\mathbf{x})\partial_i - \delta^{il}T^{mk}(\mathbf{x}')\partial_k - \delta^{im}T^{lk}(\mathbf{x}')\partial_k]\delta(\mathbf{x} - \mathbf{x}') - 4\partial_k\frac{\delta^2\theta_0^0}{\delta g_{lm}(\mathbf{x})\delta g_{ik}(\mathbf{x}')}, \quad (7e)$$

[9] For higher spins, this simple form no longer holds. While this is a great difficulty with respect to gravitational coupling [C. Aragone, Nuovo Cimento (to be published)], it is known that, say, the free spin 2 system is Lorentz invariant (for example, by direct evaluation of the Poincaré commutators). However, the derivations of this section are not applicable, because reduction to the $2S + 1$ independent variables of such systems introduces spatial nonlocality. Indeed, there would be a paradox otherwise, since a deviation from Eq. (3) would result in apparent time nonlocality through the $\partial_0(\delta\mathfrak{T}/\delta g)$ terms in Eq. (2).

[10] See R. Arnowitt, S. Deser, and C. W. Misner, J. Math. Phys. **1**, 434 (1960), and Phys. Rev. **120**, 1313 (1960), where the necessity of this form for coupled gravitational and matter fields is treated.

[11] In I, a complementary form for these commutators is obtained from another approach. The specific action, Eq. (3), is not used, but only the requirement of time locality [i.e., that no $\partial_0\delta(x^0 - x^{0\prime})$ terms appear on the right].

where reference to the purely quantum ($\partial^3\delta$) terms has been dropped. Thus, all commutators but the last are entirely model independent. As it happens, for the free Maxwell scalar field, the $\delta^2\theta_0^0/\delta g_{lm}\delta g'_{ik}$ term is also expressible in terms of $T^{\mu\nu}$ alone.[12] The above result may, of course, be obtained directly in each case by use of the CCR, taking all constraints properly into account.

In view of the special role played by the first commutator in Eq. (7a), we recall that it is equivalent to the simple requirement that in a weak external g_{00} [i.e., $g_{ij} = \delta_{ij}$, $N_i = 0$, $N \sim (-g_{00})^{\frac{1}{2}}$]

$$\frac{\delta[(-g_{00})^{\frac{1}{2}}\mathfrak{T}^{00}]}{\delta g_{00}} = 0 = \frac{\delta\mathfrak{T}^{0i}}{\delta g_{00}}. \tag{8}$$

In the next section, we treat nonlinear electrodynamics and other systems in a simple way to show that this criterion is satisfied.

III. ENERGY DENSITY COMMUTATORS

In this section, we deal with self-coupled spin-one systems in an introductory way to show that the Dirac–Schwinger condition, Eq. (1), or alternatively Eq. (8), holds there. We shall work in the usual second-order formalism, where the special form, Eq. (3), is not made explicit. In Sec. IV, we use first-order formalism and deal with the full set of stress-tensor commutators given in Eq. (7). The physics is, of course, independent of the formalism used.

We consider generalizations of electrodynamics involving arbitrary powers of the field strength, but not its derivatives. The Lagrangian will now be a function of the two independent parity and Lorentz scalars α and β^2, where

$$\alpha \equiv \tfrac{1}{2}F_{\mu\nu}F_{\alpha\beta}\eta^{\mu\alpha}\eta^{\nu\beta} = \mathfrak{B}^2 - \mathfrak{E}^2, \tag{9a}$$

$$\beta \equiv \tfrac{1}{8}\epsilon^{\mu\nu\alpha\beta}F_{\mu\nu}F_{\alpha\beta} = \tfrac{1}{8}F^{\alpha\beta}F_{\alpha\beta} = -\mathfrak{E}\cdot\mathfrak{B}. \tag{9b}$$

Familiar examples are the phenomenological Euler Lagrangian[13]

$$\mathfrak{L} = -\alpha + \lambda\alpha^2 + \mu\beta^2 \tag{10}$$

and the Born–Infeld theory[6]

$$\mathfrak{L}_B = -[-\det(\eta_{\mu\nu} + F_{\mu\nu})]^{\frac{1}{2}} + 1$$
$$= -(1 + \alpha - \beta^2)^{\frac{1}{2}} + 1 \equiv -\gamma + 1, \tag{11}$$

where an appropriate dimensional constant is absorbed into the definition of $F_{\mu\nu}$.

In going over to the covariant forms, several points must be borne in mind. The first is that there is no freedom of choice of the fundamental tensor variables, in terms of which the prescription $\eta_{\mu\nu} \to g_{\mu\nu}$ will be made: One *must* express everything in terms of the covariant tensor $F_{\mu\nu}$ and then insert $g_{\mu\nu}$ and $(-g)^{\frac{1}{2}}$ factors as needed, for $F_{\mu\nu}$ is merely a shorthand for the expression $\partial_\mu A_\nu - \partial_\nu A_\mu$, which *is* a general tensor. On the other hand, $Q^{\mu\nu} \equiv \partial^\mu A^\nu - \partial^\nu A^\mu$ (note that $Q^{\mu\nu} \neq F^{\mu\nu} \equiv F_{\alpha\beta}g^{\alpha\mu}g^{\beta\nu}$) or $\partial_\mu\mathcal{A}_\nu - \partial_\nu\mathcal{A}_\mu$ (where $\mathcal{A}_\mu$ is a vector density of any nonzero weight) are not tensors. This fixes the explicit $g_{\mu\nu}$ dependence (and, hence, also $\mathfrak{T}^{\mu\nu}$) once we recall that the totally antisymmetric symbol $\epsilon^{\mu\nu\alpha\beta}$ transforms[14] as a fourth-rank tensor density of weight one, but is explicitly metric *independent* (though its covariant form $\epsilon_{\mu\nu\alpha\beta}$ is not). Consequently, α is a scalar with explicit metric dependence, while β is a metric-independent scalar density. Of course, every Lagrangian must be made a scalar density by appropriate insertions of $(-g)^{\frac{1}{2}}$ factors. The Born–Infeld theory is unique in this respect in not requiring any explicit $(-g)^{\frac{1}{2}}$ factors, since the square root of the determinant of any covariant second-rank tensor (such as $g_{\mu\nu} + F_{\mu\nu}$) is automatically a scalar density.[14]

The above discussion suffices to specify the first metric variation of the action, but its second variation $\delta\mathfrak{T}/\delta g$, which enters in the stress-tensor commutators, cannot be evaluated without further refinement. For example, if we were to take $\delta\mathfrak{T}^{0i}/\delta g_{00}$ at $g_{\mu\nu} = \eta_{\mu\nu}$ for the Maxwell field directly from the $\mathfrak{T}^{\mu\nu}$ form, it would not vanish, since for weak g_{00} it is easy to see that $\mathfrak{T}^{0i} \sim (-g_{00})^{-\frac{1}{2}}F_{0j}F_{ji}$. The point is that there is an implicit assumption in the derivation of Eq. (2) that $\mathfrak{T}^{\mu\nu}$ is expressed in terms of variables which are independent of the metric (i.e., of a weak g_{00} for present purposes). Such variables must then be appropriate Cauchy data independent of changes of coordinates off the initial surface. We have already discussed (in Sec. II), in an essentially Hamiltonian framework, the appropriate choice in terms of the generators of canonical transformations (or, alternatively, the CCR) being metric independent. For spin-one fields it is known[10,3] that the correct initial variables are not the

[12] For the Maxwell case in particular, we find

$$i[T^{0k}(\mathbf{x}), T^{mn}(\mathbf{x}')] = [T^{mn}(\mathbf{x})\partial_k - \delta^{km}T^{ln}(\mathbf{x}')\partial_l - \delta^{kn}T^{lm}(\mathbf{x}')\partial_l]$$
$$\times \delta(\mathbf{x} - \mathbf{x}') + [\delta^{mn}T^{kl}(\mathbf{x}')\partial_l + T^{mn}(\mathbf{x}')\partial_k$$
$$+ T^{00}(\mathbf{x}')(\delta^{km}\partial_n + \delta^{kn}\partial_m - \delta^{mn}\partial_k)]\delta(\mathbf{x} - \mathbf{x}')$$

(see Appendix A). This result disagrees with earlier statements [see A. Peres, Nuovo Cimento **34**, 340 (1964); I. Bialynicki-Birula, Nuovo Cimento **35**, 697 (1964)] that one cannot express this commutator in terms of $T^{\mu\nu}$ alone. The same commutator for the spin-zero field is also expressible entirely as a (different) function of $T^{\mu\nu}$. This is because it depends only on $\phi_{,i}\,\phi_{,j}$ and π^2, both of which are expressible in terms of T_{ij} and T_{00}.

[13] W. Heisenberg and H. Euler, Z. Physik **38**, 714 (1936).

[14] Indeed, $\epsilon^{\mu\nu\alpha\beta}$ may even be defined in nonmetric spaces; see, for example, E. Schrödinger, *Space–Time Structure* (Cambridge University Press, Cambridge, England, 1963).

$F_{\mu\nu}$ (i.e., F_{0i} and F_{ij}) but rather

$$\mathcal{E}^i \equiv (-g)^{\frac{1}{2}}F^{0i} \equiv N({}^3g)^{\frac{1}{2}}F^{0i}$$

and

$$F_{ij} \text{ or } \mathfrak{B}^i \equiv \tfrac{1}{2}\epsilon^{ijk}F_{jk} \tag{12}$$

(or any appropriate functions of $\mathcal{E}$ and $\mathfrak{B}$). The usual Maxwell CCR

$$i[\mathcal{E}^i(\mathbf{r}), \mathfrak{B}^j(\mathbf{r}')] = \epsilon^{ijk}\partial_k\delta(\mathbf{r} - \mathbf{r}') \tag{13}$$

is clearly metric independent [$\delta(\mathbf{r})$ is a scalar density] and reflects the transformation properties of the $\mathcal{E}$ and $\mathfrak{B}$ under coordinate changes at $t = 0$. For self-coupled theories, $\mathcal{E}$ will no longer be conjugate to $\mathbf{A}$ (or equivalently $\mathfrak{B}$), as a result of the derivative nature of the coupling. However, the correct conjugate variable $\mathbf{\Pi}$ will be seen later to depend on $(\mathcal{E}, \mathfrak{B})$ but with no further $g_{0\mu}$ dependence (or it would not be a "good" Cauchy variable!); thus, we need not express things in terms of it for present purposes. The $\mathfrak{B}$ on the other hand will remain the canonical amplitude.

Now a term in the action of the form $\int \alpha(-g)^{\frac{1}{2}}\,d^4x$ will have a variation

$$\delta\int\alpha(-g)^{\frac{1}{2}}\,d^4x = \tfrac{1}{2}\delta\int F_{\mu\nu}F_{\alpha\beta}g^{\mu\alpha}g^{\nu\beta}(-g)^{\frac{1}{2}}\,d^4x$$
$$= \int(-F^{\mu\alpha}F^{\nu}{}_{\alpha} + \tfrac{1}{2}g^{\mu\nu}\alpha)(-g)^{\frac{1}{2}}\delta g_{\mu\nu}\,d^4x \tag{14a}$$

(where $F^{\mu\nu} \equiv F_{\alpha\beta}g^{\alpha\mu}g^{\beta\nu}$), namely, the usual Maxwell stress tensor. For any power α^n, we will have

$$\delta\int\alpha^n(-g)^{\frac{1}{2}}\,d^4x$$
$$= \int\alpha^{n-1}(-nF^{\mu\alpha}F^{\nu}{}_{\alpha} + \tfrac{1}{2}g^{\mu\nu}\alpha)(-g)^{\frac{1}{2}}\delta g_{\mu\nu}\,d^4x. \tag{14b}$$

For β^2, which is a scalar density of weight two, we have

$$\delta\int\beta^2(-g)^{-\frac{1}{2}}\,d^4x = -\tfrac{1}{2}\int\beta^2(-g)^{-\frac{1}{2}}g^{\mu\nu}\delta g_{\mu\nu}\,d^4x \tag{14c}$$

since the only metric dependence is in the factor $(-g)^{-\frac{1}{2}}$. Likewise,

$$\delta\int\beta^{2n}(-g)^{-n+\frac{1}{2}}\,d^4x$$
$$= (-n + \tfrac{1}{2})\int\beta^{2n}(-g)^{-n+\frac{1}{2}}g^{\mu\nu}\delta g_{\mu\nu}\,d^4x. \tag{14d}$$

In particular, such terms will not contribute to either T^{0k} or $\delta\mathfrak{T}^{0k}/\delta g_{00}$ at $g_{\mu\nu} = \eta_{\mu\nu}$. To compute the $\delta\mathfrak{T}/\delta g$ in general, we must rewrite the α, β, and $F^{\mu}{}_{\alpha}F^{\nu\alpha}$ in terms of $(\mathcal{E}, \mathfrak{B})$, keeping only a weak g_{00} in the present context. Thus, we let $g^{00} \to (g_{00})^{-1}$ and $(-g)^{\frac{1}{2}} \to (-g_{00})^{\frac{1}{2}}$. Although only the g_{00} dependence will be needed at this time, we will display the full N_μ dependence for future reference:

$$F_{0i} = -\not{\mathcal{E}}^{0j}g_{ij}N({}^3g)^{-\frac{1}{2}} + F_{ij}N^j, \tag{15a}$$

$$F^{ij} = (\not{\mathcal{E}}^{0i}N^j - \not{\mathcal{E}}^{0j}N^i)N^{-1}({}^3g)^{-\frac{1}{2}} + F_{kl}\,{}^3g^{ki}\,{}^3g^{lj}, \tag{15b}$$

where $\not{\mathcal{E}}^{0i} \equiv (-g)^{\frac{1}{2}}F^{0i} = \mathcal{E}^i$. In terms of the canonical variables, we find

$$\alpha \sim \mathfrak{B}^2 - \mathcal{E}^2,$$
$$\beta \sim F_{0i}F_{jk}\epsilon^{0ijk} \sim g_{00}(-g)^{-\frac{1}{2}}\mathcal{E}\cdot\mathfrak{B} \sim (-g_{00})^{\frac{1}{2}}\mathcal{E}\cdot\mathfrak{B}, \tag{16}$$
$$F^{0\alpha}F^{0}{}_{\alpha} \sim \mathcal{E}^2(-g_{00})^{-1},$$

and

$$F^{0\alpha}F^{k}{}_{\alpha} \sim (\mathcal{E}\times\mathfrak{B})^k(-g_{00})^{-\frac{1}{2}}.$$

We can now check whether $\mathfrak{T}^{0\mu}$ has the correct g_{00} dependence to satisfy Eq. (8). Since α and $\beta(-g_{00})^{-\frac{1}{2}}$ are g_{00} independent, any g_{00} dependence will come from the explicit metrics in (14) once the $\mathcal{E}$ and $\mathfrak{B}$ are substituted in. It is easy to verify that Eq. (8) is indeed satisfied for any $\alpha^n\beta^{2m}$ term of the original action. For the Born–Infeld action, we can do this in closed form:

$$\mathfrak{T}_B^{\mu\nu} = [\mathfrak{T}_M^{\mu\nu} - g^{\mu\nu}(\tfrac{1}{2}\alpha + 1)(-g)^{\frac{1}{2}}]\gamma^{-1} + (-g)^{\frac{1}{2}}g^{\mu\nu}, \tag{17}$$

where $\mathfrak{T}_M^{\mu\nu}$ is the Maxwell tensor and

$$\gamma = (1 + \alpha + \beta^2 g^{-1})^{\frac{1}{2}}. \tag{18}$$

It follows that

$$T_B^{00} = (1 + \mathfrak{B}^2)\gamma^{-1} - 1,$$
$$T_B^{0i} = (\mathcal{E}\times\mathfrak{B})^i\gamma^{-1} \tag{19}$$

have the correct $[T_B^{00}, T_B^{00'}]$ relation. This has also been done directly for the Euler Lagrangian, Eq. (10), with the appropriate $\mathbf{\Pi}$ and CCR. One may also verify the Born–Infeld T^{00} commutator using the correct $\mathbf{\Pi}$, namely,

$$\Pi^i = -(\mathcal{E}^i - \beta\mathfrak{B}^i)\gamma^{-1},$$
$$[\Pi^i(\mathbf{r}), \mathfrak{B}^j(\mathbf{r}')] = i\epsilon^{ijk}\partial_k\delta(\mathbf{r} - \mathbf{r}'). \tag{20}$$

It is, in fact, $\mathbf{\Pi}$ rather than $\mathcal{E}$ which now satisfies the Gauss equation, $\nabla\cdot\mathbf{\Pi} = 0$, and is transverse; this will be clearer when we treat the theory in first-order form later. The momentum density now has the simple form

$$T_B^{0k} = (\mathfrak{B}\times\mathbf{\Pi})^k \sim -\Pi^i\partial^k A_i + \partial_i(\Pi^i A^k) \tag{21}$$

(one could have arrived at the $\gamma^{-1}\mathcal{E}$ part of $\mathbf{\Pi}$ from the necessity that T^{0k} retain its simple form). To express T_B^{00} in terms of $\mathbf{\Pi}$, one simply uses the identities

$$1 + \mathbf{\Pi}^2 = \gamma^{-2}(1 + \beta^2)(1 + \mathfrak{B}^2),$$
$$(\mathbf{\Pi}\cdot\mathfrak{B})^2 = \gamma^{-2}\beta^2(1 + \mathfrak{B}^2)^2 \tag{22}$$

to obtain

$$T_B^{00} = [(1 + \mathbf{\Pi}^2)(1 + \mathfrak{B}^2) - (\mathbf{\Pi} \cdot \mathfrak{B})^2]^{\frac{1}{2}} - 1$$
$$\equiv [1 + \mathbf{\Pi}^2 + \mathfrak{B}^2 + (\mathbf{\Pi} \times \mathfrak{B})^2]^{\frac{1}{2}} - 1, \quad (23)$$

which is, aside from a constant, the form given by Dirac.[1] It is easy to see that the Poisson bracket of T_B^{00} with itself gives the desired $T^{0i}\partial_i\delta(\mathbf{r})$ structure. As Dirac remarked, the difference between the Poisson bracket treatment and a correct use of commutators lies in ordering problems, but these involve only $[\mathbf{\Pi}(\mathbf{0}), \mathfrak{B}(\mathbf{0})] \sim \partial\delta(\mathbf{0})$, which may be dropped by oddness. There will be difficulties when other interactions (e.g., a dynamical metric field), are present in these terms, or for non-Abelian fields.

The discussion is unaltered for massive fields: a mass term in the action,

$$W_m \sim \int A_\mu A_\nu g^{\mu\nu}(-g)^{\frac{1}{2}}\, d^4x, \quad (24)$$

gives the following contribution to the stress tensor:

$$\mathfrak{T}_m^{\mu\nu} \sim (-A^\mu A^\nu + \tfrac{1}{2}g^{\mu\nu}A_\alpha A^\alpha)(-g)^{\frac{1}{2}}. \quad (25)$$

The appropriate Cauchy variables are not A_μ, but A_i and $\mathcal{A}^0 \equiv (-g)^{\frac{1}{2}}A^0$, the covariant spatial and contravariant density time components. [The necessity of using $\mathcal{A}^0$ may be seen from the constraint equation $\partial_i \mathcal{F}^{0i} = m^2\mathcal{A}^0$, since $\mathcal{F}^{0i} \equiv (-g)^{\frac{1}{2}}F^{0i}$ is the correct variable; any other choice, say A_0, would be expressed in terms of $\mathcal{F}^{0i}$ and the metric g_{00}.] The contribution of the mass term to $\mathfrak{T}^{0\mu}$ is

$$(-g_{00})^{\frac{1}{2}}\mathfrak{T}_m^{00} \sim (\mathcal{A}^0)^2 + \mathbf{A}^2,$$
$$\mathfrak{T}_m^{0k} \sim \mathcal{A}^0 A^k, \quad (26)$$

where both terms are g_{00} independent.[15] Likewise, any power of A^2 in the action, corresponding to direct self-coupling, also satisfies this condition, since each further power behaves as

$$A_\mu A_\nu g^{\mu\nu} \sim A_0 g^{00} + \mathbf{A}^2 = -(\mathcal{A}^0)^2 + \mathbf{A}^2,$$

which is metric independent. The same holds for terms like $A_\mu F^\mu{}_\alpha F^{\alpha\nu}A_\nu$, etc., so we may conclude that (to within quantum-ordering problems) the Lorentz invariance condition, Eq. (1), is satisfied irrespective of the self-coupling for vector theories.

In the above, we have concentrated on g_{00} dependence only. For the other $T^{\mu\nu}$ commutators, we would need N_i and g_{ij} dependence as well. The former can be traced through in general; the g_{ij} will, of course, occur in many places (it also enters in the $\mathbf{\Pi}$-$\mathfrak{E}$ relation) and depends very much on the model. Fortunately, the g_{ij} dependence is only relevant to the commutator $[T^{0i}, T^{lm'}]$.

[15] Note that, in the massive case, there is a "conspiracy" between the $\mathfrak{E} \times \mathfrak{B}$ and $\mathcal{A}^0A^k = \nabla \cdot \mathfrak{E}A^k$ terms to yield the same form as in electrodynamics,

$$T^{0k} = \mathfrak{E}^i\partial^k A_i - \partial_i(\mathfrak{E}^i A^k),$$

namely, a $\pi\nabla\phi$ term together with a metric-independent divergence.

IV. FIRST-ORDER FORMALISM

We have seen in Sec. III, using second-order formalism, that, for self-coupled spin-one fields, the energy-density commutator, Eq. (1), follows from the weak g_{00} metric dependence of the Lagrangian by means of Eq. (12). However, it was shown in Sec. II that a more powerful statement can be made if the Hamiltonian form obeys Eq. (3): All the commutators given in (6) then follow. We will now demonstrate that Eq. (3) does indeed hold for the models given.

First, we consider the Born–Infeld theory.[6] It is shown in Appendix B that, in first-order form, the Lagrangian is given in flat space by

$$\mathfrak{L}_B = -\tfrac{1}{2}\gamma^{-1}(F^{\mu\nu} - \tfrac{1}{2}\beta\tilde{F}^{\mu\nu})(\partial_\mu A_\nu - \partial_\nu A_\mu) + S(\alpha, \beta),$$
$$\gamma = (1 + \alpha - \beta^2)^{\frac{1}{2}}, \quad (27)$$
$$S(\alpha, \beta) = -\gamma + \alpha\gamma^{-1} - 2\beta^2\gamma^{-1} + 1.$$

The metric dependence is again obtained by the prescription $\eta_{\mu\nu} \to g_{\mu\nu}$, with appropriate factors of $(-g)^{\frac{1}{2}}$ introduced to make $\mathfrak{L}_B$ a scalar density. The resulting Lagrangian becomes

$$\mathfrak{L}_B = -\tfrac{1}{2}\Pi^{\mu\nu}(\partial_\mu A_\nu - \partial_\nu A_\mu) + (-g)^{\frac{1}{2}}S(\alpha, \beta(-g)^{-\frac{1}{2}}),$$
$$\Pi^{\mu\nu} \equiv (-g)^{\frac{1}{2}}\gamma^{-1}[F^{\mu\nu} + \beta\tilde{F}^{\mu\nu}(2g)^{-1}],$$
$$S(\alpha, \beta(-g)^{-\frac{1}{2}}) = -\gamma^{-1}(1 - \beta^2 g^{-1}) + 1, \quad (28)$$
$$\gamma = (1 + \alpha + \beta^2 g^{-1})^{\frac{1}{2}}.$$

By construction the field equations are as in the second-order form, with the difference that $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is now a field equation on a par with $\partial_\mu \Pi^{\mu\nu} = 0$ and is obtained by varying $F^{\mu\nu}$ (independently of A_μ). The variable conjugate to A_i is $-\Pi^i$, where

$$\Pi^{0i} \equiv \Pi^i = \gamma^{-1}[\mathfrak{E}^i - \mathfrak{B}^i\beta(-g)^{-\frac{1}{2}}]. \quad (29)$$

By using the constraints, $\mathfrak{L}_B$ may be expressed completely in terms of canonical variables, A_i (or $\mathfrak{B}^i$), Π^i, and (N, N_i), with the spatial metric g_{ij} present where needed to tie together spatial indices. Notice that Eq. (29), the relation of Π^i to $\mathfrak{E}^i$ and $\mathfrak{B}^i$, is $g_{0\mu}$ independent, as both α and $(-g)^{-\frac{1}{2}}\beta$ are $g_{0\mu}$ independent when written in terms of $(\mathfrak{E}^i, \mathfrak{B}^j)$. This is also apparent when we express the various quantities

in terms of $\mathbf{\Pi}$. In particular, we find

$$\beta(-g)^{-\frac{1}{2}} = \gamma(\mathfrak{B}\cdot\mathbf{\Pi})({}^3g)^{-1}[1-\mathfrak{B}^2({}^3g)^{-1}]^{-1}, \tag{30a}$$

$$\alpha = -\gamma^2\mathbf{p}^2({}^3g)^{-1} + \mathfrak{B}^2({}^3g)^{-1}, \tag{30b}$$

$$\gamma^2 = [1+\mathfrak{B}^2({}^3g)^{-1}]^2\{[1+\mathfrak{B}^2({}^3g)^{-1}] \times [1+\mathbf{\Pi}^2({}^3g)^{-1}] - (\mathfrak{B}\cdot\mathbf{\Pi})^2({}^3g)^{-2}\}^{-1}, \tag{30c}$$

$$\Pi^{ij} = (\Pi^i\,{}^3g^{lj} - \Pi^j\,{}^3g^{li})N_l + NH_{lm}\,{}^3g^{li}\,{}^3g^{mj}, \tag{30d}$$

$$H_{lm} \equiv -\gamma^{-1}\epsilon_{lmi}({}^3g)^{-\frac{1}{2}}[\mathfrak{B}^i + (-g)^{-\frac{1}{2}}\beta\gamma p^i], \tag{30e}$$

$$p^i = -\Pi^i + \mathfrak{B}^i(\mathfrak{B}\cdot\mathbf{\Pi})({}^3g)^{-1}[1+\mathfrak{B}^2({}^3g)^{-1}]^{-1}. \tag{30f}$$

Clearly, α, $(-g)^{-\frac{1}{2}}\beta$, γ, H_{lm}, and p^i are independent of $g_{0\mu}$ and, consequently, so is $S(\alpha, (-g)^{-\frac{1}{2}}\beta)$. The term $\int \Pi^i\partial_i A_0$ in the Born–Infeld action vanishes because Π^i is transverse ($\nabla\cdot\mathbf{\Pi}=0$). Thus, the Lagrangian can be written in the desired form, $\mathfrak{L} = \Pi^i\partial_0 A_i + N\theta_0^0 + {}^3g^{ij}N_i\theta_j^0$, where

$$\begin{aligned}\theta_i^0 &= \epsilon_{ijk}\mathfrak{B}^j\Pi^k({}^3g)^{-1}\\ &= \Pi^k(\partial_k A_i - \partial_i A_k)\\ &= -\Pi^k\partial_i A_k + \partial_k(\Pi^k A_i),\end{aligned} \tag{31a}$$

$$\theta_0^0 = \tfrac{1}{2}H_{lm}\,{}^3g^{li}\,{}^3g^{mj}\epsilon_{ijk}\mathfrak{B}^k({}^3g)^{-1} + ({}^3g)^{\frac{1}{2}}S(\alpha, (-g)^{-\frac{1}{2}}\beta). \tag{31b}$$

The momentum density $T_i^0 = \theta_i^0$ ($g=\eta$) has the correct form $\mathfrak{B}\times\mathbf{\Pi}$ given by Dirac and, furthermore, θ_i^0 is entirely metric independent, which is required to give the simple form in Eq. (7d) for the commutator $[T^{00}, T^{ij'}]$. The energy density, θ_0^0, is indeed independent of $g_{0\mu}$, but has a great deal of g_{ij} dependence:

$$\theta_0^0 = -({}^3g)^{\frac{1}{2}}\{[1+\mathbf{\Pi}^2({}^3g)^{-1}][1+\mathfrak{B}^2({}^3g)^{-1}] - (\mathfrak{B}\cdot\mathbf{\Pi})^2({}^3g)^{-2}\}^{\frac{1}{2}} + ({}^3g)^{\frac{1}{2}}, \tag{32}$$

where $\mathbf{\Pi}^2 \equiv \Pi^i\Pi^j g_{ij}$, etc., and, of course, $\theta_0^0(\eta) = -T^{00}$ is just the Dirac form, Eq. (23). One could now calculate the model-dependent commutator $[T^{0i}, T^{jk'}]$ for the Born–Infeld theory, but the result is not very instructive; it does not have the simple Maxwell form of Appendix A.

The various stress-tensor commutators may also be calculated directly by replacing the commutators with Poisson brackets. We have done this explicitly for the $[T^{0\mu}, T^{0\nu'}]$ and find agreement with the general form.[16]

One may check explicitly that the above method and results also apply to an action which is an arbitrary series of terms $\alpha^n\beta^{2m}$, using the appropriate first-order form given by Eq. (B2).

[16] The commutator $[T^{0i}, T^{0j'}]$ may actually be calculated in terms of the CCR because of the simple form of T^{0i}.

Another model we have considered is the Yang–Lee theory[17] of a charged vector meson coupled nonminimally to the electromagnetic field[18]:

$$\mathfrak{L}_{\rm Int} = i\lambda(-g)^{\frac{1}{2}}F^{\mu\nu}(\phi_\mu^*\phi_\nu - \phi_\nu^*\phi_\mu). \tag{33}$$

The canonical variables are unchanged by the coupling, but the constraint equation for $F_{\mu\nu}$ is altered:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + i\lambda(\phi_\mu^*\phi_\nu - \phi_\nu^*\phi_\mu). \tag{34}$$

Nevertheless, it can be shown by means of Eq. (8) that this model also satisfies the Dirac–Schwinger condition.

Finally, we mention that actions arising from a scalar (ϕ, π^μ)-vector $(A_\mu, G^{\mu\nu})$ theory with gradient "couplings," e.g., $\mathfrak{L}_{\rm Int} = \lambda(-g)^{\frac{1}{2}}A_\mu\pi^\mu$, also have the form given by Eq. (3).

V. SUMMARY

We have seen that the simple Lorentz invariance condition (1) on the energy-density commutator is valid for arbitrary derivative or direct coupling of spin-zero and spin-one systems and is, thus, a property of the low spin rather than of the details of the interaction, as long as ordering problems are absent. The form of the full set of stress-tensor commutators was shown to be determined explicitly for systems having the canonical form $W = \int[\pi\dot\phi + N^\mu\theta_\mu^0(\pi, \phi, g_{ij})]\,d^4\chi$. This was exemplified by the Born–Infeld and other variants of electrodynamics.

APPENDIX A

We show that, for Maxwell theory, the model-dependent commutator $[T^{0k}, T^{mn'}]$ may be written entirely in terms of $T^{\mu\nu}$ itself. We start from Eq. (7e) given in the text:

$$\begin{aligned}&i[T^{0i}(\mathbf{x}), T^{lm}(\mathbf{x}')]\delta(x^0 - x'^0)\\ &= [T^{lm}(\mathbf{x})\partial_i - \delta^{il}T^{mk}(\mathbf{x}')\partial_k - \delta^{im}T^{lk}(\mathbf{x}')\partial_k]\delta(x-x')\\ &\quad - 4\partial_k\left(\frac{\delta^2\theta_0^0}{\delta g_{lm}(x)\delta g_{ik}(x')}\right).\end{aligned} \tag{A1}$$

For Maxwell theory, we have

$$\theta_0^0 = -\tfrac{1}{2}({}^3g)^{-\frac{1}{2}}(\mathfrak{B}^2 + \mathfrak{E}^2),$$

$$\theta_i^0 = (\mathfrak{E}\times\mathfrak{B})^i({}^3g)^{-1} = \mathfrak{E}^j(\partial_i A_j - \partial_j A_i).$$

[17] C. N. Yang and T. D. Lee, Phys. Rev. **128**, 885 (1962).

[18] The above coupling in second-order formalism is a different theory. It is equivalent to a first-order theory with a coupling

$$\mathfrak{L}_{\rm Int} = i\lambda'(-g)^{\frac{1}{2}}(\partial_\mu A_\nu - \partial_\nu A_\mu)(\phi^{\mu*}\phi^\nu - \phi^{\nu*}\phi^\mu),$$

in which the constraint equation for $F_{\mu\nu}$ is the normal one, but now the variable conjugate to A_i is no longer $(-g)^{\frac{1}{2}}F^{0i}$. A slight modification of the coupling, which also obeys the Dirac–Schwinger condition (1), is that of a self-coupled, massive, charged vector field:

$$\mathfrak{L}_{\rm Int} = i\lambda''(-g)^{\frac{1}{2}}G^{\mu\nu}(\phi_\mu\phi_\nu^* - \phi_\nu\phi_\mu^*).$$

We note that $\delta\theta_i^0/\delta g_{mn}$ vanishes, while, after some manipulation, we find

$$\frac{\delta^2\theta_0^0}{\delta g_{lm}(x)\delta g_{ik}(x')} = -\tfrac{1}{4}[(\delta^{ik}T^{lm} + \delta^{lm}T^{ik}) + T^{00}(+\delta^{li}\delta^{mk} + \delta^{lk}\delta^{mi} - \delta^{ik}\delta^{lm})]\delta(x - x'), \tag{A2}$$

which yields the expression of Footnote 12, when inserted into (A1). A check is obtained by direct use of CCR, together with $\nabla \cdot \mathfrak{E} = 0 = \nabla \cdot \mathfrak{B}$. One may also check that the commutator of T^{0i} with T^{mm} is the same as that with T^{00}, in agreement with the tracelessness of $T^{\mu\nu}$ in the Maxwell case.

The commutator $[T^{ij}(\mathbf{r}), T^{lm}(\mathbf{r}')]$ is not relevant to the Poincaré algebra. It can neither be expressed in terms of $\delta\mathfrak{T}/\delta g$, nor is it a function of $T^{\mu\nu}$ alone, since it involves terms like $\mathfrak{E}^i B^m \epsilon^{jlk}\partial_k\delta$, which,unlike $\mathfrak{E}^i\mathfrak{E}^j$ or $\epsilon_{ijk}\mathfrak{E}^j B^k$, cannot be written in terms of $T^{\mu\nu}$ (although $[T^{ij}(\mathbf{r}), T^{ll}(\mathbf{0})] = [T^{ij}(\mathbf{r}), T^{00}(\mathbf{0})]$ does depend on $T^{\mu\nu}$ only).

APPENDIX B

We obtain here the first-order action (in which $F^{\mu\nu}$ and A_μ are to be varied independently), corresponding to an arbitrary second-order form

$$W = \int \mathfrak{L}(\alpha, \beta)\, d^4x,$$

in which $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$, $\alpha \equiv \frac{1}{2}F_{\mu\nu}F_{\alpha\beta}\eta^{\mu\alpha}\eta^{\nu\beta}$, and $\beta \equiv \frac{1}{8}\bar{F}^{\mu\nu}F_{\mu\nu}$. We thus wish to obtain a Lagrangian $\bar{\mathfrak{L}}(F^{\mu\nu}, \partial_\mu A_\nu - \partial_\nu A_\mu)$, which yields both the homogeneous and usual Maxwell equations as *field* equations upon varying with respect to $F^{\mu\nu}$ and A_μ, respectively. The Maxwell equations are just

$$\partial_\mu(\mathfrak{L}_\alpha F^{\mu\nu} + \tfrac{1}{4}\mathfrak{L}_\beta \bar{F}^{\mu\nu}) = 0, \tag{B1}$$

where $\mathfrak{L}_{\alpha,\beta} \equiv \delta\mathfrak{L}/\delta\alpha, \beta$. We, therefore, take

$$\bar{\mathfrak{L}} \equiv (\mathfrak{L}_\alpha F^{\mu\nu} + \tfrac{1}{4}\mathfrak{L}_\beta \bar{F}^{\mu\nu})(\partial_\mu A_\nu - \partial_\nu A_\mu) + S(\alpha, \beta). \tag{B2}$$

This guarantees the Maxwell equation (B1) upon varying A_μ. To determine S, we vary with respect to $F^{\mu\nu}$ and demand that the result implies $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. It is easily checked that this is the case with

$$S(\alpha, \beta) = \mathfrak{L} - 2(\alpha\mathfrak{L}_\alpha + \beta\mathfrak{L}_\beta). \tag{B3}$$

In the Maxwell case, $\mathfrak{L} = -\frac{1}{2}\alpha$, so that

$$\bar{\mathfrak{L}} = -\tfrac{1}{2}F^{\mu\nu}(\partial_\mu A_\nu - \partial_\nu A_\mu) + \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu},$$

the usual form.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Correlations in Plasma. I

BRAHMANANDA DASGUPTA AND PADMANABHA DASGUPTA
Saha Institute of Nuclear Physics, Calcutta-9, India

(Received 15 August 1968)

We derive for nonequilibrium isotropic plasma the correlations of fluctuation in electric-field intensity and current density. The calculation is based on the ring-approximation results of Balescu and Taylor. The results reduce to the results obtained by Rostoker in the special case when the velocity-distribution function is Maxwellian at the initial time.

1. INTRODUCTION

This work attempts to derive in the "ring approximation" the correlation functions with a plasma in general nonequilibrium states.

The importance of correlations of fluctuations obtains from the fact that these can be directly related to the results of typical probe measurement[1] on a plasma and they appear as the kernel function in the transport equation.[2] In a well-known paper, Rostoker[3] derived expressions for the fluctuations of electric-field and current density within a plasma. In his derivation, Rostoker introduces the joint probability distribution function which is indeed necessary for a correct definition of the correlation function. The calculation proceeds by taking successive moments of the Liouville equation and subsequently truncating the chain by using an expansion in the plasma parameters. Instead of proceeding that way, we shall base our calculations on the very elegant resolvent formalism developed by Prigogine, Balescu, and their group.[4,5] The advantage of this formalism lies in the fact that it allows one to deal with the formal solution of the exact Liouville equation and dominant terms can be selected out using certain simple prescriptions. It has been demonstrated by Balescu[5] that, in the case of plasma, dominant terms are those arising from the so-called "ring diagrams" (ring approximation), the summation of which is equivalent to a renormalization of the interaction potential, thereby eliminating the divergence due to the long-range nature of Coulomb forces.

In this paper we present the calculations of the electric-field and the current-density correlation functions for a plasma in absence of external fields. Calculations of these functions for plasma subjected to external magnetic field will be reported in a subsequent publication. In Sec. 2 below we discuss the definitions of these functions, while Secs. 3 and 4 are devoted to the actual derivations.

2. DEFINITION OF ELECTRIC-FIELD CORRELATION AND CURRENT-DENSITY CORRELATION

A. Electric-Field Correlation

The electric-field correlation is defined as

$$\langle E_\mu(\mathbf{x}, t)E_\nu(\mathbf{x}', t')\rangle = e^2 \sum_{i,j} \iint dx\, dv\, dx'\, dv' \times \frac{\partial}{\partial x_\mu} \frac{1}{|\mathbf{x} - \mathbf{x}_i|} \frac{\partial}{\partial x'_\nu} \frac{1}{|\mathbf{x}' - \mathbf{x}'_j|} D_2(x, v, t \mid x', v', t'), \tag{2.1}$$

where $D_2(x, v, t \mid x', v', t')$ is the joint probability function such that $D_2(x, v, t \mid x', v', t')\, dx\, dv\, dx'\, dv'$ is the probability that, at time t, the system will be within $dx\, dv$ at (x, v), and at time t', the system will be in $dx'\, dv'$ at (x', v'). x and v denote the totality of position and velocity coordinates for N particles. For a stationary random process, D_2 is a function of $t - t'$, and it is related to the N-particle distribution function $f_N(x, v, t)$ through the relation

$$f_N(x, v, t) = \int dx'\, dv' D_2(x, v, t \mid x', v', 0). \tag{2.2}$$

The N-particle distribution function satisfies the Liouville equation

$$\mathfrak{L} f_N(x, v, t) = 0, \tag{2.3}$$

which has the formal solution given by

$$f_N(x, v, t) = \int dx'\, dv' \mathfrak{G}(x, v, t \mid x', v', 0) f_N(x', v', 0), \tag{2.4}$$

where $\mathfrak{G}$ is the Green's function of the Liouville operator obeying the relation

$$\mathfrak{L}\mathfrak{G}(x, v, t \mid x', v', t') = \delta(x - x')\delta(v - v')\delta(t - t'). \tag{2.5}$$

[1] K. L. Dowles, Phys. Rev. Letters **1**, 454 (1958).
[2] T. Pradhan and B. Dasgupta, Phys. Rev. **160**, 184 (1967).
[3] Norman Rostoker, Nucl. Fusion **1**, 101 (1961).
[4] I. Prigogine, *Non-Equilibrium Statistical Mechanics* (Interscience Publishers, Inc., New York, 1962).
[5] R. Balescu, *Statistical Mechanics of Charged Particles* (Interscience Publishers, Inc., New York, 1963); references to individual papers are to be found here and in Ref. 4.

The Green's functions satisfy the causality condition

$$\mathcal{G}(x, v, t \mid x', v', t') = 0 \quad \text{for} \quad t < t'.$$

Now, we write expression (2.1) as

$$\begin{aligned}\langle E_\mu(\mathbf{x}, t)E_\nu(\mathbf{x}', 0)\rangle &= \mathfrak{E}_{\mu\nu}(\mathbf{x}, \mathbf{x}', t) \\ &= e^2 \sum_{i,j} \iint dx\, dv\, dx'\, dv' E_\mu(|\mathbf{x} - \mathbf{x}_i|)E_\nu(|\mathbf{x}' - \mathbf{x}_j'|) \\ &\quad \times \mathcal{G}(x, v, t \mid x', v', 0) f_N(x', v', 0). \qquad (2.6)\end{aligned}$$

Following Balescu's notation, we write the Laplace transform of (2.6) as

$$\begin{aligned}&\mathfrak{E}_{\mu\nu}(\mathbf{x}, \mathbf{x}', t) \\ &= \frac{e^2}{2\pi} \int d\omega \exp(-i\omega t) \sum_{i,j} \int dx\, dv \\ &\quad \times E_\mu(|\mathbf{x} - \mathbf{x}_i|)E_\nu(|\mathbf{x}' - \mathbf{x}_j'|)\mathcal{R}(\omega) f_N(x, v, 0),\end{aligned}$$

where

$$\mathcal{R}(\omega) F(x, v) = \int dx'\, dv' \langle x, v| \mathcal{R}(\omega) |x', v'\rangle F(x', v').$$

The integral equation for $\mathcal{R}(\omega)$ is

$$\mathcal{R}(\omega) = \mathcal{R}^0(\omega) - e^2\mathcal{R}^0(\omega)\mathfrak{L}'\mathcal{R}(\omega), \qquad (2.7)$$

which can be formally solved by a power series in e^2. In Eq. (2.7), $\mathcal{R}^0(\omega)$ is the value of $\mathcal{R}(\omega)$ for a system of free particles and $\mathfrak{L}'$ is the Liouville operator for the interaction part of the Hamiltonian.

Now introducing the Fourier components

$$f_N(x, v, t) = \sum_{\{k\}} \rho_{\{k\}}(v, t) |\{k\}\rangle, \qquad (2.8)$$

with

$$|\{k\}\rangle = V^{-N/2} \exp\left(i \sum_j \mathbf{k}_j \cdot \mathbf{x}_j\right)$$

and

$$E_\mu(|\mathbf{x} - \mathbf{x}_i|) = \sum_{\mathbf{q}} \frac{iq_\mu}{q^2} \exp\{i\mathbf{q} \cdot (\mathbf{x} - \mathbf{x}_i)\}, \qquad (2.9)$$

we rewrite Eq. (2.6) as

$$\begin{aligned}&\mathfrak{E}_{\mu\nu}(\mathbf{x}, \mathbf{x}', t) \\ &= \frac{e^2}{2\pi} \sum_{i,j} \sum_{\mathbf{q}} \sum_{\mathbf{q}'} \sum_{\{k\}} \sum_{\{k'\}} \int d\omega \exp(-i\omega t) \int dx\, dv \left(-\frac{q_\mu q_\nu}{q^2}\right) \\ &\quad \times \exp\{i\mathbf{q} \cdot (\mathbf{x} - \mathbf{x}_i) + i\mathbf{q}' \cdot (\mathbf{x}' - \mathbf{x}_j')\} |\{k\}\rangle \\ &\quad \times \langle\{k\}| \mathcal{R}(\omega) |\{k'\}\rangle \rho_{\{k'\}}(v, 0). \qquad (2.10)\end{aligned}$$

Although in the above expressions the Fourier resolution is carried out in a finite volume, we intend to pass over to the limit $V \to \infty$, $N \to \infty$; $N/V \to$ const, later on.

We now impose the condition of homogeneity:

$$\begin{aligned}\mathfrak{E}_{\mu\nu}(\mathbf{x}, \mathbf{x}', t) &= \mathfrak{E}_{\mu\nu}(\mathbf{x} - \mathbf{x}', t) \\ &= \sum_{\mathbf{k}} \tilde{\mathfrak{E}}_{\mu\nu}(\mathbf{k}, t) \exp\{i\mathbf{k} \cdot (\mathbf{x} - \mathbf{x}')\}. \qquad (2.11)\end{aligned}$$

From (2.10) one can separate $\tilde{\mathfrak{E}}$ into two parts, namely $\tilde{\mathfrak{E}}^{(1)}$, which consists of terms $i \neq j$, and $\tilde{\mathfrak{E}}^{(0)}$, which consists of terms $i = j$, wherefore

$$\begin{aligned}&\tilde{\mathfrak{E}}_{\mu\nu}^{(1)}(\mathbf{k}, t) \\ &= e^2 n^2 \left(-\frac{k_\mu k_\nu}{k^2}\right) V^N \left(\frac{8\pi^3}{V}\right)^{-1} \sum_{\{k'\}} \int d\omega \exp(-i\omega t) \int dv \\ &\quad \times \langle 0, \cdots, \mathbf{k}, -\mathbf{k}, \cdots, 0| \mathcal{R}(\omega) |\{k'\}\rangle \rho_{\{k'\}}(v, 0). \qquad (2.12)\end{aligned}$$

One may further rewrite (2.12) in terms of $\rho_{\mathbf{k},-\mathbf{k}}(i, j, t)$, the Fourier transform of the pair-correlation function

$$\tilde{\mathfrak{E}}_{\mu\nu}^{(1)}(\mathbf{k}, t) = e^2 n^2 \left(-\frac{k_\mu k_\nu}{k^4}\right) \int d\mathbf{v}_i\, d\mathbf{v}_j \rho_{\mathbf{k},-\mathbf{k}}(i, j, t). \qquad (2.13)$$

B. Current-Density Correlation

The current density is defined as

$$\mathbf{j}(\mathbf{x}, t) = e \sum_i \mathbf{v}_i \delta(\mathbf{x} - \mathbf{x}_i(t)).$$

So the current-density correlation is

$$\begin{aligned}\mathfrak{J}_{\mu\nu}(\mathbf{x}, \mathbf{x}', t) &= \langle j_\mu(\mathbf{x}, t) j_\nu(\mathbf{x}', t')\rangle \\ &= e^2 \sum_{i,j} \iint dx\, dv\, dx'\, dv' v_{i\mu} \delta(\mathbf{x} - \mathbf{x}_i) v_{j\nu} \delta(\mathbf{x} - \mathbf{x}_j') \\ &\quad \times D_2(x, v, t \mid x', v', t'). \qquad (2.14)\end{aligned}$$

Defining the Fourier components

$$\mathfrak{J}_{\mu\nu}(\mathbf{x} - \mathbf{x}', t) = \sum_{\mathbf{k}} \tilde{\mathfrak{J}}_{\mu\nu}(\mathbf{k}, t) \exp\{i\mathbf{k} \cdot (\mathbf{x} - \mathbf{x}')\},$$

where we have again assumed the homogeneity condition, we can separate $\tilde{\mathfrak{J}}_{\mu\nu}$ into two parts, $\mathfrak{J}_{\mu\nu}^{(1)}$ and $\mathfrak{J}_{\mu\nu}^{(0)}$, the first one arising from pair correlation, exactly as in the case of the electric field.

Analogous to Eq. (2.10) for the electric-field case, we have in this case

$$\begin{aligned}&\tilde{\mathfrak{J}}_{\mu\nu}(\mathbf{k}, t) \\ &= \sum_{i,j} \sum_{\{k'\}} \int d\omega \exp(-i\omega t) \int dv v_{i\mu} v_{j\nu} \\ &\quad \times \langle 0 \cdots \mathbf{k}_i = \mathbf{k}, \mathbf{k}_j = -\mathbf{k} \cdots 0| \mathcal{R}(\omega) |\{k'\}\rangle \\ &\quad \times \rho_{\{k'\}}(v, 0). \qquad (2.15)\end{aligned}$$

$\tilde{\mathfrak{J}}^{(0)}$ is obtained from the terms with $i = j$ in (2.15) and the terms with $i \neq j$ give rise to $\tilde{\mathfrak{J}}^{(1)}$. Again, introducing the pair-correlation function $\rho_{\mathbf{k},-\mathbf{k}}$, we may write finally

$$\tilde{\mathfrak{J}}^{(1)}_{\mu\nu}(\mathbf{k}, t) = n^2 \int d\mathbf{v}_i \, d\mathbf{v}_j v_{i\mu} v_{j\nu} \rho_{\mathbf{k},-\mathbf{k}}(i, j, t). \quad (2.16)$$

3. CALCULATION OF ELECTRIC-FIELD CORRELATION

As it is explicit in Eqs. (2.13) and (2.16), we have to have complete knowledge of the pair-correlation function $\rho_{\mathbf{k},-\mathbf{k}}$ for calculating $\mathfrak{E}_{\mu\nu}$ and $\mathfrak{J}_{\mu\nu}$. This basic function has been obtained by Balescu and Taylor[6] in the so-called "ring approximation," and we shall use their result.

The perturbation expansion for $\rho_{\mathbf{k},-\mathbf{k}}$ reads

$$\begin{aligned}
&\rho_{\mathbf{k},-\mathbf{k}}(i, j, t) \\
&= \int' dv \int d\omega \exp(-i\omega t) \\
&\times \Bigg[\sum_{\{k'\}} \Big\{ \langle 0 \cdots \mathbf{k}, -\mathbf{k} \cdots 0 | \, \mathcal{R}^0(\omega) \, | \{k'\} \rangle \, \rho_{\{k'\}}(v, 0) \\
&+ \sum_{\{k''\}} \sum_{\{k''\}} \langle \cdots \mathbf{k}, -\mathbf{k} \cdots | \, \mathcal{R}^0(\omega) \, | \{k''\} \rangle \langle \{k''\} | \, \mathfrak{L}' \, | \{k'''\} \rangle \\
&\times \langle \{k'''\} | \, \mathcal{R}^0(\omega) \, | \{k'\} \rangle \, \rho_{\{k'\}}(v, 0) + \cdots \text{etc.} \Big\} \Bigg], \quad (3.1)
\end{aligned}$$

where the prime over the integral sign excludes the integration over $\mathbf{v}_i$ and $\mathbf{v}_j$. The terms occurring in (3.1) may be grouped according to their long-time behavior and this grouping is most conveniently done with the help of diagrams. As shown by Balescu, the dominant terms are those corresponding to the so-called terminal creation fragments, and, by retaining only these terms in (3.1), they constitute the ring approximation for $\rho_{\mathbf{k},-\mathbf{k}}$. Thus the actual summation of (3.1) proceeds in two steps: (i) summation of all possible ring diagrams ending with a particular creation fragment to obtain a terminal creation fragment, and (ii) summation of all possible terminal creation fragments (Fig. 1). This leads to the result

$$\begin{aligned}
\rho_{\mathbf{k},-\mathbf{k}}(i, j, t) = \delta^{ij}_{-}\{&d_{ji}\varphi(i)\varphi(j) + d_j\varphi(j)F(i) \\
&+ d_i^*\varphi(i)F^*(j)\}, \quad (3.2)
\end{aligned}$$

where

$$d_i = -\frac{\omega_p^2}{k^2} i\mathbf{k} \cdot \frac{\partial}{\partial \mathbf{v}_i}, \quad (3.3)$$

$$d_{ij} = -\frac{1}{8\pi^3 n}[d_i - d_j], \quad (3.4)$$

[6] R. Balescu and H. S. Taylor, Phys. Fluids 4, 85 (1961).

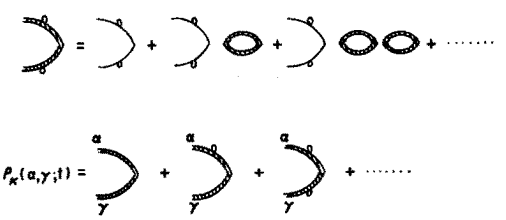


FIG. 1. Summation of terminal creation fragments. A typical terminal creation fragment is shown above.

$$\delta^{ij}_{-} = \pi\delta_{-}\{\mathbf{k} \cdot (\mathbf{v}_i - \mathbf{v}_j)\}, \quad (3.5)$$

$$\delta_{\pm}(x) = \delta(x) \pm \frac{i}{\pi} p \frac{1}{x}. \quad (3.6)$$

$\varphi(i)$ is the single-particle distribution function which satisfies the Balescu–Lenard kinetic equation. The function $F(i)$ satisfies the integral equation

$$\begin{aligned}
\epsilon(\alpha)F(\alpha) = q(\alpha) - \pi \, d_\alpha^* \varphi(\alpha) \\
\times \int d\mathbf{v}_1 \delta(\mathbf{k} \cdot \mathbf{v}_\alpha - \mathbf{k} \cdot \mathbf{v}_1) F^*(1), \quad (3.7)
\end{aligned}$$

with

$$\epsilon(\alpha) = 1 - \int d\mathbf{v}_1 \delta^{\alpha 1}_{-} d_1 \varphi(1), \quad (3.8)$$

$$q(\alpha) = \int d\mathbf{v}_1 \delta^{\alpha 1}_{-} d_{1\alpha} \varphi(\alpha)\varphi(1). \quad (3.9)$$

This integral equation can be solved in terms of the Vlassov (Van Kampen–Case) eigenfunctions $\bar{\chi}_{\nu'}(\nu)$:

$$F(\nu) = \int_{-\infty}^{+\infty} d\nu' \bar{\chi}_{\nu'}(\nu) \frac{\bar{q}(\nu')}{|\epsilon(\nu')|^2}. \quad (3.10)$$

The barred functions used in (3.10) are obtained as

$$\bar{f}(\nu) = \int_{-\infty}^{+\infty} d\mathbf{v} \delta\left(\nu - \frac{\mathbf{k} \cdot \mathbf{v}}{k}\right) f(\mathbf{v}). \quad (3.11)$$

The Vlassov eigenfunctions have the explicit form

$$\bar{\chi}_{\nu'}(\nu) = \frac{\epsilon_2(\nu)}{\pi} P \frac{1}{\nu - \nu'} + \epsilon_1(\nu)\delta(\nu - \nu'), \quad (3.12)$$

where $\epsilon_1(\nu)$ and $\epsilon_2(\nu)$ are the real and imaginary parts of ϵ, respectively. It can be shown that, for a stable plasma, the Vlassov eigenfunctions form a complete set with the normalization

$$\int_{-\infty}^{+\infty} d\nu \bar{\chi}_{\nu'}(\nu) = 1. \quad (3.13)$$

The adjoint eigenfunction is

$$\tilde{\chi}_{\nu'}(\nu) = P \frac{1}{\nu - \nu'} + \pi \frac{\epsilon_1(\nu)}{\epsilon_2(\nu)} \delta(\nu - \nu'), \quad (3.14)$$

with the orthogonality property

$$\int d\nu \bar{\chi}_{\nu'}(\nu)\tilde{\chi}_{\nu''}(\nu) = \pi \frac{|\epsilon(\nu')|^2}{\epsilon_2(\nu')}\delta(\nu' - \nu''); \quad (3.15)$$

From (3.2) and (3.7), we get

$$\tilde{\mathfrak{E}}^{(1)}_{\mu\nu}(k, t) = e^2n^2\left(-\frac{k_\mu k_\nu}{k^2}\right)\int d\nu \bar{F}(\nu). \quad (3.16)$$

Substitution of the solution (3.10), together with (3.13), yields

$$\tilde{\mathfrak{E}}^{(1)}_{\mu\nu}(k, t) = e^2n^2\left(-\frac{k_\mu k_\nu}{k^2}\right)\int \frac{d\omega}{k}\frac{\bar{q}(\omega)}{|\epsilon(\omega)|^2}. \quad (3.17)$$

We now evaluate the expression (3.17) at $t = 0$ with the assumption that at $t = 0$, $\varphi(v) = f_0(v)$, the Maxwellian velocity-distribution function. Then

$$\tilde{\mathfrak{E}}^{(1)}_{\mu\nu}(k, 0) = e^2n^2\left(-\frac{k_\mu k_\nu}{k^2}\right)\int \frac{d\omega}{k^2}\frac{q_0(\omega)}{|\epsilon^*_0(\omega)|^2}$$

$$= e^2n^2\left(-\frac{k_\mu k_\nu}{k^2}\right)g^0_k, \quad (3.18)$$

with

$$q^0(\omega) = -\frac{e^2\beta}{2\pi^2 k}\int d\mathbf{v}\delta(\omega - \mathbf{k}\cdot\mathbf{v})f_0(v), \quad (3.19)$$

$$\epsilon^*_0(\omega) = 1 - \frac{4\pi^2ne^2\beta}{k^2}\, i\omega\int d\mathbf{v}\delta_+(\omega - \mathbf{k}\cdot\mathbf{v})f_0(v). \quad (3.20)$$

From (3.19) and (3.20), one gets

$$q_0(\omega) = -\frac{k}{8\pi^4 n\omega}\,\mathrm{Im}\,\epsilon^*_0(\omega). \quad (3.21)$$

From (3.21) and (3.18), one obtains

$$g^0_k = \frac{1}{8\pi n}P\int \frac{d\omega}{k}\,\mathrm{Im}\frac{1}{\epsilon^*_0(\omega)}. \quad (3.22)$$

We note that the function $[\epsilon^*_0(k, \omega)]^{-1} - 1$ satisfies the Kramers–Kronig relation

$$\mathrm{Re}\left[\frac{1}{\epsilon^*_0(k, \omega)} - 1\right] = \frac{P}{\pi}\int \frac{d\omega'}{\omega' - \omega}\,\mathrm{Im}\frac{1}{\epsilon^*_0(k, \omega')}; \quad (3.23)$$

also

$$\mathrm{Re}\,\epsilon^*_0(k, \omega) = 1 + \frac{1}{(kL_D)^2}, \quad (3.24)$$

L_D being the Debye length. Therefore, from (3.22), (3.23), and (3.24), we get[7]

$$8\pi^3 n g^0_k = -\left[1 + \frac{(kL_D)^2}{\pi}P\int \frac{d\omega}{\omega}\,\mathrm{Im}\frac{1}{\epsilon^*_0(k, \omega)}\right]. \quad (3.25)$$

[7] B. Dasgupta and P. Dasgupta, Physica 32, 878 (1966).

The next step is to evaluate $\tilde{\mathfrak{E}}^{(0)}_{\mu\nu}(k, 0)$, which contains the contribution of the terms with $i = j$. Since these correspond to $\mathbf{k}_i = \mathbf{k}_j = 0$, the corresponding contribution may easily be evaluated to obtain

$$\tilde{\mathfrak{E}}^{(0)}_{\mu\nu}(k, 0) = \frac{e^2n}{8\pi^3}\left(-\frac{k_\mu k_\nu}{k^4}\right). \quad (3.26)$$

Hence the complete electric-field correlation is obtained as

$$\tilde{\mathfrak{E}}_{\mu\nu}(k, 0) = (4\pi e)^2 n\frac{k_\mu k_\nu}{k^4}\frac{(kL_D)^2}{\pi}P\int \frac{d\omega}{\omega}\,\mathrm{Im}\frac{1}{\epsilon^*_0(k, \omega)}. \quad (3.27)$$

Now comparing this with

$$\tilde{\mathfrak{E}}_{\mu\nu}(k, 0) = \frac{1}{2\pi}\int d\omega S_{\mu\nu}(k, \omega), \quad (3.28)$$

where $S_{\mu\nu}(k, \omega)$ is the spectral function,

$$S_{\mu\nu}(k, \omega) = 2(4\pi e)^2 n\frac{k_\mu k_\nu}{k^4}\frac{(kL_D)^2}{\omega}\,\mathrm{Im}\frac{1}{\epsilon^*_0(k, \omega)}. \quad (3.29)$$

Finally, we introduce a function

$$U_\pm(k, p) = \lim_{\lambda\to 0}\frac{k}{\pi}\int d\mathbf{v}\frac{f_0(v)}{p + i\mathbf{k}\cdot\mathbf{v} \pm \lambda}, \quad (3.30)$$

whence

$$\mathrm{Im}\,\epsilon^*_0(k, \omega) = \frac{\pi\omega}{k(kL_D)^2}\,\mathrm{Re}\,U_-(k, -i\omega). \quad (3.31)$$

In terms of this function, the spectral function is given by

$$S_{\mu\nu}(k, \omega) = (4\pi e)^2 n\frac{k_\mu k_\nu}{k^4}\frac{2\pi}{k}\frac{\mathrm{Re}\,U_-(k, -i\omega)}{|\epsilon_0(k, \omega)|^2}. \quad (3.32)$$

Having obtained this spectral function, we can compute

$$\tilde{\mathfrak{E}}_{\mu\nu}(k, t), \quad t > 0,$$

from the Laplace integral

$$\tilde{\mathfrak{E}}_{\mu\nu}(k, t) = \frac{1}{2\pi}\int_{i\omega_0-\infty}^{i\omega_0+\infty} d\omega e^{-i\omega t}S_{\mu\nu}(k, \omega), \quad (3.33)$$

where ω_0 must be chosen so as to keep the contour of integration C (Fig. 2) within the strip of holomorphy

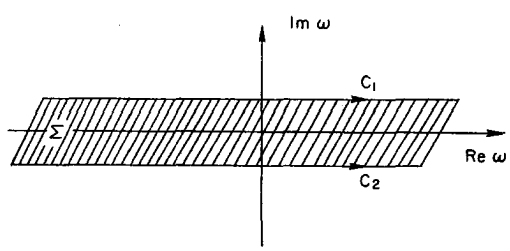


FIG. 2. Strip of holomorphy Σ.

Σ of $S_{\mu\nu}(k, \omega)$. One can separate $S_{\mu\nu}(k, \omega)$ into the so-called plus and minus functions in the following way:

$$S^{+}_{\mu\nu}(k, \omega) = \frac{S_{\mu\nu}(k, \omega)}{2} + \frac{i}{2\pi} P\int_{C_2} \frac{d\omega'}{\omega' - \omega} S_{\mu\nu}(k, \omega'), \tag{3.34}$$

$$S^{-}_{\mu\nu}(k, \omega) = \frac{S_{\mu\nu}(k, \omega)}{2} - \frac{i}{2\pi} P\int_{C_1} \frac{d\omega'}{\omega' - \omega} S_{\mu\nu}(k, \omega'), \tag{3.35}$$

$$S_{\mu\nu}(k, \omega) = S^{+}_{\mu\nu}(k, \omega) + S^{-}_{\mu\nu}(k, \omega). \tag{3.36}$$

While evaluating (3.33) for $t > 0$, one can close the contour of integration with a semicircle in the lower half ω plane because of the factor $\exp(-i\omega t)$. Since $S^{+}_{\mu\nu}$ is analytic in the upper half-plane above C_2, and $S^{-}_{\mu\nu}$ is analytic in the lower half-plane below C_1, the contribution from $S^{-}_{\mu\nu}$ vanishes. Thus

$$\tilde{\mathfrak{E}}_{\mu\nu}(k, t) = \frac{1}{2\pi}\int_C d\omega \exp(-i\omega t) S^{+}_{\mu\nu}(k, \omega), \quad t > 0. \tag{3.37}$$

The expression on the right side of (3.37) thus defines the casual correlation function. $S^{+}_{\mu\nu}(k, \omega)$ can be evaluated directly by substituting (3.32) in (3.34):

$$S^{+}_{\mu\nu}(k, \omega) = (4\pi e)^2 n \frac{k_\mu k_\nu}{k^4} (kL_D)^2 \times \int d\omega' \delta_{+}(\omega' - \omega) \frac{1}{\omega} \operatorname{Im} \frac{1}{\epsilon_0^*(k, \omega')}. \tag{3.38}$$

Using the once-subtracted dispersion relation

$$\operatorname{Re}\left[\frac{1}{\epsilon_0(k, \omega)} - \frac{1}{\epsilon_0(k, 0)}\right] = -\frac{\omega}{\pi} P\int \frac{d\omega'}{\omega'(\omega' - \omega)} \operatorname{Im} \frac{1}{\epsilon_0(k, \omega)} \tag{3.39}$$

and the relation

$$\epsilon_0(k, \omega) = \epsilon_0(k, 0) + \frac{i\omega\pi}{k(kL_D)^2} U_{-}(k, i\omega), \tag{3.40}$$

we finally get

$$S^{+}_{\mu\nu}(k, \omega) = (4\pi e)^2 n \frac{k_\mu k_\nu}{k^4} \frac{\pi}{k} \frac{U_{-}(k, i\omega)}{\epsilon_0(k, \omega)\epsilon_0(k, 0)}. \tag{3.41}$$

4. CALCULATION OF CURRENT-DENSITY CORRELATION

In order to calculate the current-density correlation, we substitute the expression for pair-correlation function (3.2) in (2.16). The resulting integral can be simplified greatly by transforming to a system with the vector $\mathbf{k}$ parallel to the polar axis. Such a transformation may be written as

$$v_\mu = \sum_\nu C_{\mu\nu} \bar{v}_\nu, \tag{4.1}$$

where $\bar{v}_\nu$ refers to the system with $\mathbf{k}$ as the third axis. The third column of the matrix $\|C_{\mu\nu}\|$ is given by

$$C_{\mu 3} = k_\mu / k. \tag{4.2}$$

Thus we write (2.16) in the form

$$\tilde{\mathfrak{J}}^{(1)}_{\mu\nu}(\mathbf{k}, t) = n^2 \sum_{\lambda,\sigma} C_{\mu\lambda} C_{\nu\sigma} \int d\bar{\mathbf{v}}_i \, d\bar{\mathbf{v}}_j \bar{v}_{i\lambda} \bar{v}_{j\sigma} \rho_{\mathbf{k},-\mathbf{k}}(\bar{\mathbf{v}}_i, \bar{\mathbf{v}}_j, t). \tag{4.3}$$

Making explicit use of the isotropy condition which guarantees that $\varphi(\mathbf{v})$ is an even function of individual Cartesian components of velocity, we notice that the only nonvanishing terms on the right side of (4.3) corresponds to $\lambda = 3$ and $\sigma = 3$, when we substitute the expression (3.2) for $\rho_k(\bar{\mathbf{v}}_i, \bar{\mathbf{v}}_j, t)$. Therefore, we have, using (4.2),

$$\tilde{\mathfrak{J}}^{(1)}_{\mu\nu}(\mathbf{k}, t) = n^2 \frac{k_\mu k_\nu}{k^2} \int d\bar{\mathbf{v}}_i \, d\bar{\mathbf{v}}_j \bar{v}_{i3} \bar{v}_{j3} \rho_k(\mathbf{v}_i, \mathbf{v}_j, t). \tag{4.4}$$

Introducing

$$\frac{\mathbf{k} \cdot \mathbf{v}_i}{k} = \bar{v}_{i3} = \nu, \quad \frac{\mathbf{k} \cdot \mathbf{v}_j}{k} = \bar{v}_{j3} = \nu',$$

$$R(\nu, \nu') = \int d\bar{\mathbf{v}}_i \, d\bar{\mathbf{v}}_j \rho_k(\bar{\mathbf{v}}_i, \bar{\mathbf{v}}_j, t) \times \delta\left(\nu - \frac{\mathbf{k} \cdot \bar{\mathbf{v}}_i}{k}\right) \delta\left(\nu' - \frac{\mathbf{k} \cdot \bar{\mathbf{v}}_j}{k}\right), \tag{4.5}$$

(4.4) may be written as

$$\tilde{\mathfrak{J}}^{(1)}_{\mu\nu}(k, t) = n^2 \frac{k_\mu k_\nu}{k^2} \int d\nu \, d\nu' \nu\nu' R(\nu, \nu'). \tag{4.6}$$

From (3.2) and (4.5), one gets

$$R(\nu, \nu') = T(\nu, \nu') + \frac{1}{\nu - \nu'} [\nu_p^2 \bar{\varphi}'(\nu) \bar{F}^*(\nu') - \nu_p^2 \bar{\varphi}'(\nu') \bar{F}(\nu)], \tag{4.7}$$

where

$$T(\nu, \nu') = -\frac{1}{8\pi^3 n} \cdot \frac{\nu_p^2}{\nu - \nu'} [\bar{\varphi}'(\nu') \bar{\varphi}(\nu) - \bar{\varphi}'(\nu) \bar{\varphi}(\nu')], \tag{4.8}$$

$$\nu_p^2 = \omega_p^2 / k^2. \tag{4.9}$$

Remembering $\epsilon_2(\nu) = \pi \nu_p^2 \bar{\varphi}'(\nu)$ and making use of the relations (3.12) and (3.14), we get from (4.7)

$$R(\nu, \nu') = T(\nu, \nu') + \bar{F}(\nu) \bar{\chi}_\nu(\nu') - \frac{1}{\pi} \bar{F}(\nu') \epsilon_2(\nu) \bar{\chi}_\nu(\nu'). \tag{4.10}$$

Therefore, from (4.6)

$$\tilde{\mathfrak{J}}^{(1)}_{\mu\nu}(k,t) = n^2 \frac{k_\mu k_\nu}{k^2}\Big[\int dv\Theta(v)v/n + \int dv\,dv'\bar{F}(v)\bar{\chi}_\nu(v')vv' - \frac{1}{\pi}\int dv\,dv'\bar{F}(v')vv'\epsilon_2(v)\tilde{\chi}_\nu(v')\Big], \quad (4.11)$$

with

$$\Theta(v) = n\int dv'T(v,v')v'. \quad (4.12)$$

We shall now make use of the following relations, satisfying the Vlassov eigenfunction and its adjoint:

$$v'\bar{\chi}_{v'}(v) = v\bar{\chi}_{v'}(v) - v_p^2\bar{\varphi}'(v), \quad (4.13)$$

$$v'\tilde{\chi}_v(v') = v\tilde{\chi}_v(v') + 1. \quad (4.14)$$

Equation (4.13) gives

$$\int vF(v)v'\bar{\chi}_v(v')\,dv\,dv' = \int dv\,dv'v^2\bar{F}(v)\bar{\chi}_v(v') + v_p^2\int dv\,dv'v\bar{F}(v)\bar{\varphi}'(v'). \quad (4.15)$$

The boundary condition at ∞, together with (3.13) and (3.10), may be used to obtain

$$\int dv\,dv'vv'\bar{F}(v)\bar{\chi}_v(v') = \int dv(v^2 - v_p^2)\frac{\bar{q}(v)}{|\epsilon(v)|^2}. \quad (4.16)$$

Equation (4.14) gives

$$\int dv\,dv'vv'\bar{F}(v')\epsilon_2(v)\tilde{\chi}_v(v') = \int dv\,dv'v^2\epsilon_2(v)\bar{F}(v')\tilde{\chi}_v(v') + \int dv\,dv'v\epsilon_2(v)\bar{F}(v'). \quad (4.17)$$

Using (3.10) and (3.15), one can derive

$$\int dv'\bar{F}(v')\tilde{\chi}_v(v') = \pi\frac{\bar{q}(v)}{\epsilon_2(v)}, \quad (4.18)$$

so that (4.17) may be written as

$$\int dv\,dv'vv'\bar{F}(v')\epsilon_2(v)\tilde{\chi}_v(v') = -\int dvv^2\bar{q}(v) + v_p^2\int dv\bar{F}(v). \quad (4.19)$$

Therefore,

$$\tilde{\mathfrak{J}}^{(1)}_{\mu\nu}(k,t) = n^2\frac{k_\mu k_\nu}{k^2}\int dv\Big[\frac{v}{n}\Theta(v) + v^2\frac{\bar{q}(v)}{|\epsilon(v)|^2} - v^2\bar{q}(v)\Big]. \quad (4.20)$$

The evaluation of $\tilde{\mathfrak{J}}^{(0)}_{\mu\nu}(k,t)$ proceeds by summing the terms with $i = j$ in (2.15):

$$\tilde{\mathfrak{J}}^{(0)}_{\mu\nu}(k,t) = n\int v_\mu v_\nu\varphi(\mathbf{v},t)\,d\mathbf{v}. \quad (4.21)$$

Again from isotropy,

$$\tilde{\mathfrak{J}}^{(0)}_{\mu\nu}(k,t) = n\delta_{\mu\nu}\int dvv_\mu^2\varphi(v,t). \quad (4.22)$$

To obtain the spectral function, we evaluate (4.20) and (4.22) at $t = 0$. Introducing

$$\varphi_0(v) = \varphi(v, t=0),\quad \bar{q}_0(v) = \bar{q}(v, t=0),\quad \epsilon_0(v) = \epsilon(v, t=0),\quad \Theta_0(v) = \Theta(v, t=0), \quad (4.23)$$

and

$$\Delta_0 = \int v_\mu^2\varphi_0(v)\,d\mathbf{v},$$

we get

$$\tilde{\mathfrak{J}}^{(1)}_{\mu\nu}(k,0) = n^2\frac{k_\mu k_\nu}{k^2}\times\int dv\Big[\frac{v}{n}\Theta_0(v) + v^2\frac{\bar{q}_0(v)}{|\epsilon_0(v)|^2} - v^2\bar{q}_0(v)\Big], \quad (4.22')$$

where

$$\tilde{\mathfrak{J}}^{(0)}_{\mu\nu}(k,0) = n\delta_{\mu\nu}\Delta_0,$$

$$\tilde{\mathfrak{J}}(k,0) = \tilde{\mathfrak{J}}^{(1)}(k,0) + \tilde{\mathfrak{J}}^{(0)}(k,0). \quad (4.24)$$

Now, one may easily recognize $\frac{1}{2}\Delta_0$ as the average energy per unit mass per degree of freedom for a general isotropic system, and therefore Δ_0 is constant throughout the time of evolution of the plasma towards equilibrium. Hence,

$$\mathfrak{J}^{(0)}_{\mu\nu}(k,0) = n\delta_{\mu\nu}(\beta m)^{-1}\int dv\bar{\varphi}_0(v), \quad (4.25)$$

and we can, therefore, write for the spectral function

$$S_{\mu\nu}(k,\omega) = \Big[(\beta m)^{-1}\bar{\varphi}_0(\omega/k)\delta_{\mu\nu} - \frac{\omega^2 n}{k^2}\bar{q}_0(\omega/k)\frac{k_\mu k_\nu}{k^2} + \frac{\omega^2 n}{k^2}\frac{\bar{q}_0(\omega/k)}{|\epsilon_0(\omega)|^2}\frac{k_\mu k_\nu}{k^2} + \frac{\omega}{k}\Theta_0\Big(\frac{\omega}{k}\Big)\cdot\frac{k_\mu k_\nu}{k^2}\Big]\frac{2\pi n}{k}. \quad (4.26)$$

The current-density correlation for arbitrary time is then obtained from

$$\tilde{\mathfrak{J}}_{\mu\nu}(k,t) = \frac{1}{2\pi}\int d\omega\exp(-i\omega t)S_{\mu\nu}(k,\omega). \quad (4.27)$$

Again, the consideration of causality retains only the contribution from $S^+_{\mu\nu}(k,\omega)$, the plus-function part of $S_{\mu\nu}(k,\omega)$, which is obtained from

$$S^+_{\mu\nu}(k,\omega) = \int d\omega'\delta_+(\omega-\omega')S_{\mu\nu}(k,\omega'). \quad (4.28)$$

The special case of equilibrium state is obtained by assuming a Maxwellian distribution for φ_0. The simplifications that result in this case are the following. First of all, since $\bar{\varphi}_0(\nu)$ is proportional to $\varphi_0(\nu)$, $\Theta(\nu)$ vanishes. Secondly, $\bar{q}_0$ simplifies into (3.19), and hence the contribution from the term $\nu^2\bar{q}_0(\nu)$ in (4.22′) again has the significance of average energy per unit mass per degree of freedom. The resulting spectral function is

$$S^0_{\mu\nu}(k,\omega) = 2\pi n\left[(m\beta)^{-1}\frac{\bar{\varphi}_0(\omega/k)}{k}\left(\delta_{\mu\nu}-\frac{k_\mu k_\nu}{k^2}\right)+\frac{\omega^2}{k}\frac{\bar{\varphi}_0(\omega/k)}{|\epsilon_0(k,\omega)|^2}\frac{k_\mu k_\nu}{k^4}\right], \quad (4.29)$$

which agrees with Rostoker's result.[3]

5. CONCLUSION

Expressions for spectral functions have been derived for isotropic nonequilibrium system in terms of the initial condition, which determines the evolution through the Balescu–Lenard kinetic equation. Extensive use of Van Kampen–Case eigenfunctions has been made in our derivation of the correlation function. We wish to emphasize that these eigenfunctions indeed provide a very convenient mathematical basis to deal with the typical problems of the statistical mechanics of charged particles. Similar set of generalized eigenfunctions has been obtained[8] for plasma in an external magnetic field, and use has been made of these to study the correlations in presence of magnetic field, which will be reported in a sequel to this paper.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to Professor T. Pradhan for his kind interest in this work. One of us (BD) wants to thank the Department of Atomic Energy, Government of India, for financial assistance during a part of this work.

[8] P. Dasgupta and B. Dasgupta, "Vlassov Equation in a Magnetic Field" (to be published).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Exponential Hilbert Space: Fock Space Revisited

JOHN R. KLAUDER
Bell Telephone Laboratories, Murray Hill, New Jersey 07974

(Received 8 May 1969)

An exponential Hilbert space, which is an abstraction of the familiar Fock space for bosons, provides a natural framework to discuss a wide class of field-operator representations. This framework is especially convenient when wide invariance groups, such as a unique translationally invariant state, are involved. In this paper, we develop the theory of exponential Hilbert spaces in a functional fashion suitable to discuss representations of field operators enjoying such invariance features. Representations of both current algebras and canonical field operators are discussed, and it is shown that these representations are natural generalizations of those characterizing infinitely divisible random processes. Questions of reducibility and equivalence are treated, and we prove that our construction gives rise to infinitely many unitarily inequivalent representations. Nevertheless, an extremely simple expression, bilinear in annihilation and creation operators, abstractly characterizes the operators of both the current algebras and canonical fields. Dynamical applications to quantum field theory will be treated in subsequent papers.

1. INTRODUCTION

In numerous applications, and especially for dynamical considerations in quantum field theory, the representation of basic field operators becomes important. In a canonically formulated theory, for example, a commonly used representation is the familiar Fock representation, although it is known to be inapplicable for interacting theories possessing wide invariance groups.[1] Compatibility with the invariance requirements is the minimum demand we can impose on a representation for it to be relevant; and for our purposes, we demand that there exists a unique, translationally invariant state in the representation space. The general representation of the appropriate algebra, such as the commutation relations or a current algebra, lacks a translationally invariant state or at least a unique one. Thus, it is of some interest that numerous field-operator representations, consistent with the invariance conditions, can be constructed in the framework of exponential Hilbert spaces.[2] Such representations—which may

[1] R. Streater and A. S. Wightman, *PCT, Spin and Statistics, and All That* (W. A. Benjamin, New York, 1964); A. S. Wightman, *Lecture Notes at the French Summer School of Theoretical Physics, Cargese Corsica, July, 1964* (Gordon & Breach, New York, 1967); L. Streit, Bull. Am. Phys. Soc. **14**, 86 (1969).

[2] Rudimentary ideas regarding exponential Hilbert spaces appear in K. O. Friedrichs, *Mathematical Aspects of the Quantum Theory of Fields* (Interscience, New York, 1953), and in J. R. Klauder, Ann. Phys. (N.Y.) **11**, 123 (1960), pp. 133 and 134; for a more explicit construction, see H. Araki and E. J. Woods, Publ. Res. Inst. Math. Sci. (Kyoto), Ser.A, **2**, 157 (1966).

The special case of equilibrium state is obtained by assuming a Maxwellian distribution for φ_0. The simplifications that result in this case are the following. First of all, since $\bar{\varphi}_0(\nu)$ is proportional to $\varphi_0(\nu)$, $\Theta(\nu)$ vanishes. Secondly, $\bar{q}_0$ simplifies into (3.19), and hence the contribution from the term $\nu^2\bar{q}_0(\nu)$ in (4.22′) again has the significance of average energy per unit mass per degree of freedom. The resulting spectral function is

$$S^0_{\mu\nu}(k,\omega) = 2\pi n\left[(m\beta)^{-1}\frac{\bar{\varphi}_0(\omega/k)}{k}\left(\delta_{\mu\nu}-\frac{k_\mu k_\nu}{k^2}\right)+\frac{\omega^2}{k}\frac{\bar{\varphi}_0(\omega/k)}{|\epsilon_0(k,\omega)|^2}\frac{k_\mu k_\nu}{k^4}\right], \quad (4.29)$$

which agrees with Rostoker's result.[3]

5. CONCLUSION

Expressions for spectral functions have been derived for isotropic nonequilibrium system in terms of the initial condition, which determines the evolution through the Balescu–Lenard kinetic equation. Extensive use of Van Kampen–Case eigenfunctions has been made in our derivation of the correlation function. We wish to emphasize that these eigenfunctions indeed provide a very convenient mathematical basis to deal with the typical problems of the statistical mechanics of charged particles. Similar set of generalized eigenfunctions has been obtained[8] for plasma in an external magnetic field, and use has been made of these to study the correlations in presence of magnetic field, which will be reported in a sequel to this paper.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to Professor T. Pradhan for his kind interest in this work. One of us (BD) wants to thank the Department of Atomic Energy, Government of India, for financial assistance during a part of this work.

[8] P. Dasgupta and B. Dasgupta, "Vlassov Equation in a Magnetic Field" (to be published).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Exponential Hilbert Space: Fock Space Revisited

JOHN R. KLAUDER
Bell Telephone Laboratories, Murray Hill, New Jersey 07974

(Received 8 May 1969)

An exponential Hilbert space, which is an abstraction of the familiar Fock space for bosons, provides a natural framework to discuss a wide class of field-operator representations. This framework is especially convenient when wide invariance groups, such as a unique translationally invariant state, are involved. In this paper, we develop the theory of exponential Hilbert spaces in a functional fashion suitable to discuss representations of field operators enjoying such invariance features. Representations of both current algebras and canonical field operators are discussed, and it is shown that these representations are natural generalizations of those characterizing infinitely divisible random processes. Questions of reducibility and equivalence are treated, and we prove that our construction gives rise to infinitely many unitarily inequivalent representations. Nevertheless, an extremely simple expression, bilinear in annihilation and creation operators, abstractly characterizes the operators of both the current algebras and canonical fields. Dynamical applications to quantum field theory will be treated in subsequent papers.

1. INTRODUCTION

In numerous applications, and especially for dynamical considerations in quantum field theory, the representation of basic field operators becomes important. In a canonically formulated theory, for example, a commonly used representation is the familiar Fock representation, although it is known to be inapplicable for interacting theories possessing wide invariance groups.[1] Compatibility with the invariance requirements is the minimum demand we can impose on a representation for it to be relevant; and for our purposes, we demand that there exists a unique, translationally invariant state in the representation space. The general representation of the appropriate algebra, such as the commutation relations or a current algebra, lacks a translationally invariant state or at least a unique one. Thus, it is of some interest that numerous field-operator representations, consistent with the invariance conditions, can be constructed in the framework of exponential Hilbert spaces.[2] Such representations—which may

[1] R. Streater and A. S. Wightman, *PCT, Spin and Statistics, and All That* (W. A. Benjamin, New York, 1964); A. S. Wightman, *Lecture Notes at the French Summer School of Theoretical Physics, Cargese Corsica, July, 1964* (Gordon & Breach, New York, 1967); L. Streit, Bull. Am. Phys. Soc. **14**, 86 (1969).

[2] Rudimentary ideas regarding exponential Hilbert spaces appear in K. O. Friedrichs, *Mathematical Aspects of the Quantum Theory of Fields* (Interscience, New York, 1953), and in J. R. Klauder, Ann. Phys. (N.Y.) **11**, 123 (1960), pp. 133 and 134; for a more explicit construction, see H. Araki and E. J. Woods, Publ. Res. Inst. Math. Sci. (Kyoto), Ser.A, **2**, 157 (1966).

be called exponential representations—are the subject of this paper, along with the intimately related theory of exponential Hilbert spaces that "support" them. Sections 2 and 3 are devoted to our formulation of the exponential Hilbert space by appealing to familiar and intuitive properties of the usual Fock representation.[3] In Sec. 4 we develop the exponential representations of current algebras, while in Sec. 5 we discuss exponential solutions of the canonical commutation relations. In each case, questions of reducibility and equivalence of various representations are treated.

The analysis of these representations within the context of exponential Hilbert spaces has several advantages. In the first place, questions of invariance and uniqueness of a state in the representation space are easily dealt with. Secondly, it provides a natural class of representations with which to deal, a class which is a natural generalization of those characterizing the so-called infinitely divisible random variables.[4] It lends itself to a natural and unified approach to the representation of diverse algebras like current algebras and canonical operators. Moreover, it is difficult to envisage a representation consistent with having a unique translationally invariant state that does not fit into our framework, although, of course, such a wild conjecture is undoubtedly false.

In spite of the fact that there are infinitely many inequivalent representations of operator algebras—and, indeed, infinitely many inequivalent exponential representations—there is, nevertheless, a comparatively simple abstract operator solution that covers all our cases. To present this operator solution, suppose we consider as an example the canonical field operators $\varphi(\mathbf{x})$ and $\pi(\mathbf{y})$ for which

$$[\varphi(\mathbf{x}), \pi(\mathbf{y})] = i\delta(\mathbf{x} - \mathbf{y}). \tag{1.1}$$

The general form of exponential representation we are led to for these operators may be given as follows: Initially, let $\varphi_1(\mathbf{x})$, $\pi_1(\mathbf{y})$ denote a Fock representation (or a two-fold direct sum of such representations) which is obtained from a suitable linear combination of annihilation and creation operators, $A_1(\mathbf{x})$ and $A_1^\dagger(\mathbf{x})$. In addition, let us introduce an auxiliary, independent set of annihilation and creation operators, A_{2r} and $A_{2r}^\dagger$, $r = 1, 2, \cdots$. Then the class of solutions we are led to has the basic form given by

$$\varphi(\mathbf{x}) = \varphi_1(\mathbf{x}) + A_2^\dagger\tilde{\varphi}(\mathbf{x})A_2, \tag{1.2a}$$

$$\pi(\mathbf{x}) = \pi_1(\mathbf{x}) + A_2^\dagger\tilde{\pi}(\mathbf{x})A_2. \tag{1.2b}$$

Here $\tilde{\varphi}(\mathbf{x}) = \{\tilde{\varphi}_{rs}(\mathbf{x})\}$ and $\tilde{\pi}(\mathbf{x}) = \{\tilde{\pi}_{rs}(\mathbf{x})\}$ are general, commuting, formally self-adjoint operators defined on the index space of A_2, and

$$A_2^\dagger\tilde{\varphi}(\mathbf{x})A_2 \equiv \sum_{r,s} A_{2r}^\dagger\tilde{\varphi}_{rs}(\mathbf{x})A_{2s}, \quad \text{etc.} \tag{1.3}$$

Provided that the operators A_2 differ from a Fock representation by a suitable additive multiple of the identity operator, the resultant operators φ and π will have all the desired features. Suitable solutions to a current algebra have essentially the same structure in which the first (linear) terms are generally absent, and the analogs of the operators $\tilde{\varphi}(\mathbf{x})$ and $\tilde{\pi}(\mathbf{x})$ also fulfill the current algebra. To achieve the desired invariance, it is important that the representation of the operators A_2 is unitarily inequivalent to a Fock representation, but differs from one only by an additive multiple of unity. Such operators give rise to "translated Fock representations," which were among the earliest non-Fock representations to be studied.[5] Utilizing this fact, we may summarize by noting that all the basic field operators we consider admit a bilinear expansion in terms of annihilation and creation operators of an embedding, or "parental" Fock representation.

This comparatively simple and completely general form of solution makes it practical to study such representations for possible application to quantum dynamical models. Elsewhere, we shall study our solutions from this point of view and show that nontrivial model field theories having wide invariance groups can be constructed with their help.[6] Here we content ourselves with the formulation of exponential Hilbert spaces and the exponential representations they so naturally contain.

Earlier studies along these lines have been directed at continuous tensor product representations,[7] which are extensions of the notion of direct product representations. The form of inner product adopted therein did not always generate a Hilbert space with positive-definite metric and gave rise to only a limited class of solutions of the basic field algebras. Our formulation and results are more general in character, always yielding proper Hilbert spaces and yielding far larger classes of solutions of the basic equations. In turn, we feel, this has been achieved with considerable

[3] For an introductory treatment, see J. R. Klauder, *Coherence, Correlation and Quantum Field Theory, Brandeis Summer School, 1967* (Gordon & Breach, New York, to be published), Sec. 4.3.

[4] E. Lukacs, *Characteristic Functions* (Charles Griffin, London, 1960), p. 79; I. M. Gel'fand and N. Y. Vilenkin, *Generalized Functions, Vol. 4: Applications of Harmonic Analysis*, translated by A. Feinstein (Academic Press, Inc., New York, 1964), Chap. III.

[5] L. van Hove, Physica **18**, 145 (1952); K. O. Friedrichs, Ref. 2.

[6] J. R. Klauder, See also Ref. 3, Sec. 6.

[7] R. F. Streater, Nuovo Cimento **53A**, 487 (1968); D. A. Dubin and R. F. Streater, Nuovo Cimento **50**, 154 (1967); R. F. Streater, *Lectures at 1968 Karpacz Winter School, Karpacz, Poland* (to be published); R. F. Streater and A. Wulfsohn, Nuovo Cimento **57B**, 330 (1968); R. F. Streater, *Lectures at Varenna Summer School, 1968* (to be published). See also H. Araki and E. J. Woods, Publ. Res. Inst. Math. Sci. (Kyoto), Ser. A, **2**, 157 (1966), and A. Guichardet, Commun. Math. Phys. **5**, 262 (1967), for the introduction of continuous tensor products.

simplification in the mathematical prerequisites, since, in essence, we merely exploit Fock-space methods to the hilt. This does not mean that our notation remains uncomplicated; simpler notational choices could be made, but, in the interests of maintaining some over-all unity in our presentation, we felt it necessary to resist that temptation.

2. EXPONENTIAL HILBERT SPACE: MOTIVATION AND ABSTRACTION

A. Fock Representation and Coherent States

We begin with a brief review of well-known properties of the Fock representation of a countable collection of annihilation and creation operators, A_l, $A_m^\dagger$, respectively, which fulfill the commutation relations

$$[A_l, A_m^\dagger] = \delta_{lm}. \tag{2.1}$$

The Fock representation is singled out by the requirement that the collection of operators A_l and $A_m^\dagger$ is irreducible and that there is a normed state $|0\rangle$, the no-particle state, for which

$$A_l |0\rangle = 0 \tag{2.2}$$

for all l. States in which n excitations are present are defined by

$$|l_1, l_2, \cdots, l_n\rangle \equiv A_{l_1}^\dagger A_{l_2}^\dagger \cdots A_{l_n}^\dagger |0\rangle, \tag{2.3}$$

and the collection of such states for all l_j and all n span the Hilbert space $\mathfrak{H}$. If $\{z_l\}$ is a square-summable sequence of complex numbers, then the state

$$|\{z_l\}\rangle \equiv \exp(-\tfrac{1}{2}\textstyle\sum |z_l|^2) \exp(\sum z_l A_l^\dagger) |0\rangle \tag{2.4}$$

defines a normalized vector in $\mathfrak{H}$ depending on $\{z_l\}$. These vectors constitute a specific overcomplete family of states (OFS), the so-called coherent states, and their properties are well known.[8] In particular, we note that

$$A_m |\{z_l\}\rangle = z_m |\{z_l\}\rangle, \tag{2.5}$$

i.e., that these states are eigenstates of the annihilation operators with eigenvalues given by the parameters z_l. The overlap of two such states is given by the formula

$$\begin{aligned}\langle\{z_k'\} \mid \{z_k\}\rangle &= \exp(-\tfrac{1}{2}\textstyle\sum |z_k'|^2 - \tfrac{1}{2}\sum |z_k|^2 + \sum z_k^{*\prime} z_k) \\ &= \exp(-\tfrac{1}{2}\textstyle\sum |z_k' - z_k|^2 + i \operatorname{Im} \sum z_k^{*\prime} z_k),\end{aligned} \tag{2.6}$$

which never vanishes. Although the vectors $|\{z_k\}\rangle$ are never orthogonal pairwise, they span the Hilbert space $\mathfrak{H}$, or, as we shall say, they constitute a total set $\mathfrak{T}$. A total set $\mathfrak{T}$ is characterized by the equivalent properties that every vector may be given as a linear combination (possibly not unique) of the vectors in $\mathfrak{T}$, or that a vector is the zero vector if its inner product vanishes with every element of $\mathfrak{T}$. In fact, various subsets of the coherent states, which we may call characteristic sets, also yield total sets. Analyticity arguments[9] indicate, for example, that as far as, say, the parameter z_1 goes, either a line segment or a set with a finite accumulation point are sufficient to generate a total set. Analyticity combined with growth restrictions[10] lead to the sufficiency of a lattice in $\operatorname{Re} z_1$ and $\operatorname{Im} z_1$, with density greater than one point per unit circle (i.e., an area of π). Square-integrability leads to the sufficiency of such a lattice with density equal to one point per unit circle,[11] which just turns out to be one point in phase space per Planck cell (i.e., an area $h = 2\pi\hbar$). Lattices with a density less than this value do not yield a total set for $\mathfrak{H}$.[12] From these examples it should be apparent that there is a wide variety of characteristic subsets of coherent states each of which yields a total set for $\mathfrak{H}$.

Coherent-state matrix elements of normally ordered operators are particularly simple. If $:B(A_l^\dagger, A_m):$ denotes a normally ordered operator—all creation operators to the left of all destruction operators—then it follows from (2.5) that

$$\langle\{z_k'\}| :B(A_l^\dagger, A_m): |\{z_k\}\rangle = B(z_l^{*\prime}, z_m)\langle\{z_k'\} \mid \{z_k\}\rangle. \tag{2.7}$$

The diagonal matrix elements become

$$\langle\{z_k\}| :B(A_l^\dagger, A_m): |\{z_k\}\rangle = B(z_l^*, z_m). \tag{2.8}$$

From this expression it is clear that the diagonal matrix elements actually determine the operator, since we may regard z_l^* and z_m as independent variables in the argument of B. Subsets of diagonal coherent state matrix elements may also suffice to uniquely determine the operator. As far as the variable z_1 goes, for example, any open interval in the complex plane suffices, or a set of points of the form $z_1 \equiv x_{1n} + iy_{1m}$, where $\{x_{1n}\}$ and $\{y_{1m}\}$ are two sets with a finite accumulation point.[13] On the other hand, the dependence of just $x_1 = \operatorname{Re} z_1$ does not suffice to determine B; information from both the real and imaginary parts must be provided.

[8] J. R. Klauder and E. C. G. Sudarshan, *Fundamentals of Quantum Optics* (W. A. Benjamin, New York, 1968), Chap. 7.

[9] V. Bargmann, Commun. Pure Appl. Math. **14**, 187 (1961).

[10] P. Butera and L. Girardello, "On the Completeness of the Coherent States" (University of Milan, preprint No. IFUM 084/FT); V. Bargmann (private communication).

[11] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, N.J., 1955), p. 406; J. R. Klauder (unpublished).

[12] V. Bargmann (private communication).

[13] K. E. Cahill, Phys. Rev. **138**, B1566 (1965).

We conclude this elementary review with a discussion of Hermitian operators of the form

$$W = \sum_{l,m} A_l^\dagger w_{lm} A_m . \tag{2.9a}$$

For simplicity, we confine our initial remarks to the specialized operators for which

$$W = \sum_l w_l A_l^\dagger A_l . \tag{2.9b}$$

Since these involve a weighted sum of number operators $N_l = A_l^\dagger A_l$, we can directly determine that

$$\begin{aligned} e^{-iWt}|\{z_k\}\rangle &= \exp(-\tfrac{1}{2}\textstyle\sum |z_k|^2) \exp(-iWt) \exp(\sum z_k A_k^\dagger)|0\rangle \\ &= \exp(-\tfrac{1}{2}\textstyle\sum |z_k|^2) \exp(\sum z_k e^{-iWt} A_k^\dagger e^{iWt})|0\rangle \\ &= \exp(-\tfrac{1}{2}\textstyle\sum |z_k|^2) \exp(\sum z_k e^{-iw_k t} A_k^\dagger)|0\rangle \\ &\equiv |\{e^{-iw_k t} z_k\}\rangle . \end{aligned} \tag{2.10}$$

Consequently, we learn that

$$\begin{aligned} \langle\{z_k'\}|\, e^{-iWt}|\{z_k\}\rangle &= \exp(-\tfrac{1}{2}\textstyle\sum |z_k'|^2 - \tfrac{1}{2}\sum |z_k|^2 + \sum z_k'^* e^{-iw_k t} z_k). \end{aligned} \tag{2.11}$$

Evidently, if we had considered the more general operator (2.9a), then the last sum in the exponent of (2.11) would be replaced by

$$\sum z_k'^* (e^{-iwt})_{kl} z_l . \tag{2.12}$$

If an operator of the form (2.9) represented the Hamiltonian $\mathcal{H}$, then the determination of the dynamics is simple and straightforward. More interesting, however, is the dynamics based on Hamiltonians of the form

$$\begin{aligned} \mathcal{H} = &\sum A_k^\dagger \omega_{kl} A_l \\ &+ \sum g_{klm} A_k^\dagger A_l^\dagger A_m + \text{h.c.} \\ &+ \sum v_{klmp} A_k^\dagger A_l^\dagger A_m A_p , \end{aligned} \tag{2.13}$$

including terms representing production and decay, scattering, etc. As we shall subsequently see, the exploitation of Fock space for "embedding" non-Fock representations of the basic field operators will permit the study of Hamiltonians of the complexity of (2.13) which exhibit nontrivial invariance groups, such as the Euclidean group of rotations and translations.

B. Abstraction to General Hilbert Space

The next stage in our development amounts to a relatively simple abstraction of the foregoing presentation. We note that square-summable, complex-valued sequences $\{z_l\}$ themselves constitute a Hilbert space, the so-called l^2. The vectors $|\{z_l\}\rangle \in \mathfrak{H}$ are then images of vectors in l^2 as elements in $\mathfrak{H}$. We may make these notions more general in character in the following way: Rather than restrict ourselves to l^2, let us consider a general Hilbert space $\mathfrak{h}$ with elements ψ, etc., denoted by lower case Greek letters. The inner product in this space will be denoted by (ψ', ψ) and takes the place of the expression $\sum z_k'^* z_k$ used earlier. The image vectors, previously denoted by $|\{z_k\}\rangle$, may now be called Ψ, the capital Greek letter associated with the corresponding element in $\mathfrak{h}$. The appropriate inner product in $\mathfrak{H}$ will be denoted by (Ψ', Ψ). This is, of course, just one of many possible notational conventions characterizing our abstraction. The essential point is that there are two Hilbert spaces involved, $\mathfrak{h}$ and $\mathfrak{H}$, and that there is a correspondence, i.e., a map, between the elements $\psi \in \mathfrak{h}$ and a *subset*, a *total set* $\mathfrak{T}$ of unit vectors $\Psi \in \mathfrak{H}$. To make the correspondence precise, we may restate the overlap (2.6) in the form

$$\begin{aligned} (\Psi', \Psi) &= e^{-\frac{1}{2}\|\psi'\|^2 - \frac{1}{2}\|\psi\|^2 + (\psi', \psi)} \\ &= N' N e^{(\psi', \psi)} . \end{aligned} \tag{2.14}$$

Here

$$N \equiv e^{-\frac{1}{2}\|\psi\|^2}, \tag{2.15a}$$

$$N' \equiv e^{-\frac{1}{2}\|\psi'\|^2} \tag{2.15b}$$

are normalization factors to ensure that

$$(\Psi, \Psi) = 1 \tag{2.16}$$

for all $\Psi \in \mathfrak{T}$.

Since $\mathfrak{T}$ is a total set, we may represent an arbitrary vector $\mathrm{X} \in \mathfrak{H}$ as a linear sum of elements in $\mathfrak{T}$, such as

$$\mathrm{X} = \sum_{n=1}^{\infty} c_n \Psi_n , \quad \Psi_n \in \mathfrak{T}. \tag{2.17}$$

The inner product of two such elements follows from (2.14), namely,

$$(\mathrm{X}', \mathrm{X}) = \sum c_m'^* c_n (\Psi_m', \Psi_n), \tag{2.18}$$

or the matrix elements of an arbitrary operator $\mathcal{O}$ as

$$(\mathrm{X}', \mathcal{O}\mathrm{X}) = \sum c_m'^* c_n (\Psi_m', \mathcal{O}\Psi_n). \tag{2.19}$$

For simplicity, we shall generally confine our attention to matrix elements between the special vectors in the total set $\mathfrak{T}$.

We may express the correspondence between ψ and Ψ in other ways as well. Let

$$\mathfrak{H}_{(n)} \equiv \bigotimes_{s=1}^{n} \mathfrak{h}_s|_S \tag{2.20}$$

denote the symmetrized (hence the subscript S), n-fold tensor product of $\mathfrak{h}$ with itself. For $n = 0$, $\mathfrak{H}_{(0)} \equiv C$,

the set of complex numbers. Then the connection between $\mathfrak{H}$ and $\mathfrak{h}$ may be stated essentially as

$$\mathfrak{H} = \bigoplus_{n=0}^{\infty} \mathfrak{H}_{(n)} = \bigoplus_{n=0}^{\infty} \left(\bigotimes_{s=1}^{n} \mathfrak{h}_s \right) \Big|_S . \tag{2.21}$$

In particular, the vectors $\Psi \in \mathfrak{T}$ and the vectors $\psi \in \mathfrak{h}$ are related as follows: For $\psi \in \mathfrak{h}$, we first set

$$\Psi'_{(n)} = \bigotimes_{s=1}^{n} \psi_s|_S \in \bigotimes_{s=1}^{n} \mathfrak{h}_s|_S = \mathfrak{H}_{(n)}, \tag{2.22}$$

where by definition, if $n = 0$, we choose $\Psi'_{(0)} = 1$. Then the appropriate Ψ is given by

$$\begin{aligned} \Psi &= N \bigoplus_{n=0}^{\infty} \frac{1}{\sqrt{n!}} \Psi'_{(n)} \\ &= N \bigoplus_{n=0}^{\infty} \frac{1}{\sqrt{n!}} \left(\bigotimes_{s=1}^{n} \psi_s \right) \Big|_S . \end{aligned} \tag{2.23}$$

Here, as before, N is a normalization factor given by $\exp(-\frac{1}{2}\|\psi\|^2)$. It is evident that the inner product of two such vectors has the form given in (2.14). While all the vectors $\Psi \in \mathfrak{T}$ imaged by all vectors $\psi \in \mathfrak{h}$ constitute a total set, there are, according to our earlier discussion, "characteristic" subsets of $\mathfrak{h}$ for which the image vectors would still constitute a total set in $\mathfrak{H}$. We suggestively set $\mathfrak{H} = \exp \mathfrak{h}$ and refer to $\mathfrak{H}$ as an exponential Hilbert space.

Along with our abstraction of the Hilbert spaces involved we shall generalize the operators. We take Eq. (2.5) as an essential defining property of the annihilation operators and abstract that relation as follows: To every $\lambda \in \mathfrak{h}$ we associate an operator $A(\lambda)$, such that

$$A(\lambda)\Psi = (\lambda, \psi)\Psi \tag{2.24}$$

for every $\Psi \in \mathfrak{T}$ imaged by $\psi \in \mathfrak{h}$. We note that $A(\lambda)$ is *antilinear* in λ, and, permitting ourselves a notational abuse, we could suggestively write (λ, A) for $A(\lambda)$. It is clear that, as straightforward abstractions of earlier relations,

$$(\Psi', A(\lambda)\Psi) = (\lambda, \psi)(\Psi', \Psi), \tag{2.25a}$$

$$(\Psi', A(\lambda)^\dagger \Psi) = (\psi', \lambda)(\Psi', \Psi), \tag{2.25b}$$

$$\begin{aligned} (\Psi', A(\lambda_1)^\dagger \cdots A(\lambda_n)^\dagger A(\mu_1) \cdots A(\mu_m)\Psi) \\ = (\psi', \lambda_1) \cdots (\psi', \lambda_n)(\mu_1, \psi) \cdots (\mu_m, \psi) \\ \times (\Psi', \Psi), \end{aligned} \tag{2.25c}$$

as well as

$$[A(\lambda), A(\lambda')^\dagger] = (\lambda, \lambda') \tag{2.26}$$

for any two elements $\lambda, \lambda' \in \mathfrak{h}$.

To every bounded operator b on $\mathfrak{h}$ we can associate a particular operator B on $\mathfrak{H}$ defined as

$$B = \bigoplus_{n=0}^{\infty} B_{(n)} = \bigoplus_{n=0}^{\infty} \bigotimes_{s=1}^{n} b, \tag{2.27}$$

where $B_{(0)} \equiv 1$, which maps a vector $\Psi \in \mathfrak{T}$ into another such vector apart from normalization. Specifically,

$$(\Psi', B\Psi) = N'N e^{(\psi', b\psi)}, \tag{2.28}$$

from which we see that $B\Psi/\|B\Psi\| \in \mathfrak{T}$. If $b \equiv e^{-iwt}$, where w is a self-adjoint operator on $\mathfrak{h}$, then b is unitary. As the direct sum of direct products of unitary operators,

$$B = \bigoplus_{n=0}^{\infty} \bigotimes_{s=1}^{n} e^{-iwt} \equiv e^{-iWt} \tag{2.29}$$

defines a unitary operator on $\mathfrak{H}$ mapping $\mathfrak{T}$ onto $\mathfrak{T}$. Specifically,

$$(\Psi', e^{-iWt}\Psi) = N'N \exp(\psi', e^{-iwt}\psi), \tag{2.30}$$

which associates a self-adjoint operator W on $\mathfrak{H}$ for every self-adjoint operator w on $\mathfrak{h}$. From the relation

$$(\Psi', W\Psi) = (\psi', w\psi)(\Psi', \Psi), \tag{2.31}$$

which follows from (2.30), it is clear on comparison with (2.25c) that W is bilinear in $A^\dagger$ and A. Specifically, if $\{\lambda_n\}$ constitutes a complete orthonormal set in $\mathfrak{h}$, then

$$W = \sum_{n,m} A^\dagger(\lambda_n)(\lambda_n, w\lambda_m)A(\lambda_m), \tag{2.32}$$

which we may suggestively abbreviate by

$$W = (A, wA) \tag{2.33a}$$

or just by

$$W = A^\dagger w A, \tag{2.33b}$$

as we did in Eq. (1.3). We note that the diagonal matrix elements for such operators are given simply by

$$(\Psi, W\Psi) = (\psi, w\psi), \tag{2.34}$$

since $\|\Psi\| = 1$. As noted in Sec. 2A, the diagonal elements $W(\Psi) \equiv (\Psi, W\Psi)$ for all $\Psi \in \mathfrak{T}$ uniquely determine W. In the present special example this is evident from the fact that we know $w(\psi) \equiv (\psi, w\psi)$ for all $\psi \in \mathfrak{h}$, and by the polarization identity[14] we therefore know $(\psi', w\psi)$ for all $\psi', \psi \in \mathfrak{h}$ (strictly speaking in the domain of w). According to (2.31), this is sufficient to fix $(\Psi', W\Psi)$.

Let us briefly examine the diagonal elements of (2.30) further. Since $N = \exp(-\frac{1}{2}\|\psi\|^2)$, it follows that

$$(\Psi, e^{-iWt}\Psi) = e^{(\psi, (e^{-iwt}-1)\psi)}. \tag{2.35}$$

The left-hand side is the characteristic function for the distribution of W in the state Ψ and acts both as a generator of the moments and of the linked moments (subscript L, sometimes called the cumulants, which

[14] F. Riesz and B. Sz.-Nagy, *Functional Analysis* (Fredrick Ungar, New York, 1955), p. 211.

are related to truncated functions and connected diagrams in field theory). In particular,

$$(\Psi, e^{-iWt}\Psi) \equiv \langle e^{-iWt}\rangle = \sum_{m=0}^{\infty} \frac{(-it)^m}{m!}\langle W^m\rangle$$
$$\equiv \exp\{\langle e^{-iWt} - 1\rangle_L\}$$
$$= \exp\left\{\sum_{m=1}^{\infty} \frac{(-it)^m}{m!}\langle W^m\rangle_L\right\}. \quad (2.36)$$

On identification with (2.35), we see that

$$(\psi, (e^{-iwt} - 1)\psi) \equiv \langle e^{-iwt} - 1\rangle = \sum_{m=1}^{\infty} \frac{(-it)^m}{m!}\langle w^m\rangle$$
$$= \langle e^{-iWt} - 1\rangle_L = \sum_{m=1}^{\infty} \frac{(-it)^m}{m!}\langle W^m\rangle_L, \quad (2.37)$$

namely, that the mth *linked* moment $\langle W^m\rangle_L$ equals the mth *ordinary* moment $\langle w^m\rangle$ as calculated in $\mathfrak{h}$.

To make these notions precise, we need to make some remarks regarding operator domains. We recall that in order for $\psi \in \mathfrak{d}_{w^p}$, i.e., for $w^p\psi \in \mathfrak{h}$, it is necessary and sufficient that

$$\frac{d^{2p}}{dt^{2p}}(\psi, e^{-iwt}\psi) \quad (2.38)$$

exist at $t = 0$.[15] A parallel criterion establishes the conditions for $W^p\Psi \in \mathfrak{H}$. In view of the direct connection between the characteristic functions afforded by (2.35), we see that $\psi \in \mathfrak{d}_{w^p}$ implies that $\Psi \in \mathfrak{D}_{W^p}$, and conversely.

From the prescribed form of the inner product (2.14), it is not difficult to establish the following inequalities:

$$2[1 - \exp(-\tfrac{1}{2}\|\psi_1 - \psi_2\|^2)] \leq \|\Psi_1 - \Psi_2\|^2$$
$$\leq 2(\|\psi_1\| + \|\psi_2\|)\|\psi_1 - \psi_2\|, \quad (2.39)$$

for any two elements $\psi_1, \psi_2 \in \mathfrak{h}$ and their exponential images.[16] It follows that a stongly convergent sequence of vectors in $\mathfrak{h}$ is imaged into a strongly convergent sequence of vectors in $\mathfrak{T}$, and conversely. Another convergence condition follows from the relation

$$(\Psi', N_1^{-1}\Psi_1 - N_2^{-1}\Psi_2) = N'[e^{(\psi', \psi_1)} - e^{(\psi', \psi_2)}]. \quad (2.40)$$

From this relation we conclude that a weakly convergent sequence of vectors in $\mathfrak{h}$ is imaged into a weakly convergent sequence of vectors in $\mathfrak{H}$ [of the special form $N^{-1}\Psi, \Psi \in \mathfrak{T}$, where $N^{-1} = \exp(\frac{1}{2}\|\psi\|^2)$, $\psi \in \mathfrak{h}$], and conversely.

[15] E. Lukacs, *Characteristic Functions* (Charles Griffin, London, 1960), p. 29, Theorem 2.3.1, Corollaries 1 and 2.

[16] The first inequality arises from the fact that $\cos\theta \leq 1$; the second inequality is derived in Ref. 8, p. 113.

3. EXPONENTIAL HILBERT SPACE: LABEL SPACES AND REDUCED PARAMETERIZATION

We develop the formalism of Sec. 2 one further stage by considering the consequences of a reduced parameterization. Rather than considering every $\psi \in \mathfrak{h}$ and its image $\Psi \in \mathfrak{T} \subset \mathfrak{H}$, let us consider the vectors, say, $\varphi \in \mathfrak{s} \subset \mathfrak{h}$ and their images $\Phi \in \mathfrak{S} \subset \mathfrak{T} \subset \mathfrak{H}$. It is especially convenient at this point to introduce a label space $\mathfrak{L}$ and to regard φ as a map from points $l \in \mathfrak{L}$ to vectors $\varphi[l] \in \mathfrak{s}$, a subset of $\mathfrak{h}$. That is, we imagine labeling the vectors in $\mathfrak{s}$ by points $l \in \mathfrak{L}$. Clearly, we can use the same label to denote the image vector $\Phi = \Phi[l] \in \mathfrak{S}$ in the exponential Hilbert space $\mathfrak{H} = \exp\mathfrak{h}$. In this notation we have

$$(\Phi[l_1], \Phi[l_2]) = N_1N_2e^{(\varphi[l_1], \varphi[l_2])}, \quad (3.1a)$$

where

$$N = e^{-\frac{1}{2}\|\varphi[l]\|^2}. \quad (3.1b)$$

We emphasize that this relation is nothing but an alternative parameterization for a subset of the vectors introduced in Sec. 2B. The question of whether or not

$$\mathfrak{S} \equiv \{\Phi[l]: l \in \mathfrak{L}\} \quad (3.2a)$$

is a total set for $\mathfrak{H}$ depends both on $\mathfrak{L}$ and on the particular labeling. Obviously, the question can be turned around to be one for

$$\mathfrak{s} \equiv \{\varphi[l]: l \in \mathfrak{L}\}. \quad (3.2b)$$

In a manner of speaking, in analogy with the discussion of the coherent states in Sec. 2A, it is sufficient if $\mathfrak{s}$ is a "characteristic set" in some sense. While we shall tacitly assume such to be the case, our main conclusions often do not hinge on this assumption, and, occasionally, it is useful to regard the total set spanned by $\mathfrak{S}$ to be a proper subset of $\mathfrak{H}$.

All of the operator equations of Sec. 2B have immediate applicability. For example, we have

$$A(\lambda)\Phi[l] = (\lambda, \varphi[l])\Phi[l], \quad (3.3)$$

$$(\Phi[l'], e^{-iWt}\Phi[l]) = N'Ne^{(\varphi[l'], e^{-iwt}\varphi[l])}, \quad \text{etc.} \quad (3.4)$$

Let us temporarily digress at this point to present an example which should help clarify some of the concepts we have introduced as well as point the way to the next stage of formal development.

Example: In the light of the formalism developed, suppose we wish to determine unitary representations of the one-parameter group

$$U(t) = e^{-iWt}, \quad (3.5)$$

which fulfills the combination law

$$U(t)U(t') = U(t + t'), \quad (3.6)$$

for all $t, t' \in R$, the real line. For the label space $\mathfrak{L}$, we choose the real line R, and, in fact, we choose the naturally induced parameterization given by

$$\Phi[t] \equiv U(t)\Phi_0, \tag{3.7}$$

where Φ_0 is some fixed fiducial vector and $U(t)$ is the representation under study. If we now insist that this representation be realized through an exponential construction, then we require that

$$(\Phi[t'], \Phi[t]) = \exp\{-\tfrac{1}{2}\|\varphi[t']\|^2 - \tfrac{1}{2}\|\varphi[t]\|^2 + (\varphi[t'], \varphi[t])\}. \tag{3.8a}$$

According to (3.6), this expression also equals

$$(\Phi_0, \Phi[t - t']) = \exp\{-\tfrac{1}{2}\|\varphi[0]\|^2 - \tfrac{1}{2}\|\varphi[t - t']\|^2 + (\varphi[0], \varphi[t - t'])\}. \tag{3.8b}$$

Equating real and imaginary parts in the exponent, we are led to the two conditions [cf. Eq. (2.6)]

$$\|\varphi[t] - \varphi[t']\|^2 = \|\varphi[0] - \varphi[t - t']\|^2, \tag{3.9a}$$

$$\operatorname{Im}(\varphi[t'], \varphi[t]) = \operatorname{Im}(\varphi[0], \varphi[t - t']). \tag{3.9b}$$

As a solution to these relations, we choose

$$\varphi[t] = a + bt \oplus \int^{\oplus} e^{-ixt}\, d\mu(x), \tag{3.10}$$

where a and b are fixed complex numbers and where the inner product is defined by

$$(\varphi[t'], \varphi[t]) = (a^* + b^*t')(a + bt) + \int e^{ix(t'-t)}\, d\mu(x). \tag{3.11}$$

Observe that

$$\|\varphi[t]\|^2 \geq \int d\mu(x) \equiv \mu(R), \tag{3.12}$$

which means [if $\mu(R) > 0$] that $\varphi[t]$ is never the zero vector. In this realization,

$$\mathfrak{h} = C \oplus \int^{\oplus} C\, d\mu(x), \tag{3.13}$$

for some measure μ, where C denotes the space of complex numbers. If we set $c \equiv \operatorname{Im} ba^*$, then the representations we have found have characteristic functions of the form

$$(\Phi_0, e^{-iWt}\Phi_0) = (\Phi_0, \Phi[t]) = \exp\left[ict - \tfrac{1}{2}|b|^2 t^2 - \int (1 - e^{-ixt})\, d\mu(x)\right], \tag{3.14}$$

which may be recognized as the characteristic function of an infinitely divisible random variable.[4] However, our derivation has not yielded the most general such characteristic function, since we necessarily have $\mu(R) < \infty$. In essence, the formal development we next take up is aimed at rectifying this deficiency and its analog in related examples.

A. Translated Parameterization

We now take up, in the abstract framework, the consequences of a translated parameterization and, subsequently, return to our example for illustration. Let us assume we are given a set $\mathfrak{s} \subset \mathfrak{h}$ and have constructed the exponential map from $\varphi[l] \in \mathfrak{s}$ to $\Phi[l] \in \mathfrak{S}$. We now introduce a family of phase-related image vectors according to the rule

$$\Phi'[l] \equiv e^{-i \operatorname{Im}(\xi, \varphi[l])}\Phi[l], \tag{3.15}$$

where ξ is some fixed element of $\mathfrak{h}$. Clearly,

$$(\Phi'[l_1], \Phi'[l_2]) = (\Phi[l_1], \Phi[l_2])e^{i \operatorname{Im}\{(\xi, \varphi[l_1]) - (\xi, \varphi[l_2])\}}, \tag{3.16}$$

which can be manipulated to become

$$(\Phi'[l_1], \Phi'[l_2]) = \exp\{-\tfrac{1}{2}\|\varphi[l_1] - \xi\|^2 - \tfrac{1}{2}\|\varphi[l_2] - \xi\|^2 + (\varphi[l_1] - \xi, \varphi[l_2] - \xi)\} \equiv N_1' N_2' e^{(\varphi'[l_1], \varphi'[l_2])}, \tag{3.17}$$

where

$$\varphi'[l] \equiv \varphi[l] - \xi. \tag{3.18}$$

From this equation we can read that a uniform translation of the set $\mathfrak{s}$, i.e., $\varphi[l] \to \varphi'[l] = \varphi[l] - \xi$, leads to an associated phase change in the image vectors in $\mathfrak{H}$. Note that

$$\operatorname{Im}(\xi, \varphi[l]) = \operatorname{Im}(\xi, \varphi[l] - \xi) = \operatorname{Im}(\xi, \varphi'[l]), \tag{3.19}$$

so that the phase change may be given by (3.15) or by

$$\Phi'[l] = e^{-i \operatorname{Im}(\xi, \varphi'[l])}\Phi[l]. \tag{3.20}$$

Certain matrix elements are insensitive to such phase changes. In particular, for any operator B it follows that

$$\frac{(\Phi'[l_1], B\Phi'[l_2])}{(\Phi'[l_1], \Phi'[l_2])} = \frac{(\Phi[l_1], B\Phi[l_2])}{(\Phi[l_1], \Phi[l_2])} \tag{3.21}$$

and, thus, for diagonal matrix elements, that

$$(\Phi'[l], B\Phi'[l]) = (\Phi[l], B\Phi[l]). \tag{3.22}$$

The annihilation operator associated with the translated parameterization is defined by the property that

$$A'(\lambda)\Phi'[l] = (\lambda, \varphi'[l])\Phi'[l]. \tag{3.23a}$$

Since a phase factor cannot affect this, we must also have

$$A'(\lambda)\Phi[l] = (\lambda, \varphi'[l])\Phi[l] = (\lambda, \varphi[l] - \xi)\Phi[l]. \tag{3.23b}$$

It follows that

$$A'(\lambda) = A(\lambda) - (\lambda, \xi) \tag{3.24a}$$

and, consequently,

$$A'(\lambda)^\dagger = A(\lambda)^\dagger - (\xi, \lambda). \tag{3.24b}$$

By direct computation we see that

$$\begin{aligned}[A'(\lambda_1), A'(\lambda_2)^\dagger] &= [A(\lambda_1), A(\lambda_2)^\dagger] \\ &= (\lambda_1, \lambda_2).\end{aligned} \tag{3.25}$$

The representation A' of the commutation relations is, in fact, unitarily equivalent to the representation A, since we have assumed $\xi \in \mathfrak{h}$.[17] Specifically,

$$V^{-1}A'(\lambda)V = A(\lambda), \quad \text{for all} \quad \lambda \in \mathfrak{h}, \tag{3.26}$$

where

$$V = \exp\,[A(\xi)^\dagger - A(\xi)]. \tag{3.27}$$

(Shortly we shall break the assumption that $\xi \in \mathfrak{h}$ and actually be dealing with unitarily inequivalent representations A' and A.)

The kind of change involved in the annihilation and creation operators carries over to many other operators as well. Suppose we consider

$$(\Phi'[l_1], e^{-iW't}\Phi'[l_2]) = N_1'N_2' \exp\,\{\varphi'[l_1], e^{-iwt}\varphi'[l_2])\}. \tag{3.28}$$

Then the same argument as before shows that

$$W' = (A', wA') \tag{3.29}$$

in keeping with the fact that

$$\frac{(\Phi'[l_1], W'\Phi'[l_2])}{(\Phi'[l_1], \Phi'[l_2])} = (\varphi'[l_1], w\varphi'[l_2]). \tag{3.30}$$

In addition,

$$(\Phi'[l], W'\,\Phi'[l]) = (\Phi[l], W'\Phi[l]) = (\varphi'[l], w\varphi'[l]), \tag{3.31}$$

which is different from the expression that would follow from (3.4), in which

$$W = (A, wA) = (A' + \xi, w[A' + \xi]). \tag{3.32}$$

Among all the possible choices of translation vectors ξ we can always choose $\xi \equiv \varphi[l_0] \in \mathfrak{s}$ for some $l_0 \in \mathfrak{L}$, so that

$$\varphi'[l_0] = \varphi[l_0] - \xi = 0. \tag{3.33}$$

That is, even if $\varphi[l] \neq 0$ for all $l \in \mathfrak{L}$, we can always choose ξ so that in the *translated* set $\mathfrak{s}' = \mathfrak{s} - \xi$ the zero vector appears. A parameterization, such that $\varphi'[l_0] = 0$, exhibits certain additional properties. For example, we note first that

$$(\Phi'[l], \Phi'[l_0]) = e^{-\frac{1}{2}\|\varphi'[l]\|^2} = N', \tag{3.34}$$

which is a *real* expression for all $l \in \mathfrak{L}$. To make the next point, let us assume that w is an arbitrary self-adjoint operator on $\mathfrak{h}$. In general, the only eigenvectors of w which are imaged directly into eigenvectors of W by the exponential map $\varphi \to \Psi$ are those with eigenvalue zero. However, when $\varphi'[l_0] = 0$ is an element of $\mathfrak{s}$, it follows from (3.28) that its image $\Phi'[l_0]$ is an eigenvector with eigenvalue zero for *every* self-adjoint operator $W' = (A', wA')$. This is a particularly important property. Moreover, if w has no eigenvectors with eigenvalue zero, then $\Phi'[l_0]$ is a nondegenerate eigenvector with eigenvalue zero for the operator $W' = (A', wA')$. In summary, we note that if $\mathfrak{s}$ does not already contain the zero vector a translated set $\mathfrak{s}' = \mathfrak{s} - \xi$ may be considered which, by suitable choice of ξ, does contain the zero vector.

1. Improper Translations

The transition to a translated set assumes its primary importance when the transformation is an improper one. Suppose that $\varphi'[l] \in \mathfrak{s}' \subset \mathfrak{h}$ constitutes a valid initial set of vectors. Let us consider, in addition, the relation

$$\varphi'[l] = \hat{\varphi}[l] - \hat{\xi}, \tag{3.35}$$

where in the present case we admit the possibility that $\hat{\xi} \notin \mathfrak{h}$, i.e., $\|\hat{\xi}\| = \infty$, in which case $\hat{\varphi}[l] \notin \mathfrak{h}$ for all $l \in \mathfrak{L}$. To distinguish this possibility, we have resorted to the carat over the usual vector symbol. Note that although $\hat{\varphi}$ and $\hat{\xi}$ may not be elements of $\mathfrak{h}$, their formal difference lies in $\mathfrak{h}$ by assumption. Such a relation can always be viewed as the limit of valid vectors in $\mathfrak{h}$, such as

$$\varphi'[l] = \hat{\varphi}[l] - \hat{\xi} \equiv \lim_{n\to\infty}\{\varphi_n[l] - \xi_n\}, \tag{3.36}$$

where $\varphi_n, \xi_n \in \mathfrak{h}$, for all n. Abstractly, we may regard $\hat{\varphi} = \lim \varphi_n$ and $\hat{\xi} = \lim \xi_n$ as elements of the distribution space (dual to the test function, or nuclear space) in a rigged Hilbert space triplet.[18] Indeed we shall loosely use $\hat{\varphi}$ and $\hat{\xi}$ in inner products in this way. In a representation on some L^2 space, it generally suffices that $\hat{\varphi}$ and $\hat{\xi}$ are functions which in the present case need not be square integrable. To make this notion clearer, let us reconsider the example of unitary

[17] K. O. Friedrichs, Ref. 2, p. 79.

[18] I. M. Gel'fand and N. Y. Vilenkin, Ref. 4, Chap. III, Sec. 4.

representations for the operators $U(t) = \exp(-iWt)$, treated earlier.

Example revisited: In the present formulation of the example we seek a set of vectors of the form

$$\Phi'[t] = e^{-i\,\mathrm{Im}\,(\hat{\xi},\varphi'[t])}U(t)\Phi_0, \tag{3.37}$$

for which an exponential construction exists, i.e.,

$$(\Phi'[t_1], \Phi'[t_2]) = N_1' N_2' e^{(\varphi'[t_1],\varphi'[t_2])}. \tag{3.38}$$

We assume $\mathrm{Im}\,(\hat{\xi}, \varphi'[0]) = 0$, and, thus, $\Phi'[0] = \Phi_0$. The combination law $U(t)U(t') = U(t + t')$ now leads to

$$(\Phi'[t_1], \Phi'[t_2]) = (\Phi_0, \Phi'[t_2 - t_1])e^{i\alpha}, \tag{3.39}$$

where

$$\alpha \equiv \mathrm{Im}\,(\hat{\xi}, \varphi'[t_2 - t_1] - \varphi'[t_2] + \varphi'[t_1]). \tag{3.40}$$

Equating real and imaginary terms in the exponent, we find that

$$\|\varphi'[t_1] - \varphi'[t_2]\|^2 = \|\varphi'[0] - \varphi'[t_2 - t_1]\|^2, \tag{3.41a}$$

while

$$\mathrm{Im}\,(\varphi'[t_1], \varphi'[t_2]) = \mathrm{Im}\,(\varphi'[0], \varphi'[t_2 - t_1]) + \mathrm{Im}\,(\hat{\xi}, \varphi'[t_2 - t_1] - \varphi'[t_2] + \varphi'[t_1]). \tag{3.41b}$$

Apart from the last term in $\hat{\xi}$, these relations are identical to those found earlier in (3.9). As a solution to the first relation, we now choose

$$\varphi'[t] = a + bt \oplus \int^{\oplus} (e^{-itx} - 1)\, d\hat{\mu}(x), \tag{3.42}$$

instead of (3.10), which leads to the expression

$$\|\varphi'[t]\|^2 = |a + bt|^2 + \int |e^{-itx} - 1|^2\, d\hat{\mu}(x). \tag{3.43}$$

Note, in the present case, that a finite norm requires only that

$$\int \left(\frac{x^2}{1 + x^2}\right) d\hat{\mu}(x) < \infty, \tag{3.44}$$

while it is possible that $\hat{\mu}(R) = \int d\hat{\mu}(x) = \infty$, in contrast with the requirement that $\mu(R) < \infty$. This is just the type of generalization in this class of representations which we sought. Let us adopt

$$\hat{\xi} = \frac{a}{c}\, Y \oplus \int^{\oplus} d\hat{\mu}(x), \tag{3.45a}$$

where $c = \mathrm{Im}\, ba^*$ as before and

$$Y \equiv \int \frac{x}{1 + x^2}\, d\hat{\mu}(x). \tag{3.45b}$$

With this choice we find that

$$\mathrm{Im}\,(\hat{\xi}, \varphi'[t]) = \int \left(\frac{xt}{1 + x^2} - \sin xt\right) d\hat{\mu}(x), \tag{3.46}$$

which is well defined in virtue of (3.44); moreover, we find that the second relation (3.41b) is satisfied.

In summary, we have determined that

$$\begin{aligned}(\Phi_0, \Phi'[t]) &= \exp\left(ict - \tfrac{1}{2}|b|^2 t^2 - \tfrac{1}{2}\int |e^{-itx} - 1|^2\, d\hat{\mu}(x)\right) \\ &= e^{-i\,\mathrm{Im}\,(\hat{\xi},\varphi'[t])}(\Phi_0, U(t)\Phi_0),\end{aligned} \tag{3.47}$$

from which it follows that

$$(\Phi_0, e^{-iWt}\Phi_0) = \exp\left[ict - \tfrac{1}{2}|b|^2 t^2 - \int \left(1 - \frac{itx}{1 + x^2} - e^{-itx}\right) d\hat{\mu}(x)\right]. \tag{3.48}$$

In this final expression we have arrived at the most general characteristic function for an infinitely divisible probability distribution (compatible with an infinite-dimensional $\mathfrak{H}$); in fact, Eq. (3.48), coupled with (3.44), is just the Lévy canonical representation for such distributions.[19] The translations $\hat{\xi}$ are not restricted to elements of $\mathfrak{h}$. This is clear since

$$\|\hat{\xi}\|^2 = |a/c|^2\, Y^2 + \hat{\mu}(R), \tag{3.49}$$

neither term of which need be finite. In this particular example the translated set $\mathfrak{s}'$ does not contain the zero vector unless $a = 0$, as can be seen from (3.43).

2. *Improper Translations and Unitary Inequivalence*

In our initial discussion of those representations of $U(t) = \exp(-iWt)$ admitting an exponential construction, our parameterization was directly associated with the group combination law. To achieve the utmost generality in such representations, it was expedient to consider a translated parameterization, a translation which may well be an improper one. Nevertheless, we have shown that an exponential construction encompasses the set of infinitely divisible probability distributions, and that any unitary, one-parameter group of operators, realized in an exponential way, is such a distribution. Of primary importance, of course, is the additional machinery we have introduced to realize and represent various associated operators.

When the translation vector $\hat{\xi}$ is improper, the operators $A'(\lambda)$ and $A(\lambda)$ are unitarily inequivalent.[17]

[19] Ref. 15, p. 90.

Heuristically, this is almost evident from the connecting relation

$$A(\lambda) = A'(\lambda) + (\lambda, \hat{\xi}). \tag{3.50}$$

For if $\hat{\xi}$ is improper, i.e., $\hat{\xi} \notin \mathfrak{h}$, then, for some $\lambda = \lambda_0 \in \mathfrak{h}$, the expression $(\lambda_0, \hat{\xi}) = \infty$ and thus, although $A'(\lambda_0)$ is an acceptable operator, $A(\lambda_0)$ is not.[20] This could not happen if the two operator sets were unitarily equivalent. The operator $A(\lambda)$ is defined now for those $\lambda \in \mathfrak{h}$ such that $|(\lambda, \hat{\xi})| < \infty$. If $\mathfrak{h}$ has infinite dimensionality, then $A(\lambda)$ is defined for a dense set of λ. It follows that the operators $A(\lambda_1)$ and $A(\lambda_2)^\dagger$ constitute an inequivalent, irreducible representation of the canonical commutation relations:

$$[A(\lambda_1), A(\lambda_2)^\dagger] = [A'(\lambda_1), A'(\lambda_2)^\dagger] = (\lambda_1, \lambda_2). \tag{3.51}$$

For completeness we quote the complete theorem on equivalence of "translated Fock representations": Given a Fock representation $A'(\lambda)$ and two translated Fock representations

$$A_1(\lambda) = A'(\lambda) + (\lambda, \hat{\xi}_1), \tag{3.52a}$$

$$A_2(\lambda) = A'(\lambda) + (\lambda, \hat{\xi}_2), \tag{3.52b}$$

defined for a set of λ dense in $\mathfrak{h}$, then $A_1(\lambda)$ is unitarily equivalent to $A_2(\lambda)$ if and only if $(\hat{\xi}_1 - \hat{\xi}_2) \in \mathfrak{h}$.[17,21] This may be made concrete by choosing an orthonormal sequence λ_m in the common dense domain of definition and requiring that

$$\sum |\hat{\xi}_{1m} - \hat{\xi}_{2m}|^2 < \infty, \tag{3.53}$$

where $\hat{\xi}_{1m} = (\lambda_m, \hat{\xi}_1)$, etc.

B. Operator-Field Representations: Space-Translation Invariance and Cluster Decomposition

Consider a field operator $W(\mathbf{x})$, where for concreteness we may take $\mathbf{x}$ as a point in Euclidean three-dimensional space R^3, which satisfies the commutation relation

$$[W(\mathbf{x}), W(\mathbf{y})] = 0. \tag{3.54}$$

We assume W to be formally self-adjoint so that the smeared operators

$$W(f) = \int W(\mathbf{x}) f(\mathbf{x})\, d\mathbf{x} \tag{3.55}$$

are, for suitable real $f(\mathbf{x})$, self-adjoint operators. As a suitable test function space we adopt Schwartz's space $\mathfrak{D}$ composed of real C^∞ functions $f(\mathbf{x})$ which have compact support.

[20] Ref. 14, p. 78.

[21] J. R. Klauder, J. McKenna, and E. J. Woods, J. Math. Phys. 7, 822 (1966).

In many applications a central question concerns the representation of the field operators $W(f)$. We approach this question, as before, by considering the expectation functional

$$E(f) \equiv (\Phi_0, \exp[-iW(f)]\Phi_0) \tag{3.56}$$

and imposing an exponential construction. Both the natural ($\hat{\xi} = 0$) and translated parameterizations ($\hat{\xi} \neq 0$) are important, and we shall illustrate them both. We adopt $\mathfrak{D}$ as the label space $\mathfrak{L}$ and $f(\mathbf{x})$ as the labels l.

If we set

$$\Phi'[f] = e^{-i\,\mathrm{Im}\,(\hat{\xi}, \varphi'[f])} e^{-iW(f)} \Phi_0, \tag{3.57}$$

then we seek representations consistent with the relation

$$(\Phi'[f_1], \Phi'[f_2]) = N_1' N_2' e^{(\varphi'[f_1], \varphi'[f_2])}, \tag{3.58a}$$

where

$$N' = e^{-\frac{1}{2}\|\varphi'[f]\|^2}. \tag{3.58b}$$

As before we assume that $\mathrm{Im}\,(\hat{\xi}, \varphi'[0]) = 0$. The combination law

$$e^{-iW(f_1)} e^{-iW(f_2)} = e^{-iW(f_1+f_2)} \tag{3.59}$$

leads, just as in the elementary case, to the two relations

$$\|\varphi'[0] - \varphi'[f_1 - f_2]\|^2 = \|\varphi'[f_2] - \varphi'[f_1]\|^2, \tag{3.60a}$$

$$\mathrm{Im}\,(\varphi'[0], \varphi'[f_1 - f_2]) = \mathrm{Im}\,(\varphi'[f_2], \varphi'[f_1]) - \mathrm{Im}\,(\hat{\xi}, \varphi'[f_1 - f_2] - \varphi'[f_1] + \varphi'[f_2]). \tag{3.60b}$$

There are various solutions to these relations, a few of which we shall illustrate. In the natural or group-oriented parameterization, we set $\hat{\xi} = 0$, and, for example, we may adopt

$$\varphi'[f] = \varphi[f] = (bf)(\mathbf{x}) \oplus \int^{\oplus} e^{-i(w,f)}\, d\mu(w). \tag{3.61}$$

Here b is a general linear operator from $\mathfrak{D}$ into L^2. The expression (w, f) is a real number for all f which implies that w is a (real) distribution in $\mathfrak{D}'$. Consequently, $\mu(w)$ is a measure on $\mathfrak{D}'$, which being the dual of a nuclear space is well defined.[22] We have arbitrarily chosen only the linear term $(bf)(\mathbf{x})$ in the present solution [cf. Eq. (3.10)]. In order for $\varphi[f]$ to have finite norm, it is clear that $\mu(\mathfrak{D}') \equiv \int d\mu(w) < \infty$. It follows that

$$(\Phi_0, e^{-iW(f)}\Phi_0) = \exp\left\{-\tfrac{1}{2}\int |bf(\mathbf{x})|^2\, d\mathbf{x} - \int [1 - e^{-i(w,f)}]\, d\mu(w)\right\}. \tag{3.62}$$

This expression characterizes a vast number of field-operator distributions. As a very simple example,

[22] I. M. Gel'fand and N. Y. Vilenkin, Ref. 4, Chap. IV, Sec. 2.

suppose that μ is concentrated on Dirac δ functions at the point $\mathbf{x} = \mathbf{z}$. In other words, let

$$\int e^{-i(w,f)}\, d\mu(w) = \int e^{-i(\lambda\delta_z,f)}\, d\sigma(\lambda) = \int e^{-i\lambda f(\mathbf{z})}\, d\sigma(\lambda), \tag{3.63a}$$

where λ is a single real variable and

$$\mu(\mathfrak{D}') = \int d\mu(w) = \int d\sigma(\lambda) < \infty. \tag{3.63b}$$

The inclusion of several derivatives of the δ function is an obvious generalization of this example—and many more come to mind. However, more important representations may be obtained if we exploit a translated parameterization with an improper translation.

For the second class of examples we choose as solution to (3.60) the vectors

$$\hat{\xi} = 0 \oplus \int^{\oplus} d\hat{\mu}(w), \tag{3.64a}$$

$$\varphi'[f] = (bf)(\mathbf{x}) \oplus \int^{\oplus} [e^{-i(w,f)} - 1]\, d\hat{\mu}(w). \tag{3.64b}$$

Note that this is a different type of translation vector from the one used before, since we principally wish to emphasize a different type of improper translation. Because $\varphi'[0] = 0$, it follows that

$$(\Phi_0, \Phi'[f]) = e^{-\frac{1}{2}\|\varphi'[f]\|^2} = \exp\left[-\tfrac{1}{2}\int |bf(\mathbf{x})|^2\, d\mathbf{x} - \tfrac{1}{2}\int |e^{-i(w,f)} - 1|^2\, d\hat{\mu}(w)\right]. \tag{3.65}$$

From this expression we learn that

$$(\Phi_0, e^{-iW(f)}\Phi_0) = \exp\left\{-\tfrac{1}{2}\int |bf(\mathbf{x})|^2\, d\mathbf{x} - \int [1 - e^{-i(w,f)}]\, d\hat{\mu}(w)\right\}. \tag{3.66}$$

These relations do not require that

$$\hat{\mu}(\mathfrak{D}') \equiv \int d\hat{\mu}(w) = \|\hat{\xi}\|^2 \tag{3.67}$$

be finite; in fact, we may choose many improper translation vectors $\hat{\xi}$.

As a very simple example, suppose $\hat{\mu}$ is again concentrated on δ functions at $\mathbf{x} = \mathbf{z}$, but now with a uniform distribution of $\mathbf{z}$ values. In particular, we assume that

$$\int [1 - e^{-i(w,f)}]\, d\hat{\mu}(w) = \int [1 - e^{-i(\lambda\delta_z,f)}]\, d\hat{\mu}_0(\lambda)\, d\mathbf{z} = \int [1 - e^{-i\lambda f(\mathbf{z})}]\, d\hat{\mu}_0(\lambda)\, d\mathbf{z}. \tag{3.68}$$

Although this corresponds to an improper translation $\hat{\xi}$, since

$$\|\hat{\xi}\|^2 = \int d\hat{\mu}(w) = \int d\hat{\mu}_0(\lambda)\, d\mathbf{z} = \infty, \tag{3.69}$$

the desired relation (3.68) is finite since $f(\mathbf{z}) = 0$ outside a compact set in R^3. As it stands, this relation is also finite so long as $\int [|\lambda|/(1 + |\lambda|)]\, d\hat{\mu}_0(\lambda) < \infty$. If $\hat{\mu}_0(-\lambda) = -\hat{\mu}_0(\lambda)$, then only the even part of (3.68) is nonzero and the condition for existence is the finiteness of

$$\int \frac{\lambda^2}{(1 + \lambda^2)}\, d\hat{\mu}_0(\lambda). \tag{3.70}$$

Even if such symmetry is not present, a different translation vector, patterned after the earlier example [cf. Eq. (3.45)], gives a more general expression. For the most part we are content to assume such symmetry for the purpose of illustration.

1. Space-Translation Invariance

In order that the vector Φ_0 possess space-translation invariance, it is necessary that

$$(\Phi_0, e^{-iW(f_a)}\Phi_0) = (\Phi_0, e^{-iW(f)}\Phi_0) \tag{3.71}$$

for all $f(\mathbf{x}) \in \mathfrak{D}$ and all $\mathbf{a} \in R^3$, where

$$f_a(\mathbf{x}) \equiv f(\mathbf{x} + \mathbf{a}). \tag{3.72}$$

An example of such behavior is given by (3.65), where b is simply a real constant and where (3.68) is adopted. An extension of this example is given by

$$(\Phi_0, e^{-iW(f)}\Phi_0) = \exp\left\{-\tfrac{1}{2}b^2\int f^2(\mathbf{x})\, d\mathbf{x} - \int [1 - e^{-i(w,f_z)}]\, d\hat{\mu}_0(w)\, d\mathbf{z}\right\}, \tag{3.73}$$

where f_z is the translated test function (3.72) and where $\hat{\mu}_0(w)$ is concentrated on distributions "at or near" $\mathbf{x} = 0$. For example, we could assume that, for each $f(\mathbf{x}) \in \mathfrak{D}$ such that $f(\mathbf{x}) = 0$ for $|\mathbf{x}| < R_0$, the distributions of interest would fulfill $(w, f) = 0$. The translation invariance of the expression (3.73) [in the sense of (3.71)] is evident.

We may profitably write $\mathfrak{h}$ as a direct integral space over R^3 when representations such as (3.73) apply.

That is, we set

$$\mathfrak{h} = \int^{\oplus} \mathfrak{h}_z \, dz \tag{3.74}$$

and

$$\varphi'[f] = \int^{\oplus} \varphi'_z[f] \, dz, \tag{3.75}$$

so that

$$\|\varphi'[f]\|^2 = \int (\varphi'_z[f], \varphi'_z[f])_z \, dz. \tag{3.76}$$

In this language, the example of (3.73) has a direct integral entry of the form

$$\varphi'_z[f] = bf(\mathbf{z}) \oplus \int^{\oplus} [e^{-i(w,f_z)} - 1] \, d\hat{\mu}_0(w). \tag{3.77}$$

Among the possible representations of this particular form are those for which

$$\varphi'_z[f] = \varphi'_z[f(\mathbf{z})], \tag{3.78}$$

i.e., only the specific value $f(\mathbf{z})$ enters. Equation (3.68) has this "ultralocal" characteristic which arises when $\hat{\mu}_0$ is concentrated on δ functions at $\mathbf{x} = 0$, and these are essentially the only such representations. This class of representations has the property that

$$E(f_1 + f_2) = E(f_1)E(f_2) \tag{3.79}$$

whenever

$$f_1(\mathbf{x})f_2(\mathbf{x}) = 0, \tag{3.80}$$

where $E(f)$ is given by (3.56). Special representations of this form have been formulated as continuous-tensor-product representations, and have been studied by several authors.[7,23]

2. *Cluster Decomposition*

The physical idea behind cluster decomposition is that equal-time field operators localized about remote regions of space are statistically independent. It is a familiar property for statistically independent variables that the probability (amplitude) and the related characteristic function factorize. Such factorization is just the statement of (3.79) for the special class of ultralocal representations. More generally, if we assume translation invariance, $E(f_a) = E(f)$, then, to ensure cluster decomposition, it is necessary that

$$\lim_{|a|\to\infty} E(f' + f_a) = E(f')E(f), \quad \text{for all } f', f \in \mathfrak{D}. \tag{3.81}$$

Examples of translationally invariant distributions that exhibit cluster decomposition are conveniently formulated in an exponential Hilbert space. If we assume $\varphi'[0] = 0$, then we have the relation

$$E(f) = (\Phi_0, \Phi'[f])e^{i \operatorname{Im}(\hat{\xi}, \varphi'[f])} = e^{-\frac{1}{2}\|\varphi'[f]\|^2 + i \operatorname{Im}(\xi, \varphi'[f])}. \tag{3.82}$$

To win the desired features, we must have invariance:

$$\|\varphi'[f_a]\|^2 = \|\varphi'[f]\|^2, \tag{3.83a}$$

$$\operatorname{Im}(\hat{\xi}, \varphi'[f_a]) = \operatorname{Im}(\hat{\xi}, \varphi'[f]); \tag{3.83b}$$

and asymptotic independence:

$$\lim_{|a|\to\infty} \|\varphi'[f' + f_a]\|^2 = \|\varphi'[f']\|^2 + \|\varphi[f]\|^2, \tag{3.84a}$$

$$\lim_{|a|\to\infty} \operatorname{Im}(\hat{\xi}, \varphi'[f' + f_a]) = \operatorname{Im}(\hat{\xi}, \varphi'[f']) + \operatorname{Im}(\hat{\xi}, \varphi'[f]). \tag{3.84b}$$

In the direct integral form of (3.74)–(3.76), we have

$$\|\varphi'[f]\|^2 = \int \|\varphi'_z[f]\|_z^2 \, dz, \tag{3.85}$$

which to exhibit invariance need only fulfill

$$\|\varphi'_{z+a}[f_a]\|_{z+a}^2 = \|\varphi'_z[f]\|_z^2. \tag{3.86a}$$

Asymptotic independence is fulfilled if

$$\lim_{|a|\to\infty} \|\varphi'_z[f' + f_a]\|_z^2 = \|\varphi'_z[f']\|_z^2. \tag{3.86b}$$

These relations are consistent with the intuitive notion that $\varphi'_z[f]$ should depend only on the values of $f(\mathbf{x})$ when $\mathbf{x}$ is "near" $\mathbf{z}$. Equation (3.73) gives an example of this type.

These examples, as well as our general discussion, should serve to demonstrate how ideal the exponential Hilbert space construction really is for representing field operators and invariant vectors exhibiting space-translation invariance and cluster decomposition. In the next two sections we discuss exponential representations of current algebras and of the familiar canonical commutation relations for scalar fields.

4. EXPONENTIAL REPRESENTATIONS OF FIELD-OPERATOR ALGEBRAS

We consider an equal-time field algebra (or "current" algebra of particle physics) which is characterized by a family of formally self-adjoint field operators $W_l(\mathbf{x})$, $l = 1, 2, \cdots, L$, with the commutation rule

$$[W_l(\mathbf{x}), W_m(\mathbf{y})] = ic_{lmn}W_n(\mathbf{x})\delta(\mathbf{x} - \mathbf{y}). \tag{4.1}$$

Here c_{lmn} are the structure constants of a Lie group, and summation over repeated indices is understood. If $f(\mathbf{x})$, $g(\mathbf{x})$ are suitable real test functions, such as elements of $\mathfrak{D}$, then

$$[W_l(f), W_m(g)] = ic_{lmn}W_n(fg) \tag{4.2}$$

[23] Such representations for canonical operators were effectively first treated by H. Araki, thesis, Princeton University, Princeton, N.J., 1960.

is the proper statement of (4.1) for self-adjoint operators. The unitary group elements in canonical coordinates are given by

$$U[f_l] = e^{iW_l(f_l)} \tag{4.3}$$

(summation understood in the exponent on the right-hand side), where $f_l(\mathbf{x}) \in \mathfrak{D}$ for all l; it is representations of these field operators that we seek.

Guided by our formulation in Sec. 3, we introduce the overcomplete family of states

$$\Phi'[f_l] \equiv e^{-i\,\mathrm{Im}\,(\hat{\xi},\varphi'[f_l])}U[f_l]\Phi_0, \tag{4.4}$$

for some choice of $\hat{\xi}$ and of $\varphi'[f_l] \in \mathfrak{h}$. We insist on an exponential construction such that

$$(\Phi'[f_l'], \Phi'[f_l]) = N'Ne^{(\varphi'[f_l'],\varphi'[f_l])} \tag{4.5}$$

The group law reads

$$U[f_l]U[\bar{f}_l] \equiv U[(f\cdot\bar{f})_l], \tag{4.6}$$

where $f\cdot\bar{f}$ symbolizes the parametric combination law characterizing the group. This law implies a functional identity within the exponential form that leads to the two relations

$$\|\varphi'[f_l'] - \varphi'[f_l]\|^2 = \|\varphi'[0] - \varphi'[(-f'\cdot f)_l]\|^2, \tag{4.7a}$$

$$\begin{aligned}\mathrm{Im}\,(\varphi'[f_l'], \varphi'[f_l]) &= \mathrm{Im}\,(\varphi'[0], \varphi'[(-f'\cdot f)_l]) \\ &\quad - \mathrm{Im}\,(\hat{\xi}, \varphi'[(-f'\cdot f)_l] - \varphi'[f_l] + \varphi'[f_l']).\end{aligned} \tag{4.7b}$$

Solutions to these equations generate the representations we seek.

As our first solution, we adopt

$$\hat{\xi} = 0, \tag{4.8a}$$

$$\varphi'[f_l] \equiv \varphi[f_l] = u[f_l]\varphi_0, \tag{4.8b}$$

where the $u[f_l]$ are unitary operators which satisfy

$$u[f_l']u[f_l] = u[(f'\cdot f)_l]; \tag{4.9}$$

i.e., $u[f_l]$ also forms a unitary representation of the sought-for group. Indeed, $U[f_l]$ is related to $u[f_l]$ in the fashion that B is related to b in Eq. (2.27), namely,

$$U[f_l] = \bigoplus_{n=0}^{\infty} \bigotimes_{s=1}^{n} u[f_l]. \tag{4.10}$$

Moreover, in canonical coordinates,

$$U[f_l] = e^{iW_l(f_l)}, \tag{4.11a}$$

$$u[f_l] = e^{iw_l(f_l)}, \tag{4.11b}$$

so that an equation like (2.29) holds as well. Thus, in this solution it is clear that the representation $U[f_l]$ is always *reducible*, whether $u[f_l]$ is irreducible or not. Equivalent (inequivalent) representations of $u[f_l]$ lead to equivalent (inequivalent) representations of $U[f_l]$. In order that $U[f_l]$ be a cyclic representation with Φ_0 a cyclic vector, it is necessary that $u[f_l]$ be a cyclic representation (on $\mathfrak{h}$) with φ_0 a cyclic vector. However, it is not *a priori* clear that this is sufficient in the general case.

With the solution (4.8) the expectation functional takes the form

$$\begin{aligned}E(f_l) &\equiv (\Phi_0, e^{iW_l(f_l)}\Phi_0) \\ &= \exp\{(\varphi_0, [e^{iw_l(f_l)} - 1]\varphi_0)\},\end{aligned} \tag{4.12}$$

as follows from (2.35). In the present parameterization $A(\lambda)\Phi_0 = (\lambda, \varphi_0)\Phi_0$ and thus, according to (2.32) and (2.33), we find

$$W_l(f_l) = (A, w_l(f_l)A). \tag{4.13a}$$

Hence, for each l we have

$$W_l(f) = (A, w_l(f)A). \tag{4.13b}$$

It is a well-known property of creation and annihilation operators that

$$\begin{aligned}[W_l(f), W_m(g)] &= [(A, w_l(f)A), (A, w_m(g)A)] \\ &= (A, [w_l(f), w_m(g)]A).\end{aligned} \tag{4.14}$$

Thus, the validity of the commutation relations (4.2) follows from (4.14) and the fact that the $w_l(f)$ fulfill the same algebra on $\mathfrak{h}$.

The reducible nature of the representations given by $W_l(f)$ may be seen another way. Let

$$N = (A, A) \tag{4.15}$$

denote the total number operator in the Fock representation. If

$$\mathbf{X} = \bigoplus_{n=0}^{\infty} \mathbf{X}_{(n)} \in \mathfrak{H}, \tag{4.16}$$

then

$$N\mathbf{X} \equiv \bigoplus_{n=0}^{\infty} n\mathbf{X}_{(n)}, \tag{4.17}$$

so that N has zero and the positive integers for eigenvalues. Now, no matter what the operators w_l are, it follows that

$$\begin{aligned}[W_l(f), N] &= [(A, w_l(f)A), (A, A)] \\ &= (A, [w_l(f), I]A) = 0.\end{aligned} \tag{4.18}$$

Implicitly, we have demonstrated reducibility by showing the existence of an operator N different from the identity operator, which commutes with all the generators.

As a second solution to relations (4.7), let us select

$$\hat{\xi} = \hat{\varphi}_0, \tag{4.19a}$$

$$\varphi'[f_l] = (u[f_l] - 1)\hat{\varphi}_0, \tag{4.19b}$$

where again $u[f_l]$ is a unitary group representation on $\mathfrak{h}$. Although we may give up the finiteness of $\|\hat{\xi}\| = \|\hat{\varphi}_0\|$, we assume that $\varphi'[f_l] \in \mathfrak{h}$ and that $(\hat{\varphi}_0, \varphi'[f_l])$ is well defined. In particular, we are immediately led to representations which are essentially characterized by

$$(\Phi_0, e^{iW_l(f_l)}\Phi_0) = \exp\{(\hat{\varphi}_0, [e^{iw_l(f_l)} - 1]\hat{\varphi}_0)\}. \tag{4.20}$$

These representations, of course, are not fundamentally different than those given above unless $\|\hat{\varphi}_0\| = \infty$. As an important example of the difference that this condition makes, let us suppose that $w \equiv w_l(f_l)$, for some choice of f_l, only has an absolutely continuous spectrum. Then

$$\text{w-lim}_{t\to\infty} e^{iwt} = 0, \tag{4.21}$$

where the weak operator limit is meant. If $\|\hat{\varphi}_0\| < \infty$, then

$$\text{w-lim}_{t\to\infty} e^{iWt} = P_0, \tag{4.22a}$$

where $W = W_l(f_l)$ and where P_0 is a projection operator onto the subspace of zero-W eigenvalue. This subspace always includes the vector $\hat{\Phi}_0$, which is characterized by the fact that $A(\lambda)\hat{\Phi}_0 = 0$ for all λ. On the other hand, for suitable choices of $\hat{\varphi}_0 \notin \mathfrak{h}$, it follows that

$$\text{w-lim}_{t\to\infty} e^{iWt} = 0, \tag{4.22b}$$

so that, like w, the associated operator W has only an absolutely continuous spectrum. This is a common and frequently desirable property of field operators and should be incorporated whenever suitable. An example of the two cases in (4.22) is given later in this section.

In the translated form appropriate to (4.19), we have, according to (3.23a), the basic annihilation operators $A'(\lambda)$ such that

$$A'(\lambda)\Phi'[f_l] = (\lambda, \varphi'[f_l])\Phi'[f_l].$$

Since $\varphi'[f_l] \in \mathfrak{h}$, this relation is defined for all $\lambda \in \mathfrak{h}$. Indeed, $\varphi'[0] = 0$, so that $A'(\lambda)\Phi'[0] = 0$, which shows that $A'(\lambda)$ is surely the Fock representation. On the other hand, the group generators W_l are still given by (4.13) in terms of the operators A. The relation between the two, as in (3.50), reads $A(\lambda) = A'(\lambda) + (\lambda, \hat{\varphi}_0)$, which, if $\hat{\varphi}_0 \notin \mathfrak{h}$, is not defined for all λ. By hypothesis, however, $\varphi'[f_l] = (e^{iw_l(f_l)} - 1)\hat{\varphi}_0 \in \mathfrak{h}$ and is, in fact, "enough" within $\mathfrak{h}$, we assume, so that $(\hat{\varphi}_0, \varphi'[f_l])$ makes sense; this notion will be made clearer below through an example. We assume for simplicity that even $w_l(f_l)\hat{\varphi}_0 \in \mathfrak{h}$, and, moreover, that $(\hat{\varphi}_0, w_l(f_l)\hat{\varphi}_0)$ makes sense. With this simplification we have

$$\begin{aligned} W_l(f) &= (A, w_l(f)A) \\ &= (A' + \hat{\varphi}_0, w_l(f)[A' + \hat{\varphi}_0]) \\ &= (A', w_l(f)A') + (A', w_l(f)\hat{\varphi}_0) \\ &\quad + (w_l(f)\hat{\varphi}_0, A') + (\hat{\varphi}_0, w_l(f)\hat{\varphi}_0). \end{aligned} \tag{4.23}$$

While this equation still has the same appearance as (4.13b), it is, of course, quite different, since A is no longer a Fock representation. This inequivalence in representations suggests that, when $\hat{\varphi}_0 \notin \mathfrak{h}$, we are dealing with an *inequivalent* representation of $U[f_l]$ (with no change of the representation $u[f_l]$). More generally, consider two such representations

$$W_{l1}(f) = (A_1, w_l(f)A_1), \tag{4.24a}$$

$$W_{l2}(f) = (A_2, w_l(f)A_2), \tag{4.24b}$$

where A_1 and A_2 are related to a standard Fock representation A', as in Eq. (3.52). The first representation W_{l1} is unitarily equivalent to the second W_{l2}, if A_1 is unitarily equivalent to A_2; but this is not the only possibility. Unitary equivalence of W_{l1} and W_{l2} is also assured, if A_1 is unitarily equivalent to $e^{i\gamma}A_2$; here, γ is any over-all phase factor, since such a factor obviously drops out of (4.24b). In turn, this holds if and only if $\hat{\xi}_1 - e^{i\gamma}\hat{\xi}_2 \in \mathfrak{h}$, for some real γ. This exhausts the equivalence class of unitarily equivalent representations, unless there are additional unitary operators v on $\mathfrak{h}$ which commute with all the $w_l(f)$. For any v such that

$$v^\dagger w_l(f)v = w_l(f), \tag{4.25}$$

the representations W_{l1} and W_{l2} are unitarily equivalent, provided that

$$\hat{\xi}_1 - v\hat{\xi}_2 \in \mathfrak{h}. \tag{4.26}$$

The existence of such v (not simply $e^{i\gamma}$) depends on whether or not the $w_l(f)$ representation is irreducible.

According to (4.26), whenever $\hat{\xi}_1 = v\hat{\xi}_2$ where the unitary operator v fulfills (4.25), the two representations $W_{l1}(f)$ and $W_{l2}(f)$ are unitarily equivalent. This is also evident from the equality of their expectation functionals, which follows from the form of (4.20) for $\hat{\varphi}_0 = \hat{\xi}_1 = v\hat{\xi}_2$ and from the commutation property (4.25). It is well known that the expectation functional only determines a cyclic representation up to unitary equivalence and that equal expectation functionals correspond to unitarily equivalent representations.[24]

[24] M. A. Naimark, *Normed Rings*, translated by L. F. Boron (P. Noordhoff Ltd., Groningen, The Netherlands, 1964), p. 242.

Conversely, unequal expectation functionals may or may not correspond to equivalent representations, depending on whether (4.26) is fulfilled or not.

Whenever *no* unitary v exists such that (4.26) holds, the two representations W_{l1} and W_{l2} are necessarily inequivalent. In particular, for the cases illustrated above where $\hat{\xi}_1 = 0$ and $\hat{\xi}_2 = \hat{\xi} = \hat{\varphi}_0$, unitary inequivalence arises whenever $\hat{\varphi}_0 \notin \mathfrak{h}$. Moreover, any time $\hat{\xi} \notin \mathfrak{h}$, the decomposition of $U[f_l]$ given in (4.10) does not hold.

To examine the reducibility of a given representation $W_l(f)$, we seek an operator different from the unit operator, which commutes with the generators $W_l(f)$. It is not possible to use $N = (A, A)$, for in a non-Fock representation N is not an operator.[25] What is needed is another operator y for which $[w_l(f), y] = 0$ and for which

$$Y = (A, yA) \tag{4.27}$$

is a *bona fide* operator. In that case,

$$[W_l(f), Y] = (A, [w_l(f), y]A) = 0, \tag{4.28}$$

and reducibility would be established. Such a y can exist only if the representation w_l on $\mathfrak{h}$ is itself reducible. Inspection of the Weyl operators shows that operators of the form (4.27) generate the subalgebra of operators which commute with the $W_l(f)$. Hence, if no such y exists, then the representation of $W_l(f)$ is *irreducible*, in marked contrast to the situation when A is a Fock representation.

To illustrate some of these aspects, let us discuss a particularly simple example of a current algebra, that appropriate to the "affine fields." The field algebra we wish to examine is given formally by the commutation relation

$$[\kappa(\mathbf{x}), \pi(\mathbf{y})] = i\delta(\mathbf{x} - \mathbf{y})\pi(\mathbf{y}), \tag{4.29}$$

and we take the unitary operators as

$$U[f, g] = e^{-i\pi(g)}e^{i\kappa(f)}. \tag{4.30}$$

These are also canonical coordinates (of the so-called second kind[26]) and are chosen for reasons of convenience. The group combination law reads

$$U[f, g]U[f', g'] = U[f + f', g + e^{-f}g'], \tag{4.31}$$

where multiplication is pointwise.

This algebra is the local field analog of that based on the two-parameter Lie algebra,[27] whose commutator reads

$$[B, P] = iP. \tag{4.32}$$

If we imagine that $B = \frac{1}{2}(QP + PQ)$ and heuristically regard Q and P as Heisenberg operators, then a reasonable feel for this algebra is obtained. However, unlike the Heisenberg algebra, there are two unitarily inequivalent, irreducible representations of (4.33), one for which $P > 0$ and the other for which $P < 0$.[27] The unitary operators of the affine group,

$$u_0[r, s] = e^{-isP}e^{irB}, \tag{4.33}$$

are the analogs of $U[f, g]$. If we diagonalize P, then we have the representation

$$(u_0[r, s]\varphi)(k) = e^{-\frac{1}{2}r}e^{-isk}\varphi(e^{-r}k), \tag{4.34}$$

where $\varphi(k) \in L^2(R) = L^2(-\infty, \infty)$. Note that $L^2(0, \infty)$ and $L^2(-\infty, 0)$ form invariant subspaces; in the former $P > 0$, while in the latter $P < 0$. The irreducible representations are given by restricting attention to $L^2(0, \infty)$ [or $L^2(-\infty, 0)$] as representation spaces. For our purposes, we choose $L^2(0, \infty)$ and, thereby, obtain an irreducible representation space for u_0.

As our example of a solution of the affine field algebra, we adopt an exponential construction and choose

$$\varphi'[f, g] = \int^{\oplus} \{u_0[f(\mathbf{z}), g(\mathbf{z})] - 1\}\tilde{\varphi}_0 \, d\mathbf{z}. \tag{4.35}$$

Here we have taken

$$\mathfrak{h} = \int^{\oplus} \mathfrak{h}_z \, d\mathbf{z} \tag{4.36}$$

and may identify $\mathfrak{h}_z = L^2(0, \infty)$. Thus, $\tilde{\varphi}_0$ is (presently!) an element of $L^2(0, \infty)$, and is chosen the same for all $\mathbf{z}$. Note that

$$\hat{\xi} = \int^{\oplus} \tilde{\varphi}_0 \, d\mathbf{z}, \tag{4.37}$$

and, therefore,

$$\|\hat{\xi}\|^2 = \int \|\tilde{\varphi}_0\|_z^2 \, d\mathbf{z} = \|\tilde{\varphi}_0\|_0^2 \int d\mathbf{z}. \tag{4.38}$$

In the infinite-configuration space which we consider, $\|\hat{\xi}\| = \infty$ and corresponds to an improper translation. Moreover, $\tilde{\varphi}_0$ vectors with unequal norms manifestly lead to unitarily inequivalent representations of $U[f, g]$. As we shall see, thanks to the irreducibility assumed for u_0, each class of unitarily equivalent representations is labeled by the rays $[\tilde{\varphi}_0]$, i.e., by

$$[\tilde{\varphi}_0] \equiv \{e^{i\gamma}\tilde{\varphi}_0 : 0 \leq \gamma < 2\pi\}. \tag{4.39}$$

[25] Ref. 17, p. 141; L. Gårding and A. S. Wightman, Proc. Nat. Acad. Sci. U.S. **40**, 622 (1954); J. M. Chaiken, Ann. Phys. (N.Y.) **42**, 23 (1967).

[26] P. M. Cohn, *Lie Groups* (Cambridge University Press, Cambridge, England, 1961), p. 110.

[27] I. M. Gel'fand and M. A. Naimark, Dokl. Akad. Nauk SSSR **55**, 570 (1947); E. W. Aslaksen and J. R. Klauder, J. Math. Phys. **9**, 206 (1968).

The representation induced by (4.35) has an expectation functional given by

$$E_A(f, g) \equiv (\Phi_0, U[f, g]\Phi_0)$$
$$= \exp\left[\int (\tilde{\varphi}_0, \{u_0[f(\mathbf{z}), g(\mathbf{z})] - 1\}\tilde{\varphi}_0)_z \, d\mathbf{z}\right]$$
$$= \exp\left(\iint_0^\infty \{\tilde{\varphi}_0^*(k)[e^{-\frac{1}{2}f(\mathbf{z})-ikg(\mathbf{z})}\tilde{\varphi}_0(e^{-f(\mathbf{z})}k) - \tilde{\varphi}_0(k)]\} \, dk \, d\mathbf{z}\right). \quad (4.40)$$

Observe that the vector Φ_0 is translation invariant and that the representation fulfills cluster decomposition. Indeed, this example is ultralocal in the sense of (3.79) and (3.80), although not every solution of (4.31) need be ultralocal by any means. Note also that if $f(\mathbf{z}) \equiv 0$, $U[0, g]U[0, g'] = U[0, g + g']$ and we are necessarily led to an exponential representation for the single commuting field operator $\pi(\mathbf{x})$. Clearly, the resultant representation (4.40) that arises when $f = 0$ has the general form derived in Sec. 3. This property is also fulfilled if we instead set $g = 0$, although the form of (4.40) does not make that result immediately obvious. Clearly, this is a general feature of exponential group representations in canonical coordinates.

The presence or absence of a linear term in the exponent of the expectation functional [such as appears in (3.66)] is dictated by requirements of the group representation. For example, in the case of the affine field we could legitimately adopt

$$\varphi'[f, g] = \int^\oplus (bf(\mathbf{z}) \oplus \{u_0[f(\mathbf{z}), g(\mathbf{z})] - 1\}\tilde{\varphi}_0) \, d\mathbf{z}, \quad (4.41)$$

where, say, b is a real constant. The modified expectation functional is

$$E'_A(f, g) = \exp\left(-\tfrac{1}{2}b^2 \int f^2(\mathbf{z}) \, d\mathbf{z}\right) E_A(f, g), \quad (4.42)$$

where $E_A(f, g)$ is given by (4.40). Again, this representation fulfills cluster decomposition and has a translationally invariant Φ_0. In addition, distinct b values lead to inequivalent representations. However, there is an unfaithful subrepresentation of the field operator $\pi(\mathbf{x})$, which may or may not be important in applications.

To see this feature most simply, let us return to our single degree of freedom case where $[B, P] = iP$. One set of solutions is to assume $P = 0$! In that case

$$u_0[r, s] = e^{-isP}e^{irB} = e^{irB}. \quad (4.43)$$

In the subspace where this occurs, B can be diagonalized and each distinct eigenvalue b corresponds to a unitarily inequivalent representation of u_0. Each such representation is one-dimensional. For our purposes, we choose not to consider such representations either in the one-dimensional case or in the field analog in which Eq. (4.29) is satisfied by assuming a subspace where $\pi(\mathbf{y}) \equiv 0$ while $\kappa(\mathbf{x}) \neq 0$. Thus, we ignore the possibility of a linear term in $\varphi'[f, g]$. In other group representations it may be appropriate to include such terms. When we discuss the canonical commutation relations in the next section, we shall see, in a way, that linear terms are required.

Recall that the annihilation operator $A'(\lambda)$ is defined by

$$A'(\lambda)\Phi'[f, g] = (\lambda, \varphi'[f, g])\Phi'[f, g] \quad (4.44)$$

and satisfies

$$[A'(\lambda_1), A'(\lambda_2)^\dagger] = (\lambda_1, \lambda_2). \quad (4.45)$$

Since in this example $\mathfrak{h} = \int^\oplus \mathfrak{h}_z \, d\mathbf{z}$, it is convenient to introduce the operators $\mathcal{A}'(\lambda_z)$ and their formal (and nonexistent!) adjoints $\mathcal{A}'(\lambda_z)^\dagger$ defined as follows: For each $\lambda_z \in \mathfrak{h}_z$, we set

$$\mathcal{A}'(\lambda_z)\Phi'[f, g] = (\lambda_z, \varphi'_z[f, g])_z\Phi'[f, g] \quad (4.46)$$

and

$$[\mathcal{A}'(\lambda_{1z}), \mathcal{A}'(\lambda_{2z'})^\dagger] = \delta(\mathbf{z} - \mathbf{z}')(\lambda_{1z}, \lambda_{2z})_z. \quad (4.47)$$

In the manner of (2.32) and (2.33a), we let

$$(\mathcal{A}', w\mathcal{A}')_z \equiv \sum_{n,m} \mathcal{A}'(\lambda_{nz})^\dagger(\lambda_{nz}, w\lambda_{mz})_z\mathcal{A}'(\lambda_{mz}), \quad (4.48)$$

where $\{\lambda_{nz}\}$ is a complete orthonormal set in $\mathfrak{h}_z$. This form is not an operator but becomes one on integration, rather like the number-density operator in usual quantum field theory.

Along with $\mathcal{A}'$ let us introduce

$$\mathcal{A}(\lambda_z) = \mathcal{A}'(\lambda_z) + (\lambda_z, \hat{\xi}_z)_z, \quad (4.49a)$$

which in our example becomes

$$\mathcal{A}(\lambda_z) = \mathcal{A}'(\lambda_z) + (\lambda_z, \tilde{\varphi}_0)_z. \quad (4.49b)$$

Since A' is the Fock representation, so too, we may say, is $\mathcal{A}'$. Although A is not the Fock representation because $\|\hat{\xi}\| = \infty$, we may say that A is "locally Fock" whenever $\mathcal{A}(\lambda_z)$ is unitarily equivalent to $\mathcal{A}'(\lambda_z)$. This occurs in our case, provided that $\tilde{\varphi}_0 \in \mathfrak{h}_z$. In other words, the representations would be equivalent if we dealt with a finite-configuration-space volume (quantization in a box). However, we can also take $\tilde{\varphi}_0 \notin \mathfrak{h}_z$, as we shall see, in which case the representation A is not even locally Fock.

In essentially the same manner as (4.13b), we may realize the group generators for ultralocal representations as

$$W_i(f) = \int f(\mathbf{z})(\mathcal{A}, \omega_i\mathcal{A})_z \, d\mathbf{z}. \quad (4.50)$$

Note that on $\mathfrak{h}_z$ the operators ω_l simply satisfy

$$[\omega_l, \omega_m] = ic_{lmn}\omega_n, \tag{4.51}$$

rather like the "algebra of charges" associated with the "algebra of currents." If the representation A is locally Fock, then, for $g(\mathbf{z}) \in \mathfrak{D}$,

$$N(g) \equiv \int g(\mathbf{z})(\mathcal{A}, \mathcal{A})_z \, d\mathbf{z} \tag{4.52}$$

is a meaningful operator, although

$$N = N(1) = \int (\mathcal{A}, \mathcal{A})_z \, d\mathbf{z} = (A, A) \tag{4.53}$$

is not defined. However, it is clear that

$$\begin{aligned}[W_l(f), N(g)] &= \iint f(\mathbf{z})g(\mathbf{x}) \\ &\quad \times [(\mathcal{A}, \omega_l\mathcal{A})_z, (\mathcal{A}, \mathcal{A})_x] \, d\mathbf{z} \, d\mathbf{x} \\ &= \int f(\mathbf{z})g(\mathbf{z})(\mathcal{A}, [\omega_l, I]\mathcal{A})_z \, d\mathbf{z} \\ &= 0. \end{aligned} \tag{4.54}$$

Thus, the operator $N(g)$ plays the role of Y in Eq. (4.27). Since $N(g)$, for any $g(\mathbf{z}) \in \mathfrak{D}$, is not a multiple of the identity, we have established the reducibility of ultralocal exponential representations of $U[f_l]$ which are based on locally Fock representations of the operators A and $A^\dagger$.

Finally, we consider those representations for which $\tilde{\varphi}_0 \notin \mathfrak{h}_z$. To help visualize this situation, we appeal to our example of the affine fields. [Not all groups admit a generalization to $\tilde{\varphi}_0 \notin \mathfrak{h}_z$; in particular, the "algebra of charges" (4.51) cannot correspond to a compact group.] On reference to (4.40), the quantity which must be well defined is obviously

$$\int_0^\infty \tilde{\varphi}_0^*(k)[e^{-\frac{1}{2}r - isk}\tilde{\varphi}_0(e^{-r}k) - \tilde{\varphi}_0(k)] \, dk. \tag{4.55}$$

Since this must be meaningful for all r and s, it is clear that $\tilde{\varphi}_0$ can fail to be L^2 only near $k = 0$. For example, let us assume that $\tilde{\varphi}_0(k)$ is square integrable for $K \leq k < \infty$, for some $K > 0$. To complete the specification of $\tilde{\varphi}_0(k)$, we assume that

$$\tilde{\varphi}_0(k) \equiv k^{-\frac{1}{2}}e(k), \tag{4.56}$$

where $e(k)$ fulfills the Lipschitz condition of order δ,

$$|e(k') - e(k)| \leq C\,|k' - k|^\delta, \tag{4.57}$$

for $0 \leq k, k' \leq K$, and for some C and $\delta > 0$. From this condition it readily follows that (4.55) is well defined for all r and s. Two examples of suitable $\tilde{\varphi}_0(k) \notin L^2(0, \infty)$ are given by $(k + k^4)^{-\frac{1}{2}}$ and $k^{-\frac{1}{2}}e^{-k}$.

With such a choice for $\tilde{\varphi}_0$, the representation A is not even locally Fock, so that although $W_l(f)$ is still given by (4.50), the reducing operator $N(g)$ is no longer an operator. In fact, there is no operator different from unity which commutes with all the $W_l(f)$. As a consequence, the ultralocal exponential representations based on nonlocally Fock representations of A and $A^\dagger$ are irreducible representations of $U[f_l]$. In symbols, if $\tilde{\varphi}_0 \notin \mathfrak{h}_z$, the representation $U[f_l]$ is irreducible. In our example, since ω_l $(= B, P)$ form an irreducible representation in $\mathfrak{h}_z$, only those representations for which $\tilde{\varphi}_{01} - e^{i\gamma}\tilde{\varphi}_{02} \in L^2(0, \infty)$, for some real γ, are locally equivalent. Full equivalence of two representations of the form (4.40) requires not only local equivalence but the condition that $\hat{\varphi}_{01} - e^{i\gamma}\hat{\varphi}_{02} \in \mathfrak{h}$, where $\hat{\varphi}_{0l} = \int^\oplus \tilde{\varphi}_{0l} \, d\mathbf{z}$, $l = 1, 2$. The infinite volume of configuration space requires that $\tilde{\varphi}_{01} = e^{i\gamma}\tilde{\varphi}_{02}$ for equivalence, which establishes that the rays (4.39) label inequivalent representations of (4.40). The validity of this result may also be seen by the "tag test." [28] In this approach one first shows for (4.40) that

$$\text{w-lim}_{|a| \to \infty} U[f_a, g_a] = E_A(f, g). \tag{4.58}$$

That is, these operators have a weak limit which is a multiple of the identity, and that multiple—the "tag"—is simply $E_A(f, g)$. As is easily seen, distinct tags label inequivalent representations. Unless $\tilde{\varphi}_{01} = e^{i\gamma}\tilde{\varphi}_{02}$, the tags are evidently unequal for some choice of $f, g \in \mathfrak{D}$. Clearly, this kind of technique applies directly to exponential representations fulfilling translation invariance and cluster decomposition for general current algebras.

Let us briefly discuss some spectral properties of the field operator $\pi(\mathbf{x})$ in the locally and nonlocally Fock-representation cases. We define[29]

$$P_\Delta \equiv \int_\Delta \pi(\mathbf{x}) \, d\mathbf{x}, \tag{4.59a}$$

$$B_\Delta \equiv \int_\Delta \kappa(\mathbf{x}) \, d\mathbf{x} \tag{4.59b}$$

for some compact set $\Delta \subset R^3$. It follows from (4.29) that

$$[B_\Delta, P_\Delta] = iP_\Delta, \tag{4.60}$$

which is just the commutation relation (4.32). The representation of P_Δ is determined by our expectation functional; it will be highly reducible in $\mathfrak{H}$, but we may

[28] J. R. Klauder and J. McKenna, J. Math. Phys. 6, 68 (1965), Sec. 4.C.

[29] Although a characteristic function is not an element of the test function space $\mathfrak{D}$, it is a limit of such functions. Its appropriateness as a smearing function can be determined directly from the expectation functional for ultralocal representations.

well ask the question whether P_Δ only has an absolutely continuous spectrum, $P_\Delta > 0$ in our case, or whether there is also a subspace in which $P_\Delta = 0$. We may examine this question most simply in the context of

$$(\Phi_0, e^{-isP}\Phi_0) = \exp\left[\int_\Delta \int_0^\infty (e^{-iks} - 1)\,|\tilde{\varphi}_0(k)|^2\,dk\,d\mathbf{z}\right] = \exp\left[-\Delta\int_0^\infty (1 - e^{-iks})\,|\tilde{\varphi}_0(k)|^2\,dk\right]. \quad (4.61)$$

If $\tilde{\varphi}_0(k) \in L^2(0, \infty)$, then the Riemann–Lebesque lemma insures that

$$\lim_{s\to\infty} (\Phi_0, e^{-isP}\Phi_0) = \exp(-\Delta\,\|\tilde{\varphi}_0\|^2) \neq 0, \quad (4.62a)$$

and, consequently, there is necessarily a subspace where $P_\Delta = 0$. The bilinear form of W_l assures us that in this subspace $B_\Delta = 0$ as well. Thus, we are confronted with a representation for B_Δ and P_Δ, a portion of which is unfaithful, and in which every group element is represented by unity.[30] However, if $\tilde{\varphi}_0(k) \notin L^2(0, \infty)$, e.g., in the manner described previously, then clearly

$$\lim_{s\to\infty} (\Phi_0, e^{-isP}\Phi_0) = 0. \quad (4.62b)$$

In fact, it can be easily shown that

$$\lim_{s\to\infty} (\Phi[f_1, g_1], e^{-isP}\Phi[f_2, g_2]) = 0, \quad (4.63)$$

from which it follows that

$$\text{w-lim}_{s\to\infty}\, e^{-isP_\Delta} = 0. \quad (4.64)$$

This is just the condition that P_Δ only has an absolutely continuous spectrum. Depending on the proposed application of the representation, the spectral properties of a given smeared field (P_Δ, say) may determine that certain field representations are appropriate. Elsewhere, in an application of the local affine fields to physically motivated model problems, we shall discuss this question further.[31]

The non-Abelian nature of the combination law (4.31) imposes restrictions on the allowed representations. To demonstrate that there are representations, other than the ultralocal ones, exhibiting translation invariance and cluster decomposition, it suffices to give a class of different examples. Let $0 \leq k_q < \infty$, $q = 1, \cdots, Q$, denote Q integration variables; $f(\mathbf{z}_q) \equiv f(\mathbf{a}_q + \mathbf{z})$, $g(\mathbf{z}_q) \equiv g(\mathbf{a}_q + \mathbf{z})$, where $\mathbf{a}_q$ denote Q fixed, distinct points in R^3 and where $\tilde{\varphi}_0(k_q) = \tilde{\varphi}_0(k_1, \cdots, k_Q)$ is an element of L^2 or possibly more general, roughly in the manner of (4.57). Then, it is essentially clear that a representation of (4.31) exists such that

$$E_A(f, g) = \exp\left(\int d\mathbf{z} \int_0^\infty \cdots \int_0^\infty \tilde{\varphi}_0^*(k_q) \times \{\exp[-\textstyle\sum \tfrac{1}{2} f(\mathbf{z}_q) - i\sum k_q g(\mathbf{z}_q)] \times \tilde{\varphi}_0(e^{-f(\mathbf{z}_q)}k_q) - \tilde{\varphi}_0(k_q)\}\,dk^Q\right), \quad (4.65)$$

which is inequivalent to all those discussed previously.

5. EXPONENTIAL REPRESENTATIONS OF CANONICAL FIELD OPERATORS

In this section we consider exponential representations of scalar field and momentum operators which fulfill the equal-time-commutation relation

$$[\varphi(\mathbf{x}), \pi(\mathbf{y})] = i\delta(\mathbf{x} - \mathbf{y}). \quad (5.1)$$

Unlike (4.1), the canonical operators do not form a closed algebra, so that our analysis will necessarily be somewhat different than in Sec. 4. If $f(\mathbf{x})$ and $g(\mathbf{x})$ denote suitable real test functions, such as elements of $\mathfrak{D}$, then

$$[\varphi(f), \pi(g)] = i(f, g) \quad (5.2)$$

is the proper statement of (5.1) for self-adjoint operators on a suitable domain. We shall be particularly interested in the unitary Weyl operators

$$U[f, g] = e^{i[\varphi(f) - \pi(g)]}, \quad (5.3)$$

where we have chosen our sign convention for convenience. These operators obey the relation

$$U[f', g']U[f, g] = e^{(\frac{1}{2}i)[(f', g) - (g', f)]}U[f' + f, g' + g], \quad (5.4)$$

which is the Weyl form of the commutation relations.

We introduce the overcomplete family of states

$$\Phi'[f, g] \equiv e^{-i\,\mathrm{Im}(\hat{\xi}, \varphi'[f,g])}U[f, g]\Phi_0 \quad (5.5)$$

for some choice of $\hat{\xi}$ and $\varphi'[f, g] \in \mathfrak{h}$. We assume that $\mathrm{Im}\,(\hat{\xi}, \varphi'[0, 0]) = 0$. Once again we insist on an exponential construction such that

$$(\Phi'[f', g'], \Phi'[f, g]) = N'Ne^{(\varphi'[f',g'], \varphi'[f,g])} = e^{-\frac{1}{2}\|\varphi'[f',g'] - \varphi'[f,g]\|^2} \times e^{i\,\mathrm{Im}\,(\varphi'[f',g'], \varphi'[f,g])}. \quad (5.6)$$

[30] For example, see E. P. Wigner, *Group Theory*, translated by J. J. Griffin (Academic Press, Inc., New York, 1959), p. 72. We use this term to denote representations having subrepresentations in which one or more generators are represented by the zero operator.

[31] E. W. Aslaksen, thesis, Lehigh University, Bethlehem, Pa., 1968; J. R. Klauder, *5th International Conference on Gravitation and the Theory of Relativity* (Tbilisi University, Tbilisi, USSR, to be published); J. R. Klauder, *Proceedings of the Relativity Conference in the Midwest* (Plenum Press, New York, to be published).

If we employ the combination law (5.4), then we learn [cf. Eq. (4.7)] that

$$\|\varphi'[f', g'] - \varphi'[f, g]\|^2 = \|\varphi'[0, 0] - \varphi'[f - f', g - g']\|^2, \quad (5.7a)$$

$$\begin{aligned}\mathrm{Im}\,(\varphi'[f', g'], \varphi'[f, g]) &= \mathrm{Im}\,(\varphi'[0, 0], \varphi'[f - f', g - g']) \\ &\quad - \mathrm{Im}\,(\hat{\xi}, \varphi'[f - f', g - g'] \\ &\quad - \varphi'[f, g] + \varphi'[f', g']) \\ &\quad - \tfrac{1}{2}[(f', g) - (g', f)]. \end{aligned} \quad (5.7b)$$

The basic solution to these relations gives rise to the Fock representation of Sec. 2A, which motivated our study in the first place. Somewhat more generally, we adopt as our first solution $\mathfrak{h} \equiv \mathfrak{h}_1$, $\hat{\xi} = 0$, and

$$\begin{aligned}\varphi'[f, g] &\equiv \varphi_1[f, g] \\ &= 2^{-\frac{1}{2}}\{[(\Omega^{-\frac{1}{2}}\tilde{f})(\mathbf{k}) - i(\Omega^{\frac{1}{2}}\tilde{g})(\mathbf{k})] \oplus (\Gamma^{\frac{1}{2}}\tilde{f})(\mathbf{k})\}. \end{aligned} \quad (5.8)$$

Here Ω and Γ are suitable linear operators, such as $\Omega = \Omega(k)$, $\Gamma = \Gamma(k)$, i.e., multiplication operators in momentum space such that $\Omega(k) > 0$ almost everywhere. On the other hand, $\Gamma(k) \geq 0$, and, in fact, Γ may be chosen identically zero. While (5.8) gives a convenient specific realization, we shall also treat $\mathfrak{h}_1$ abstractly. It is clear that $\varphi_1[f, g]$ already fulfills (5.7) and thus generates an exponential representation of the canonical commutation relations. The resultant expectation functional is given by

$$\begin{aligned}E_1(f, g) &= (\Phi_0, \Phi'[f, g]) = e^{-\frac{1}{2}\|\varphi'[f,g]\|^2} \\ &= e^{-\frac{1}{4}[(\tilde{f}, \Lambda\tilde{f}) + (\tilde{g}, \Omega\tilde{g})]}, \end{aligned} \quad (5.9a)$$

where

$$\Lambda \equiv \Omega^{-1} + \Gamma. \quad (5.9b)$$

This solution is translationally invariant and satisfies cluster decomposition whenever $\Lambda = \Lambda(k)$ and $\Omega = \Omega(k)$ are polynomially bounded. In this case, distinct function pairs $\Lambda(k)$ and $\Omega(k)$ lead to unitarily inequivalent representations.

The representation characterized by $E_1(f, g)$ is just the direct product of two independent Fock representations; it reduces to only one such representation if $\Gamma \equiv 0$.[32] Whether one or two Fock representations are involved, the field and momentum operators are given by suitable linear combinations of annihilation and creation operators. According to (3.3), we set

$$A_1(\lambda)\Phi[f, g] = (\lambda, \varphi_1[f, g])\Phi[f, g]. \quad (5.10)$$

It can be shown[32] that the smeared momentum operator $\pi_1(e)$ is given by

$$\pi_1(e) = (i/2^{\frac{1}{2}})[A_1(\lambda_{\pi(e)})^\dagger - A_1(\lambda_{\pi(e)})], \quad (5.11a)$$

where

$$\lambda_{\pi(e)} \equiv -i(\Omega^{\frac{1}{2}}\tilde{e})(\mathbf{k}) \oplus 0. \quad (5.11b)$$

In like manner, the smeared field operator $\varphi_1(e)$ is given by

$$\varphi_1(e) = 2^{-\frac{1}{2}}[A_1(\lambda_{\varphi(e)})^\dagger + A_1(\lambda_{\varphi(e)})], \quad (5.12a)$$

where

$$\lambda_{\varphi(e)} = -i(\Omega^{-\frac{1}{2}}\tilde{e})(\mathbf{k}) \oplus -i(\Gamma^{\frac{1}{2}}\tilde{e})(\mathbf{k}). \quad (5.12b)$$

Consequently, we find that

$$[\varphi_1(e), \pi_1(e)] = i(\lambda_{\varphi(e)}, \lambda_{\pi(e)}) = i(e, e), \quad (5.13)$$

as desired. The representation of φ_1 and π_1 determined by the above construction is reducible whenever $\Gamma \not\equiv 0$.

Although the solution presented above is rather special, all the other solutions build on it. For our second solution, we still choose $\hat{\xi} = 0$, but set

$$\mathfrak{h} = \mathfrak{h}_1 \oplus \mathfrak{h}_2, \quad (5.14a)$$

$$\varphi'[f, g] \equiv \varphi[f, g] = \varphi_1[f, g] \oplus \varphi_2[f, g], \quad (5.14b)$$

where $\mathfrak{h}_1$ and $\varphi_1[f, g]$ are as before. Since φ_1 already fulfills Eq. (5.7), it follows that $\varphi_2[f, g]$ satisfies those same relations, with the last term in (5.7) absent. Consequently, we adopt as our solution

$$\varphi_2[f, g] = \tilde{u}_0[f, g]\varphi_0, \quad (5.15a)$$

where the unitary operators $\tilde{u}_0$ fulfill

$$\tilde{u}_0[f', g']\tilde{u}_0[f, g] = \tilde{u}_0[f' + f, g' + g], \quad (5.15b)$$

a strictly Abelian combination law. Accordingly, we may set

$$\tilde{u}_0[f, g] = e^{i[\tilde{\varphi}(f) - \tilde{\pi}(g)]}, \quad (5.15c)$$

where $\tilde{\varphi}(f)$ and $\tilde{\pi}(g)$ are *fully commuting* self-adjoint generators. Moreover, these operators act on $\mathfrak{h}_2$.

In terms of these expressions the resulting expectation functional has the form

$$\begin{aligned}E(f, g) &= E_1(f, g)e^{(\varphi_0, \{\tilde{u}_0[f,g]-1\}\varphi_0)} \\ &= E_1(f, g)\exp(\varphi_0, \{e^{i[\tilde{\varphi}(f) - \tilde{\pi}(g)]} - 1\}\varphi_0), \end{aligned} \quad (5.16)$$

where E_1 is still given by (5.9). So long as $\varphi_0 \in \mathfrak{h}_2$, as is presently the case, none of these combined solutions can exhibit translational invariance and cluster decomposition. This defect will be remedied later when we treat improper translations.

The Weyl operators for this solution may be given as follows: Let $U_1[f, g]$ denote the Weyl operators

[32] H. Araki and E. J. Woods, J. Math. Phys. **4**, 637 (1965); J. R. Klauder and L. Streit, J. Math. Phys. **10**, 1661 (1969).

characterized by Eq. (5.9), and define

$$\tilde{U}_0[f, g] = \bigoplus_{n=0}^{\infty} \bigotimes_{s=1}^{n} \tilde{u}_0[f, g]. \tag{5.17}$$

Then the solution characterized by Eq. (5.16) is given by

$$U[f, g] = U_1[f, g] \otimes \tilde{U}_0[f, g]. \tag{5.18}$$

If we extend the operators U_1 and $\tilde{U}_0$ in obvious fashion to the direct product space, then we may drop the direct product and observe that

$$U[f, g] = U_1[f, g]\tilde{U}_0[f, g] \equiv e^{i[\varphi(f)-\pi(g)]}. \tag{5.19}$$

Since $\tilde{U}_0[f, g]$ obeys an Abelian combination law, in virtue of (5.15b), it is clear that $U[f, g]$ will satisfy the Weyl form of the commutation relations (5.4) and that the generators $\varphi(f)$ and $\pi(g)$ will obey the usual Heisenberg rules.

In the present solution, the annihilation operators $A(\lambda)$, $\lambda \in \mathfrak{h}_1 \oplus \mathfrak{h}_2$, fulfill the relation

$$A(\lambda)\Phi[f, g] = (\lambda, \varphi[f, g])\Phi[f, g]. \tag{5.20}$$

If we set $\lambda = \lambda_1 \oplus \lambda_2$, $\lambda_j \in \mathfrak{h}_j$, then we define

$$\begin{aligned} A(\lambda) = A(\lambda_1 \oplus \lambda_2) &= A(\lambda_1 \oplus 0) + A(0 \oplus \lambda_2) \\ &\equiv A_1(\lambda) + A_2(\lambda), \end{aligned} \tag{5.21}$$

which may be seen to be in accord with (5.10). Note that A_1 commutes with A_2 and $A_2^\dagger$. Equation (5.20) leads to the relations

$$A_j(\lambda)\Phi[f, g] = (\lambda_j, \varphi_j[f, g])\Phi[f, g], \tag{5.22}$$

where we have put $\varphi_1 = \varphi_1 \oplus 0$ and $\varphi_2 = 0 \oplus \varphi_2$. The basic field and momentum operators in this representation are given as a combination of two different terms. On comparison with (4.13b), we determine that

$$\varphi(f) = \varphi_1(f) + (A_2, \tilde{\varphi}(f)A_2), \tag{5.23a}$$

$$\pi(g) = \pi_1(g) + (A_2, \tilde{\pi}(g)A_2), \tag{5.23b}$$

where φ_1 and π_1 are given by Eqs. (5.11a) and (5.12a). Note that the field operators involve one term which is *linear* in A and $A^\dagger$ and another term which is *quadratic*. This is a characteristic feature of such representations. Since

$$\begin{aligned} [(A_2, \tilde{\varphi}(f)A_2), (A_2, \tilde{\pi}(g)A_2)] &= (A_2, [\tilde{\varphi}(f), \tilde{\pi}(g)]A_2) \\ &= 0, \end{aligned} \tag{5.24}$$

it follows that

$$[\varphi(f), \pi(g)] = [\varphi_1(f), \pi_1(g)] = i(f, g), \tag{5.25}$$

as desired. The reducibility of this representation is evident, assuming that $\tilde{\varphi}$ and $\tilde{\pi}$ are not both identically zero.

Before we pass to another solution of the basic relations (5.7), we note the fact that since $\tilde{\varphi}$ and $\tilde{\pi}$ commute, they may both be simultaneously diagonalized. Alternatively stated, we may write

$$\mathfrak{h}_2 = \int^{\oplus} C \, d\mu(a, b), \tag{5.26a}$$

$$e^{i[\tilde{\varphi}(f)-\tilde{\pi}(g)]} = \int^{\oplus} e^{i[(a,f)-(b,g)]} \, d\mu(a, b), \tag{5.26b}$$

such that

$$\varphi_0 = \int^{\oplus} d\mu(a, b). \tag{5.26c}$$

Here $a = a(\mathbf{x})$ and $b = b(\mathbf{x})$ are elements of $\mathfrak{D}'$ and the integration is over the space $\mathfrak{D}' \times \mathfrak{D}'$. Note that

$$\|\varphi_0\|^2 = \int d\mu(a, b) = \mu(\mathfrak{D}' \times \mathfrak{D}'), \tag{5.27}$$

which must be finite in the present case. In this language, the expectation functional (5.16), for example, takes the form

$$E(f, g) = E_1(f, g) \exp\left[\int (e^{i[(a,f)-(b,g)]} - 1) \, d\mu(a, b)\right]. \tag{5.28}$$

Evidently, other expressions may be "diagonalized" in $\mathfrak{h}_2$ and given a concrete representation in similar fashion. As a specific example, it may be assumed that μ is concentrated on δ functions at a specific point $\mathbf{z}$, just as was assumed for Eq. (3.63).

We turn now to another set of solutions of the basic relations (5.7) in which the translation vector $\hat{\xi} \neq 0$. In essence, we will retain the second form of our solution, in which $\mathfrak{h} = \mathfrak{h}_1 \oplus \mathfrak{h}_2$, and consider $\hat{\xi} = \hat{\xi}_1 \oplus \hat{\xi}_2$. The role of $\hat{\xi}_1$ is especially simple. From the form of $\varphi_1[f, g]$ in (5.8), it is clear that $\hat{\xi}_1$ does not influence Eq. (5.7) in any way. The only appearance $\hat{\xi}_1$ makes is in the phase factor which accompanies (5.5). We can relate any solution with $\hat{\xi}_1 \neq 0$ to the corresponding solution with $\hat{\xi}_1 = 0$ by appending the phase factor $e^{-i \operatorname{Im}(\hat{\xi}_1, \varphi_1[f,g])}$, where $\varphi_1[f, g]$ is given in (5.8), to the expectation functional for the case when $\hat{\xi}_1 = 0$. The effect of $\hat{\xi}_1$, stated otherwise, is to add to φ_1 and π_1 constant multiples of the identity. For convenience, we shall confine our attention to the consequences of $\hat{\xi}_2$, and we set $\hat{\xi}_1 = 0$.

With these remarks in mind we choose as our solution to (5.7) the relations

$$\hat{\xi} = 0 \oplus \hat{\xi}_2 \equiv 0 \oplus \hat{\varphi}_0, \tag{5.29a}$$

$$\varphi'[f, g] \equiv \varphi_1[f, g] \oplus \varphi_2'[f, g], \tag{5.29b}$$

where

$$\varphi_2'[f, g] = (\tilde{u}_0[f, g] - 1)\hat{\varphi}_0. \tag{5.30}$$

Here $\varphi_1[f, g] \in \mathfrak{h}_1$ has the same form as before, and likewise for the Abelian group $\tilde{u}_0[f, g]$. The expectation functional to which this solution leads,

$$E(f, g) = E_1(f, g) \exp(\hat{\varphi}_0, \{\tilde{u}_0[f, g] - 1\}\hat{\varphi}_0)$$
$$= E_1(f, g) \exp(\hat{\varphi}_0, \{e^{i[\tilde{\varphi}(f) - \tilde{\pi}(g)]} - 1\}\hat{\varphi}_0), \tag{5.31}$$

is similar to the one in (5.16). A real difference arises only when $\|\hat{\varphi}_0\| = \infty$, corresponding to an improper translation. Since this is the most interesting case, we shall concentrate on improper translations.

Although $\hat{\varphi}_0 \notin \mathfrak{h}_2$, we assume that $\varphi_2'[f, g] \in \mathfrak{h}_2$ and that $(\hat{\varphi}_0, \varphi_2'[f, g])$ makes sense. Moreover, for convenience, let us assume that $\tilde{\varphi}(f)\hat{\varphi}_0$ and $\tilde{\pi}(g)\hat{\varphi}_0$ are in $\mathfrak{h}_2$, and that $(\hat{\varphi}_0, \tilde{\varphi}(f)\hat{\varphi}_0)$ and $(\hat{\varphi}_0, \tilde{\pi}(g)\hat{\varphi}_0)$ make sense. With these assumptions, which are like those made in conjunction with Eq. (4.23), we can determine expressions for the field operators $\varphi(f)$ and $\pi(g)$.

We define the operators $A'(\lambda)$, $\lambda \in \mathfrak{h}_1 \oplus \mathfrak{h}_2$, according to the relation

$$A'(\lambda)\Phi'[f, g] = (\lambda, \varphi'[f, g])\Phi'[f, g]. \tag{5.32}$$

If we set $\lambda = \lambda_1 \oplus \lambda_2$, then we define

$$A'(\lambda) = A'(\lambda_1 \oplus \lambda_2) = A'(\lambda_1 \oplus 0) + A'(0 \oplus \lambda_2)$$
$$\equiv A_1(\lambda) + A_2'(\lambda). \tag{5.33}$$

Since

$$A_2'(\lambda)\Phi'[f, g] = (\lambda_2, \varphi_2'[f, g])\Phi'[f, g]$$
$$= (\lambda_2, \{\tilde{u}_0[f, g] - 1\}\hat{\varphi}_0)\Phi'[f, g], \tag{5.34}$$

it is clear that

$$A_2'(\lambda) = A_2(\lambda) - (\lambda, \hat{\varphi}_0), \tag{5.35}$$

where here $\hat{\varphi}_0 = 0 \oplus \hat{\varphi}_0$ is understood. Since $\|\hat{\varphi}_0\| = \infty$, it follows that A_2 and A_2' are unitarily inequivalent and, in particular, that A_2 is (now) inequivalent to the Fock representation A_2'.

The basic field and momentum operators of this solution have the *same form* as in (5.23), namely,

$$\varphi(f) = \varphi_1(f) + (A_2, \tilde{\varphi}(f)A_2), \tag{5.36a}$$

$$\pi(g) = \pi_1(g) + (A_2, \tilde{\pi}(g)A_2), \tag{5.36b}$$

but are inequivalent to those previous solutions. The latter terms in each expression may be interpreted in the manner of (4.23). General criteria for equivalence of two such representations follow the discussion pertaining to Eq. (4.24) (in addition to taking into account the equivalence of the first term). In particular (assuming Ω and Γ remain fixed), two such representations are equivalent, provided that there exists a unitary v which commutes with $\tilde{\varphi}$ and $\tilde{\pi}$ such that

$$\hat{\varphi}_{01} - v\hat{\varphi}_{02} \in \mathfrak{h}_2. \tag{5.37}$$

To demonstrate the reducibility of these representations, we need only exhibit one operator, different from unity, which commutes with φ and π. Although (A_2, A_2) fails to be an operator when A_2 is not equivalent to the Fock representation, we have, by assumption, the fact that

$$Y \equiv (A_2, \tilde{\varphi}(e)A_2), \tag{5.38}$$

for some fixed but arbitrary test function e, is a meaningful operator. Since $\tilde{\varphi}$ and $\tilde{\pi}$ commute, it is clear that

$$[Y, \varphi(f)] = 0 = [Y, \pi(g)], \tag{5.39}$$

which demonstrates reducibility.

The commutativity of $\tilde{\varphi}$ and $\tilde{\pi}$ permits us again to diagonalize them both simultaneously. In particular, we let

$$\mathfrak{h}_2 = \int^{\oplus} C\, d\hat{\mu}(a, b), \tag{5.40a}$$

$$\hat{\varphi}_0 = \int^{\oplus} d\hat{\mu}(a, b), \tag{5.40b}$$

$$\varphi_2'[f, g] = \int^{\oplus} (e^{i[(a,f)-(b,g)]} - 1)\, d\hat{\mu}(a, b). \tag{5.40c}$$

As before, the integration is over $\mathfrak{D}' \times \mathfrak{D}'$, but now we assume that

$$\|\hat{\varphi}_0\|^2 = \int d\hat{\mu}(a, b) = \hat{\mu}(\mathfrak{D}' \times \mathfrak{D}') = \infty. \tag{5.41}$$

With this expression, it follows that

$$E(f, g) = E_1(f, g) \exp\left[\int (e^{i[(a,f)-(b,g)]} - 1)\, d\hat{\mu}(a, b)\right]. \tag{5.42}$$

Among the many possible ways in which improper translations can be used, we shall consider only two basic examples. Initially, let us create a translationally invariant expectation functional exhibiting cluster decomposition. This we may do, following the lead of Eq. (3.73), by assuming that

$$L(f, g) \equiv \int (e^{i[(a,f)-(b,g)]} - 1)\, d\hat{\mu}(a, b)$$
$$= \int (e^{i[(a,f_x)-(b,g_x)]} - 1)\, d\hat{\mu}_0(a, b)\, d\mathbf{x}, \tag{5.43}$$

where $\hat{\mu}_0$ is concentrated on distributions "at or near" the origin. As a further specialization, we may let $\hat{\mu}_0$ be concentrated on δ functions at the origin so that

$$L(f, g) = \int (e^{i[\lambda f(\mathbf{x}) - \nu g(\mathbf{x})]} - 1)\, d\hat{\sigma}_0(\lambda, \nu)\, d\mathbf{x}. \tag{5.44}$$

Such a specialization would be an ingredient in leading to an ultralocal representation [in the sense of (3.78)–(3.80)]. Note that it is not necessary that $\hat{\sigma}_0$ have a finite measure. For example, a possible choice for $\hat{\sigma}_0$

would be

$$d\hat{\sigma}_0(\lambda, \nu) = (\lambda^2 + \nu^2)^{-1} e^{-(\lambda^2+\nu^2)}\, d\lambda\, d\nu. \quad (5.45)$$

We have already noted that when $\|\hat{\varphi}_0\| = \infty$, the representation of φ and π is inequivalent to that based on a Fock representation for A_2. In light of the discussion in Sec. 4, it should be clear that in the example discussed above, if $\hat{\sigma}_0$ has infinite measure, the representation of φ and π is not even "locally Fock." The issues of unfaithful representations which arose in the group case do not arise for the canonical operators, for it is impossible to fulfill the Heisenberg commutation relation, if there is a subspace where φ or π or both act as zero. This is reflected in the fact that all solutions have the property that $\varphi(f)$ and $\pi(g)$ have a strictly absolutely continuous spectrum, as may be seen by applying the test of Eq. (4.22). In all such tests the properties of φ_1 and π_1 are controlling.[33]

6. NONTRIVIAL NATURE OF HAMILTONIAN

As our concluding remarks we want only to indicate, in a simple and heuristic fashion, that the Hamiltonians associated with exponential representations are generally far from trivial and contain terms for production, annihilation, and scattering. Elsewhere[6] we shall study these models in their own right for their physical content and predictions.

A very common feature of a canonical theory is the identity between the momentum operator $\pi(\mathbf{x})$ and the first time derivative of the field $\dot{\varphi}(\mathbf{x})$. If $\mathcal{H}$ denotes the Hamiltonian for the problem, then we require that

$$i[\mathcal{H}, \varphi(\mathbf{x})] = \pi(\mathbf{x}). \quad (6.1)$$

Roughly speaking, this means that

$$\mathcal{H} = \tfrac{1}{2}\int \pi^2(\mathbf{x})\, d\mathbf{x} + \mathfrak{W}, \quad (6.2)$$

where $[\mathfrak{W}, \varphi(\mathbf{x})] = 0$. This form holds true whether or not the canonical operators are given by an irreducible representation. In an exponential representation of the canonical variables, as we considered in Sec. 5, the momentum operator $\pi(\mathbf{x})$ is a bilinear expression in annihilation and creation operators, as is made explicit in Eq. (5.36b). Consequently, $\mathcal{H}$, being quadratic in $\pi(\mathbf{x})$, is (at least) a quartic in these operators, much as in Eq. (2.13). Thus, it is clear that theories exhibiting production, annihilation, and scattering—as these terms are conventionally understood—can be constructed with the aid of exponential representations, since such terms appear in the Hamiltonian. It is noteworthy that these terms already appear in the kinetic energy factor, usually treated as part of the free theory. This is, of course, a consequence of the uncommon representations which we have employed.

[33] H. G. Tucker, Pacific J. Math. **12**, 1125 (1962).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Galilean Relativity, Locality, and Quantum Hydrodynamics*

JOHN C. GARRISON, HARRY L. MORRISON, AND JACK WONG
University of California, Lawrence Radiation Laboratory, Livermore, California

(Received 16 May 1969)

In this work we study the consequences of locality and Galilean covariance for the operators that occur in Landau's quantum hydrodynamics. We specifically consider the following requirements: (1) Galilean covariance of the velocity field, (2) locality of the velocity field, and (3) Landau's assumption that the momentum density is a symmetrized product of the velocity and density operators. It is demonstrated that the density–velocity commutation relation of the Landau theory is essentially a direct consequence of (1) and (2). The addition of (3) is sufficient to determine the velocity–velocity commutation relation, also in agreement with Laudau. We further show that the density–velocity commutation relation, independent of (3) or any specific form for the velocity field, is inconsistent with the nonnegative character of the local density.

1. INTRODUCTION

In this paper we will be exclusively concerned with the theoretical efforts, initiated by Landau,[1] to formulate a description of superfluid helium in terms of a quantized version of the equations of classical hydrodynamics. Our work does not apply to other hydrodynamic theories such as those which use the phase of an appropriately defined condensate wavefunction as a velocity potential.[2] Landau's original proposal was to define a local velocity field in terms of the current and mass density for a many-body

* Work performed under the auspices of the U.S. Atomic Energy Commission.

[1] L. D. Landau, J. Phys. Moscow **5**, 71 (1941).

[2] P. W. Anderson, in *Quantum Fluids, 1965*, D. F. Brewer, Ed. (North-Holland Publ. Co., Amsterdam, 1966), p. 146.

would be

$$d\hat{\sigma}_0(\lambda, \nu) = (\lambda^2 + \nu^2)^{-1} e^{-(\lambda^2+\nu^2)}\, d\lambda\, d\nu. \quad (5.45)$$

We have already noted that when $\|\hat{\varphi}_0\| = \infty$, the representation of φ and π is inequivalent to that based on a Fock representation for A_2. In light of the discussion in Sec. 4, it should be clear that in the example discussed above, if $\hat{\sigma}_0$ has infinite measure, the representation of φ and π is not even "locally Fock." The issues of unfaithful representations which arose in the group case do not arise for the canonical operators, for it is impossible to fulfill the Heisenberg commutation relation, if there is a subspace where φ or π or both act as zero. This is reflected in the fact that all solutions have the property that $\varphi(f)$ and $\pi(g)$ have a strictly absolutely continuous spectrum, as may be seen by applying the test of Eq. (4.22). In all such tests the properties of φ_1 and π_1 are controlling.[33]

6. NONTRIVIAL NATURE OF HAMILTONIAN

As our concluding remarks we want only to indicate, in a simple and heuristic fashion, that the Hamiltonians associated with exponential representations are generally far from trivial and contain terms for production, annihilation, and scattering. Elsewhere[6] we shall study these models in their own right for their physical content and predictions.

A very common feature of a canonical theory is the identity between the momentum operator $\pi(\mathbf{x})$ and the first time derivative of the field $\dot{\varphi}(\mathbf{x})$. If $\mathcal{H}$ denotes the Hamiltonian for the problem, then we require that

$$i[\mathcal{H}, \varphi(\mathbf{x})] = \pi(\mathbf{x}). \quad (6.1)$$

Roughly speaking, this means that

$$\mathcal{H} = \tfrac{1}{2}\int \pi^2(\mathbf{x})\, d\mathbf{x} + \mathfrak{W}, \quad (6.2)$$

where $[\mathfrak{W}, \varphi(\mathbf{x})] = 0$. This form holds true whether or not the canonical operators are given by an irreducible representation. In an exponential representation of the canonical variables, as we considered in Sec. 5, the momentum operator $\pi(\mathbf{x})$ is a bilinear expression in annihilation and creation operators, as is made explicit in Eq. (5.36b). Consequently, $\mathcal{H}$, being quadratic in $\pi(\mathbf{x})$, is (at least) a quartic in these operators, much as in Eq. (2.13). Thus, it is clear that theories exhibiting production, annihilation, and scattering—as these terms are conventionally understood—can be constructed with the aid of exponential representations, since such terms appear in the Hamiltonian. It is noteworthy that these terms already appear in the kinetic energy factor, usually treated as part of the free theory. This is, of course, a consequence of the uncommon representations which we have employed.

[33] H. G. Tucker, Pacific J. Math. **12**, 1125 (1962).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Galilean Relativity, Locality, and Quantum Hydrodynamics*

JOHN C. GARRISON, HARRY L. MORRISON, AND JACK WONG
University of California, Lawrence Radiation Laboratory, Livermore, California

(Received 16 May 1969)

In this work we study the consequences of locality and Galilean covariance for the operators that occur in Landau's quantum hydrodynamics. We specifically consider the following requirements: (1) Galilean covariance of the velocity field, (2) locality of the velocity field, and (3) Landau's assumption that the momentum density is a symmetrized product of the velocity and density operators. It is demonstrated that the density–velocity commutation relation of the Landau theory is essentially a direct consequence of (1) and (2). The addition of (3) is sufficient to determine the velocity–velocity commutation relation, also in agreement with Laudau. We further show that the density–velocity commutation relation, independent of (3) or any specific form for the velocity field, is inconsistent with the nonnegative character of the local density.

1. INTRODUCTION

In this paper we will be exclusively concerned with the theoretical efforts, initiated by Landau,[1] to formulate a description of superfluid helium in terms of a quantized version of the equations of classical hydrodynamics. Our work does not apply to other hydrodynamic theories such as those which use the phase of an appropriately defined condensate wavefunction as a velocity potential.[2] Landau's original proposal was to define a local velocity field in terms of the current and mass density for a many-body

* Work performed under the auspices of the U.S. Atomic Energy Commission.

[1] L. D. Landau, J. Phys. Moscow **5**, 71 (1941).

[2] P. W. Anderson, in *Quantum Fluids, 1965*, D. F. Brewer, Ed. (North-Holland Publ. Co., Amsterdam, 1966), p. 146.

system. As a simple example, we may consider a system of identical, spinless bosons of mass m described by a second-quantized field operator $\psi(x)$. The mass and momentum densities are then given by[3]

$$\rho(x) \equiv m\psi^\dagger(x)\psi(x),$$
$$\mathbf{J}(x) \equiv (2i)^{-1}[\psi^\dagger(x)\nabla\psi(x) - \nabla\psi^\dagger(x)\psi(x)],$$

and Landau's velocity field is defined by

$$\mathbf{v}(x) \equiv \tfrac{1}{2}[\rho^{-1}(x)\mathbf{J}(x) + \mathbf{J}(x)\rho^{-1}(x)]. \tag{1.1}$$

By using this definition, one can, at least formally, obtain the commutation relations involving $\mathbf{v}$; they are[4]

$$[\mathbf{v}(x), \rho(x')] = -i\nabla\delta(\mathbf{x} - \mathbf{x}'), \tag{1.2}$$

$$[v_j(x), v_k(x')] = i\rho^{-1}(x)[\nabla_j v_k(x) - \nabla_k v_j(x)]\delta(\mathbf{x} - \mathbf{x}'). \tag{1.3}$$

The order of factors on the right-hand side of (1.3) is irrelevant, since (1.2) implies

$$[\rho(x'), \operatorname{curl} \mathbf{v}(x)] = 0.$$

The problem of imposing these commutation relations on the classical hydrodynamic equations has been extensively studied, but we will not review this aspect of the theory, since it does not enter into our considerations. Instead we will be concerned with difficulties which arise directly from the commutation relations. In this connection we should first mention several objections which have previously been made to Landau's program. The first is essentially mathematical: namely, it is highly unlikely that any sensible meaning can be attached to the inverse of the mass-density operator; one can always construct a state in which there are no particles in the vicinity of some point. From the physical point of view, London[5] argued that the existence of a velocity operator defined at a point would contradict the uncertainty principle. While this reasoning is suggestive, its exactness is open to question, since the argument of a field operator need not be identified with the position of any particle.

An alternative version of Landau's program was suggested by Kronig and Thellung[6]; however, their theory only allowed irrotational flow. A complete theory was formulated by Thellung,[7] Itô,[8] and Ziman.[9] The Ziman version has been applied by several authors[10,11] to obtain various properties of the excitation spectrum of superfluid helium. Instead of proposing a microscopic definition for the velocity field, Ziman remarked that the equations of classical hydrodynamics could be obtained from a variational principle. The existence of a Lagrangian then allowed the use of the general technique of canonical quantization. Since the theory has no necessary connection with an underlying particle theory, it is rather striking that the commutation relations involving $\mathbf{v}$ are exactly those obtained earlier by Landau.

A rather serious objection to the Ziman theory has been made recently by Fröhlich,[12] who showed that the existence of a quantity canonically conjugate to the mass density is inconsistent with the fact that the spectrum of the total number operator is the set of nonnegative integers.

The purpose of the present paper is to establish two results: In Sec. 2 we show that the Landau commutation relations represent the simplest solution to the constraints imposed by locality and Galilean relativity. Thus they are not necessarily associated with the specific microscopic definition used by Landau or the canonical formalism constructed by Ziman. In Sec. 3 we further show that the commutation relation (1.2) is inconsistent with the physical requirement that the expectation value of the mass density be nonnegative. In Sec. 4 we briefly discuss the meaning of these results.

2. THE LANDAU COMMUTATION RELATIONS

The only Galilean transformation we have to consider is the so-called velocity transformation

$$\mathbf{x}' = \mathbf{x} + \mathbf{u}t,$$

with the corresponding transformations for ψ and v:

$$\psi'(x') = \exp[im(\mathbf{u}\cdot\mathbf{x} + \tfrac{1}{2}u^2t)]\psi(x),$$
$$\mathbf{v}'(x') = \mathbf{v}(x) + \mathbf{u}. \tag{2.1}$$

The second of these equations constitutes the requirement of Galilean covariance for the velocity field. An application of conventional field-theory techniques[13] to the Galilei group yields a unitary operator corresponding to the transformation of ψ:

$$\psi'(x) = U\psi(x)U^{-1}, \tag{2.2}$$

[3] We put $\hbar = 1$ and use the notation $x = (\mathbf{x}, t)$. All commutators are to be understood as equal-time commutators.

[4] We should remark that the sign of the right-hand member of (1.3) is opposite to that given in Landau's original paper (Ref. 1.). This is undoubtedly due to a misprint, since the commutation relations, as given there, are inconsistent with the equations of motion, which are given correctly.

[5] F. London, Rev. Mod. Phys. **17**, 310 (1943).

[6] R. Kronig and A. Thellung, Physica **18**, 749 (1952).

[7] A. Thellung, Physica **19**, 217 (1953).

[8] I. Itô, Progr. Theoret. Phys. (Kyoto) **9**, 117 (1953).

[9] J. M. Ziman, Proc. Roy. Soc. (London) **A219**, 257 (1953).

[10] G. R. Allcock and C. G. Kuper, Proc. Roy. Soc. (London) **A231**, 226 (1955).

[11] A. Thellung, Helv. Phys. Acta **29**, 103 (1956).

[12] H. Fröhlich, Physica **34**, 47 (1967).

[13] J. Schwinger, Phys. Rev. **82**, 914 (1951).

where

$$U = \exp[i(t\mathbf{u}\cdot\mathbf{P} - \mathbf{u}\cdot\mathbf{X} - \tfrac{1}{2}Mu^2t)],$$

and

$$\mathbf{P} = \int d^3x\mathbf{J}(x),$$

$$\mathbf{X} = \int d^3x\mathbf{x}\rho(x),$$

$$M = \int d^3x\rho(x).$$

The operator $\mathbf{P}$ represents the total momentum, while $\mathbf{X}$ is related to the center-of-mass position and M is the total-mass operator. Note that (2.2) is an *active* transformation; that is, the functional form of ψ is altered while the argument remains unchanged. However, by making use of the fact that $\mathbf{P}$ is the generator of spatial translations, one can construct a unitary operator which yields the transformation (2.1) directly.[14] Thus, if we define $S(\mathbf{u})$ by

$$S(\mathbf{u}) = \exp[-i(\mathbf{u}\cdot\mathbf{X} + \tfrac{1}{2}Mu^2t)], \tag{2.3}$$

then we find

$$\psi'(x') = S(\mathbf{u})\psi(x)S^{-1}(\mathbf{u}). \tag{2.4}$$

Alternatively, the validity of (2.4) can be established by direct calculation.

We now introduce the basic assumption that $\mathbf{v}(x)$ has no explicit x dependence. Since every operator can be regarded as a function of ψ, $\psi^\dagger$, and their spatial derivatives, it follows that the transformation of the velocity field is also effected by $S(\mathbf{u})$; therefore, in view of (2.1), we have

$$v_k(x) + u_k = S(\mathbf{u})v_k(x)S^{-1}(\mathbf{u}).$$

In order to see the consequences of this equation, we first operate on both sides with $(\partial/\partial u_j)$ and use the explicit expression (2.3) to obtain

$$\delta_{kj} = iS(\mathbf{u})[v_j(x), (X_k + Mu_kt)]S^{-1}(\mathbf{u}).$$

Next operate from the left with S^{-1} and from the right with S to find

$$[v_j(x), X_k] + tu_k[v_j(x), M] = -i\delta_{kj}, \tag{2.5}$$

which yields

$$[v_j(x), M] = 0, \tag{2.6}$$

$$[v_j(x), X_k] = -i\delta_{kj}, \tag{2.7}$$

since (2.5) is an identity in $\mathbf{u}$. Now we are ready to impose a precise statement of locality. We require

$$[v_j(x), \rho(x')] = \sum_{n=0}^{N} T_{jk_1\cdots k_n}(x)\nabla_{k_1}\cdots\nabla_{k_n}\delta(\mathbf{x} - \mathbf{x}'), \tag{2.8}$$

where the coefficients in the finite-order differential operator on the right may be q-numbers.[15] The relations (2.6) and (2.7) are simply the first two moments of (2.8), so we immediately conclude that

$$T_j(x) \equiv 0,$$

$$T_{jk}(x) = -i\delta_{jk}.$$

Then we can write (2.8) in the form

$$[v_j(x), \rho(x')] = -i\nabla_j\delta(\mathbf{x} - \mathbf{x}') + O_2(x)\delta(\mathbf{x} - \mathbf{x}'), \tag{2.9}$$

where O_2 is a finite-order differential operator with order at least two. Thus the simplest solution to the constraints imposed by locality and Galilean covariance is (1.2).

In order to obtain the remaining commutation relations, we assume that the operator $\mathbf{K}$, defined by

$$\mathbf{K} = \tfrac{1}{2}\int d^3x[\rho(x)\mathbf{v}(x) + \mathbf{v}(x)\rho(x)], \tag{2.10}$$

differs from the total momentum $\mathbf{P}$ by an operator which commutes with both $\rho(x)$ and $\mathbf{v}(x)$. This condition is certainly satisfied in the Ziman theory, for which $\mathbf{K}$ and $\mathbf{P}$ are identical; and it appears to have been assumed by Landau. Our assumption therefore contains the relevant aspect of both theories, and we have

$$[K_j, v_k(x)] = i\nabla_j v_k(x).$$

Then, from (2.10) and (1.2), we obtain

$$\tfrac{1}{2}\int d^3x'\{\rho(x')[v_j(x'), v_k(x)] + [v_j(x'), v_k(x)]\rho(x')\} = i[\nabla_j v_k(x) - \nabla_k v_j(x)]. \tag{2.11}$$

The apparent difficulty caused by the symmetrized form of the integrand can be eliminated by using the Jacobi identity:

$$\{\rho(x'), [v_j(x''), v_k(x)]\} + \{v_k(x), [\rho(x'), v_j(x'')]\} + \{v_j(x''), [v_k(x), \rho(x')]\} = 0.$$

Since the commutator of ρ and $\mathbf{v}$ is a c-number, the second and third terms above vanish, and we conclude that ρ commutes with the velocity commutator. It then follows easily that the simplest local commutation relation consistent with (2.11) is

$$[v_j(x'), v_k(x)] = i\rho^{-1}(x)[\nabla_j v_k(x) - \nabla_k v_j(x)]\delta(\mathbf{x} - \mathbf{x}'), \tag{2.12}$$

which is the same as (1.3) after interchanging x and x'. So far, we have not specified the domain of definition

[14] This is easily seen by first considering infinitesimal transformations and then exponentiating.

[15] We use Dirac's terminology in which operators acting in the Hilbert space of many-body wavefunctions are called q-numbers, while ordinary (complex-valued) functions are referred to as c-numbers.

of the commutation relations; they are, as usual, assumed to hold on a common dense subspace of the physical Hilbert space. We also assume that the commutation relation (1.2) holds in the so-called Weyl form,[16] which in our case becomes

$$\exp\{i\rho(f)\}\exp\{iv(\mathbf{g})\} = \exp\{i(f, \nabla\cdot\mathbf{g})\}\exp\{iv(\mathbf{g})\}\exp\{i\rho(f)\}$$

with

$$\rho(f) \equiv \int d^3x\rho(x)f(x),$$

$$v(\mathbf{g}) \equiv \int d^3x\mathbf{g}(x)\cdot\mathbf{v}(x),$$

$$(f, \nabla\cdot\mathbf{g}) = \int d^3x f(x)\nabla\cdot\mathbf{g},$$

where f and g_k are suitable real-valued test functions. This assumption is required in order to guarantee the existence and differentiability of the unitary operator $U(\lambda)$ introduced in the following section.

Thus we see that the Landau commutation relations follow in a natural way from the requirements of Galilean relativity and locality, together with the assumption that $\mathbf{K}$ is the generator of spatial translations. Consequently, the commutation relations do not depend on the explicit, constructive definition given by Landau or on the canonical formalism of Ziman.

3. DIFFICULTIES ARISING FROM THE COMMUTATION RELATIONS

We will now establish that the Landau commutation relations are not consistent with the fact that the mass density has nonnegative expectation values. As a special case we will recover the result of Fröhlich mentioned in Sec. 1. Both results follow from a slightly more general mathematical fact which we state in the form of a theorem. First, we define the *numerical range* of an operator as the set of complex numbers obtained by forming the expectation value of the operator with respect to all normalized state vectors in the space upon which the operator acts.

Theorem: Let $A(x)$ and $B(x)$ be q-number fields with B Hermitian, and assume the following commutation relations:

$$[B(x'), A(x)] = i\sum_{n=0}^{N} C_{k_1\cdots k_n}(x)\nabla_{k_1}\cdots\nabla_{k_n}\delta(\mathbf{x}'-\mathbf{x}),$$

$$[B(x'), C_{k_1\cdots k_n}(x)] = 0,$$

where at least one of the C's is nonzero; then the numerical range of A includes the entire real axis.

Proof: Define

$$U(\lambda) = \exp\left(i\lambda\int d^3x' f(x')B(x')\right),$$

where f is a suitable real-valued c-number function (e.g., a test function in the sense of distribution theory). Since B is Hermitian, $U(\lambda)$ is unitary for any real λ. We now transform A by $U(\lambda)$ to get

$$A(x, \lambda) = U(\lambda)AU^{-1}(\lambda).$$

The new operator satisfies a simple differential equation

$$\begin{aligned}\frac{\partial A(x,\lambda)}{\partial\lambda} &= iU(\lambda)\int d^3x' f(x')[B(x'), A(x)]U^{-1}(\lambda)\\ &= -U(\lambda)\sum_n C_{k_1\cdots k_n}(x)\nabla_{k_1}\cdots\nabla_{k_n}f(x)U^{-1}(\lambda)\\ &= -\sum_n C_{k_1\cdots k_n}(x)\nabla_{k_1}\cdots\nabla_{k_n}f(x).\end{aligned}$$

The second and third lines follow from the assumed commutation relations. The solution is

$$A(x, \lambda) = A(x) - \lambda\sum_n C_{k_1\cdots k_n}(x)\nabla_{k_1}\cdots\nabla_{k_n}f(x). \tag{3.1}$$

Let Ψ be any normalized state vector and define $\Psi(\lambda)$ by

$$\Psi(\lambda) = U^{-1}(\lambda)\Psi.$$

Then (3.1) yields

$$\langle\Psi(\lambda)|A(x)|\Psi(\lambda)\rangle = \langle\Psi|A(x)|\Psi\rangle - \lambda\sum_n\langle\Psi|C_{k_1\cdots k_n}(x)|\Psi\rangle\nabla_{k_1}\cdots\nabla_{k_n}f(x).$$

We see that $\mathrm{Re}\,\langle\Psi(\lambda)|A|\Psi(\lambda)\rangle$ can be made to have any desired value by a suitable choice of λ, unless Ψ satisfies

$$\langle\Psi|C_{k_1\cdots k_n}(x)|\Psi\rangle \equiv 0. \tag{3.2}$$

This relation cannot be satisfied by all Ψ's since that would imply that all the C's are zero, which would contradict the hypothesis of the theorem. Consequently, we can always pick Ψ so that (3.2) is not satisfied, and the conclusion of the theorem follows.

The result of Fröhlich can be obtained by setting $A(x) = \rho(x)$ and $B(x) = \phi(x)$, where ϕ is canonically conjugate to ρ; that is,

$$[\phi(x'), \rho(x)] = i\delta(\mathbf{x}'-\mathbf{x}). \tag{3.3}$$

The hypothesis of the theorem is satisfied by this choice. Since the numerical range of a Hermitian

[16] H. Araki, J. Math. Phys. 1, 492 (1960).

operator is necessarily a subset of the real axis, we conclude in this case that the numerical range of $\rho(x)$ is the entire real axis. In other words, it is always possible to find a state in which the expectation value of $\rho(x)$ is any desired negative number. Thus a canonical commutation relation such as (3.3) cannot be satisfied by a nonnegative operator.

It might still be thought that the situation could be saved by giving up the canonical formalism. However, if we choose $B(x) = \mathbf{C} \cdot \mathbf{v}(x)$ with a fixed real vector $\mathbf{C}$, then (1.2) gives

$$[B(x'), \rho(x)] = i\mathbf{C} \cdot \nabla\delta(x' - x).$$

Again we can invoke the theorem to show the existence of states in which the expectation value of $\rho(x)$ is negative. Thus the difficulty lies in the relation (1.2) itself, not in the specific definition of $\mathbf{v}$ or in any underlying formalism.

4. CONCLUSION

In discussing the significance of the results we have obtained, it is necessary to differentiate between theories of the Landau type, which are directly based on an underlying microscopic theory, and those of the Ziman type, which involve the ad hoc procedure of quantizing classical hydrodynamics as a field theory.

We begin with theories of the Landau type, for which it is possible to make more definite statements. In this case the mass-density operator is certainly nonnegative; therefore, we can definitely conclude that any such theory that includes (1.2) is not internally consistent, either physically or mathematically. In particular, the formal definition (1.1) is inadmissible. It should be emphasized again that (1.2) follows from locality and Galilean convariance alone: neither (1.1) nor the identification of $\mathbf{K}$ as the generator of spatial translations is required. This situation leaves open the possibility of different definitions of $\mathbf{v}$ which do not lead to (1.2). One alternative is that $\mathbf{v}$ is still a local operator, but that some higher-order terms, as indicated in (2.9), are included in the commutation relation. As a consequence of the theorem proved in Sec. 3, the coefficients retained in the differential operator O_2 would have to be q-numbers not commuting with $\mathbf{v}$; otherwise the contradiction would still follow. Another possibility is that $\mathbf{v}$ is nonlocal; that is, it does not satisfy a local commutation relation of the form (2.9). In this connection it might be useful to look for an operator formalism analogous to the nonlocal definition of the c-number velocity field in terms of the one-particle density matrix developed by Frohlich.[17,18]

The case of theories of the Ziman type is less clear for the simple reason that the physical basis of such theories is itself unclear. Thus it might be argued that the existence of negative density states is an unfortunate but curable sickness, somewhat similar to the difficulty of states of negative norm in the Gupta–Bleuler version of electrodynamics. If such an interpretation of the Ziman theory is to succeed, it must be shown that the subsidiary conditions required to eliminate the states of negative density are consistent with the dynamics of the system. That is, one must show that there are no transitions from physical to unphysical states. In the absence of such precautions, any results obtained should be viewed with considerable skepticism.

ACKNOWLEDGMENTS

One of us (H. L. M.) wishes to acknowledge useful discussions with Dr. W. Bierter. We also wish to thank the referee for several useful comments and for supplying the reference to the work of H. Itô.

[17] H. Fröhlich, Physica 37, 215 (1967).

[18] For another alternative, which involves the study of the algebra of currents and densities, see W. Bierter and H. L. Morrison, J. Low Temp. Phys. 1, 65 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Quantum Theory on a Network.* I. A Solvable Model Whose Wavefunctions Are Elementary Functions

ELLIOTT W. MONTROLL
Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627

(Received 13 June 1969)

A network model is developed for electrons in molecules and crystals. Wavefunctions and energy levels can be calculated exactly. A network which is topologically equivalent to that of the molecule or crystal of interest is constructed. One-dimensional bonds connect various node points. The potential $V(x) = -V_0 \operatorname{sech}^2 \gamma x$ is postulated to act along each bond so that a potential well exists at each node point (atom) of the network. For some values of V_0 the wavefunctions are elementary while generally they are hypergeometric functions. When the parameter γ is very large, the model is essentially the tight binding model and, when it is very small, one has the free-electron network model investigated by Ruedenberg and Scherr, Griffith, Coulson,and others. The density of states, energy band structure, and other features are discussed.

I. INTRODUCTION

Exploratory investigations of internal vibrations of molecules and crystal lattices generally start from a network model of masses connected by springs. Those of phase transitions generally start with the Ising model, which is also a network model, but one with two possible states identified with each node of the network. The author has occasionally wondered why network models have not been used more extensively in solid-state physics courses and as a basis for the analysis of complex phenomenon in the electronic theory of solids.

Free electrons on zigzag wires and various ring structures have been used by a number of authors[1–6] for discussions of the spectra of hydrocarbon chains with conjugated double bonds and for a variety of ring-structure organic molecules. Coulson[7] (see also Hoerni[8]) has investigated free-electron network models of various lattices. He has, among other things, found the density of states for such a model of graphite. Della Riccia[9] has investigated the free-electron network model for diamond. However, in recent years the computer has taken over; very complicated calculations are being made on complex models, and the simple network theory has become neglected.

Recent conversations with Pell and Matthias about complicated materials such as amorphous semiconductors and superconducting alloys have stimulated the author to examine certain "solvable" network models of periodic lattices. It is his feeling that, before introducing complications into a system, one should have a simple model whose unperturbed properties can be easily described without requiring perturbation theory.

The model which we analyze in this paper is the microscopic scaled-down version of the ball and wire models of crystals used in lecture demonstrations and by researchers who want to visualize the structures of the crystals they are dealing with. We consider a network (periodic in the case of crystals and not necessarily so in the case of molecules) of atoms, 1, 2, $\cdots$, with bonds b_{ij} connecting the ith and jth, as shown in Fig. 1. Electrons are postulated to be restricted to move only along the bonds so that the electron wavefunction in a segment between two connected node points is a 1-dimensional wavefunction. In this paper spins are ignored as are electron correlations through Coulomb interactions. All node points will be held fixed. The influence of these neglected effects will be discussed in later papers.

We will associate a potential well with each node point and choose it so that a single electronic bound

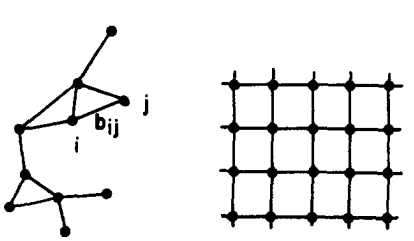


FIG. 1. Networks, aperiodic and periodic.

* This work was partially supported by the Office of Naval Research.

[1] L. Pauling, J. Chem. Phys. **4**, 673 (1936).
[2] H. Kuhn, Helv. Chim. Acta **31**, 1441 (1948); J. Chem. Phys. **18**, 840 (1948); *ibid.* **22**, 2098 (1954).
[3] J. R. Platt, J. Chem. Phys. **17**, 484 (1949); **21**, 1597 (1953).
[4] N. S. Bayliss, J. Chem. Phys. **16**, 287 (1948).
[5] J. S. Griffith, J. Chem. Phys. **21**, 174 (1953); Trans. Faraday Soc. **49**, 345 (1953).
[6] K. Ruedenberg and C. W. Scherr, J. Chem. Phys. **21**, 1565 (1953); **22**, 151 (1954); C. W. Scherr, *ibid.* **21**, 1582 (1953); K. Ruedenberg, *ibid.* **22**, 1878 (1954).
[7] C. A. Coulson, Proc. Phys. Soc. (London) **67**, 608 (1954); **68**, 1129 (1953).
[8] J. A. Hoerni, J. Chem. Phys. **34**, 508 (1961).
[9] G. Della Riccia, *Proceedings of the Semiconductor Conference, Exeter* (Institute of Physics and the Physical Society, London, 1962), p. 570.

state would exist for each node point if there were no connections to other node points. In this paper the potential will be chosen so that the wavefunctions along any bond connecting two node points are elementary functions. Energy bands, density of states, etc., will be derived for various lattice structures. All calculations will be exact and, indeed, very similar to certain lattice-vibration calculations.

II. BOUNDARY CONDITIONS AT NODE POINTS

Assume that the wavefunction at a distance x_{ij} from the ith point in the direction of the jth is

$$\phi^{ij}(x_{ij}) = a_{ij} f_{ij}(x_{ij}, \delta_{ij}, k), \tag{1}$$

where a_{ij} and δ_{ij} are the two integration constants which appear in the integration of the second-order Schrödinger equation. If l_{ij} is the length of bond b_{ij}, then

$$0 < x_{ij} < l_{ij}. \tag{1'}$$

The parameter k in (1) is the quantum number which tells us which wavefunction is being discussed.

The first boundary condition required is that the wavefunction be continuous everywhere on the net. Hence, at node point j,

$$a_{jj_1} f_{jj_1}(0, \delta_{jj_1}) = a_{jj_2} f_{jj_2}(0, \delta_{jj_2}) = \cdots = a_{jj_\beta} f_{jj_\beta}(0, \delta_{jj_\beta}), \tag{2a}$$

if $j_1 \cdots j_\beta$ is the set of β nodes to which node j is attached.

It has been shown by Ruedenberg and Scherr[6] that, as a consequence of momentum conservation at each node ("Kirchhoffs law"),

$$\sum_\alpha a_{jj_\alpha} \left(\frac{df_{jj_\alpha}}{dx_{jj_\alpha}}\right)_0 = 0, \tag{2b}$$

so that, after dividing (2b) by $a_{jj_1} f_{jj_1}$, we have

$$\sum_\alpha \left(\frac{a_{jj_\alpha}}{a_{jj_1} f_{jj_1}(0, \delta_{jj_1})}\right) \left(\frac{df_{jj_\alpha}}{dx_{jj_\alpha}}\right)_0 = 0. \tag{3}$$

The 0 subscript on the parenthesis indicates that the derivatives are to be evaluated at $x_{jj_\alpha} = 0$. By combining (2a) and (3), we find

$$\sum_\alpha \left(\frac{d \log f_{jj_\alpha}}{dx_{jj_\alpha}}\right)_0 = 0, \quad j = 1, 2, 3, \cdots. \tag{4}$$

The conservation condition (2b) is the statement that at all nodes

$$\sum \text{grad } \phi = 0,$$

where the gradient is measured along each network line away from the node and ϕ represents the wavefunction at the node.

III. FREE PARTICLES ON NET

In order to see how the boundary conditions are applied, we first consider briefly free particles on a network.[7,6]

The wavefunction of a free particle with energy $E = \hbar^2 k^2/2m$ is

$$\phi(x) = a \cos(kx + \delta) = a(\cos kx \cos \delta - \sin kx \sin \delta). \tag{5}$$

Hence,

$$\phi(0) = a \cos \delta, \tag{6a}$$

$$\phi(l) = a \cos \delta(\cos kl - \sin kl \tan \delta), \tag{6b}$$

so that

$$\phi(l) = \phi(0)(\cos kl - \sin kl \tan \delta) \tag{7}$$

and

$$\tan \delta = [\cos kl - \phi(l)/\phi(0)]/\sin kl; \tag{8}$$

also

$$a^2 = \phi^2(0)(1 + \tan^2 \delta). \tag{9}$$

Finally,

$$\phi'(x) = ak(-\sin kx \cos \delta - \cos kx \sin \delta) \tag{10}$$

and

$$\phi'(0)/\phi(0) = -k \tan \delta. \tag{11}$$

Hence, for this case, using the notation of Eqs. (2) and (4), we have

$$\left(\frac{d(\log f_{jj_\alpha})}{dx_{jj_\alpha}}\right)_0 = -k \tan \delta_{jj_\alpha}. \tag{12}$$

Then, our conservation of momentum condition (4) becomes

$$\sum_\alpha \tan \delta_{jj_\alpha} = 0, \quad j = 1, 2, \cdots, n, \tag{13}$$

or, from (8), *if all bonds have length l*,

$$n_j \phi[j] \cos kl - \sum_\alpha \phi[j_\alpha] = 0, \quad j = 1, 2, \cdots, n, \tag{14}$$

where we use the following notation: $\phi[j]$ is the wavefunction at the jth node of the network; $\phi_{jj_\alpha}(l)$ of (7) becomes $\phi[j_\alpha]$ if l is the distance j to j_α; n_j is the total number of bonds which are connected to j.

The set of equations (14) are identical with those which appear in the vibration of lattices whose atomic displacements in the x, y, and z directions are independent of each other and in which central- and noncentral-force constants are identical. The ϕ's would correspond to the spatial-dependent factor in the displacement of atoms from their equilibrium positions. The factor $m\omega^2/2\gamma$ which appears in lattice-vibration theory (m being the mass of the vibrating

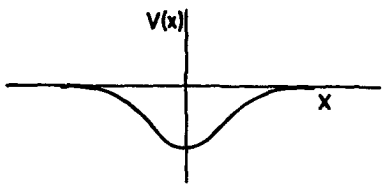


FIG. 2. Schematic form of the potential $V(x) = -V_0 \operatorname{sech}^2 x$.

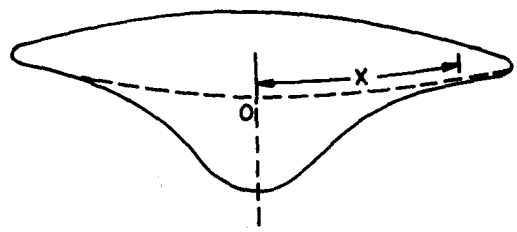


FIG. 3. Ring of circumference l with potential well centered at origin.

particle, ω the normal mode frequency, and γ the force constant) is the analog of $n_j(1 - \cos kl)$. As we shall see in the following sections, replacement of the free-electron condition by one with the electrons in an atomic force field introduces a more complicated "form factor" than $n_j \cos kl$. However, the general structure of the equations still have the form (when all $l_{ij} = l$)

$$F(k, \eta)\phi[j] = \sum_{\alpha} \phi[j_\alpha], \tag{14'}$$

where the form factor F depends on the energy through k and the parameters η which characterize the atomic potential which effects the electrons. In this form, all the mathematical apparatus which has been developed for the theory of lattice vibrations is immediately applicable.

IV. A POTENTIAL WITH A SINGLE BOUND STATE AND ELEMENTARY WAVEFUNCTIONS

Once a 2- or 3-dimensional assembly is modeled by a network, the wavefunctions are defined only along the various 1-dimensional bonds which connect node points. There are a number of 1-dimensional solvable potentials[10] whose wavefunctions are tabulated functions.

The particular potential which we use in this paper is $V(x)$, such that

$$2mV(x)/\hbar^2 = -2\gamma^2 \operatorname{sech}^2 x\gamma. \tag{15a}$$

If we set

$$k^2 = 2mE/\hbar^2, \tag{15b}$$

the Schrödinger equation becomes

$$\phi'' + (k^2 + 2\gamma^2 \operatorname{sech}^2 \gamma x)\phi = 0. \tag{15c}$$

This 1-parameter potential is a special case of the more general potential

$$V(x) = -V_0 \operatorname{sech}^2 \gamma x, \tag{15d}$$

which has the form plotted in Fig. 2. If the center of the well is located at the node positions, this makes each atomic center an attractive center to an electron in the network. The general solution of the Schrödinger equation with the potential (15d) is in terms of hypergeometric functions.[11] Its application to networks will be discussed later by Mills and the author. Here we limit ourselves to the special relation between V_0 and γ which leads to wavefunctions which depend only on circular and hyperbolic functions.[12] The main ideas of the theory can then be presented with a minimum of mathematical complexity.

The general solution of (15c) can be shown to be

$$\phi(x) = a[\cos(kx + \delta) - (\gamma/k)\sin(kx + \delta)\tanh \gamma x] \tag{16}$$

by direct substitution into (15c). We will use a circular ring configuration and set the well of the potential at the origin as shown in Fig. 3. The length of the chain is postulated to be l. The derivative of (16) is

$$\phi'(x) = -ak[\sin(kx + \delta) + (\gamma/k)\cos(kx + \delta)\tanh \gamma x + (\gamma^2/k^2)\operatorname{sech}^2 \gamma x \sin(kx + \delta)]. \tag{17}$$

If we close the ring by connecting the point at $x = \frac{1}{2}l$ with $x = -\frac{1}{2}l$, the function $\phi(x)$ as well as its derivative must be continuous at the point of closure. The continuity of $\phi(x)$ implies

$$\cos(\tfrac{1}{2}kl + \delta) - (\gamma/k)\sin(\tfrac{1}{2}kl + \delta)\tanh \tfrac{1}{2}\gamma l = \cos(\tfrac{1}{2}kl - \delta) - (\gamma/k)\sin(\tfrac{1}{2}kl - \delta)\tanh \tfrac{1}{2}\gamma l,$$

so that

$$\sin \delta[\sin \tfrac{1}{2}kl + (\gamma/k)\cos \tfrac{1}{2}kl \tanh \tfrac{1}{2}\gamma l] = 0. \tag{18}$$

The required continuity of $\phi'(x)$ at the closure point yields

$$\cos \delta\{\sin \tfrac{1}{2}kl[1 + (\gamma/k)^2 \operatorname{sech}^2 \tfrac{1}{2}\gamma l] + (\gamma/k)\cos \tfrac{1}{2}kl \tanh \tfrac{1}{2}\gamma l\} = 0. \tag{19}$$

There are then two classes of wavefunctions: those with $\sin \delta = 0$ and those with $\cos \delta = 0$. In the first case, $\delta = 0$ [while a choice $\delta = \pi$ is possible, it is equivalent to merely changing the sign of the constant a in (16) and is, therefore, not a new solution]. In the

[10] M. F. Manning, Phys. Rev. **38**, 161 (1935).

[11] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon Press, New York, 1965).

[12] E. W. Montroll and G. F. Newell, J. Appl. Phys. **23**, 184 (1952).

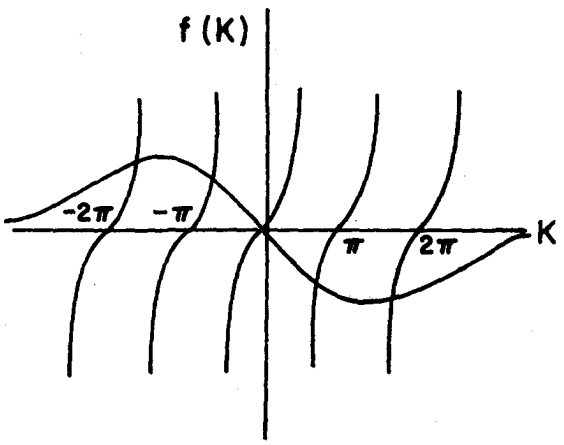


FIG. 4. Each intersection of a branch of tan K with $f(K)$ is associated with a conduction-electron energy-state.

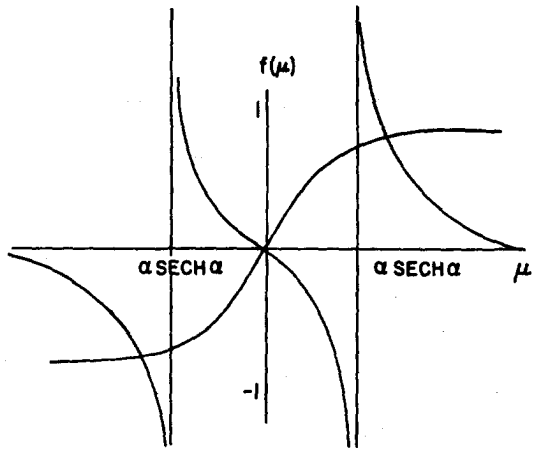


FIG. 5. Each intersection of tanh μ with the three branches of the function on the right-hand side of (24b) corresponds to a bound-state energy.

second case, $\delta = \frac{1}{2}\pi$. When $\delta = 0$, the wavefunction

$$\phi(x) = a \cos kx[1 - (\gamma/k) \tan kx \tanh \gamma x] \quad (20a)$$

is symmetrical. When $\delta = \frac{1}{2}\pi$, we find that

$$\phi(x) = b \sin kx[1 + (\gamma/k) \cot kx \tanh \gamma x], \quad (20b)$$

which is antisymmetrical.

Let us first consider the symmetrical wavefunctions by choosing $\delta = 0$. The only way both (18) and (19) can be satisfied is by letting k satisfy

$$\sin \tfrac{1}{2}kl[1 + (\gamma^2/k^2) \operatorname{sech}^2 \tfrac{1}{2}\gamma l] = -(\gamma/k) \cos \tfrac{1}{2}kl/\tanh \tfrac{1}{2}\gamma l \quad (21)$$

or, if we let

$$K = \tfrac{1}{2}kl \quad \text{and} \quad \tfrac{1}{2}\gamma l = \alpha,$$

$$\tan K = -\frac{(\alpha/K) \tanh \alpha}{1 + (\alpha^2/K^2) \operatorname{sech}^2 \alpha} = -\frac{K\alpha \tanh \alpha}{K^2 + \alpha^2 \operatorname{sech}^2 \alpha}. \quad (22)$$

To find the roots of this equation, we plot left- and right-hand sides (see Fig. 4) on the same graph and search for intersections. There is an intersection and, therefore, a root in each interval $(\pm\pi, \pm 3\pi)$, $(\pm 3\pi, \pm 5\pi)$, etc., as well as one at the origin.

The wavefunction associated with $k = 0$ is

$$\phi_0(x) = a(1 - \gamma x \tanh \gamma x), \quad \text{if} \quad -\tfrac{1}{2}l < x < \tfrac{1}{2}l \quad (23)$$

and the associated energy is $E = 0$.

A bound-state solution can be found by letting

$$K = i\mu. \quad (24a)$$

Then, (22) becomes

$$\tanh \mu = (\mu\alpha \tanh \alpha)/(\mu^2 - \alpha^2 \operatorname{sech}^2 \alpha). \quad (24b)$$

The analog of Fig. 4 is given in Fig. 5. Since $E \propto k^2$, $E \propto -\mu^2$ which corresponds to a bound state.

Notice that, as $\alpha \to \infty$ (which is the case as $l \to \infty$ or $\gamma \to \infty$), $\alpha^2 \operatorname{sech}^2 \alpha \sim (\alpha^2/4) \exp(-2\alpha) \to 0$. Hence, $\mu \to \alpha$ and the bound-symmetrical-state wavefunction becomes

$$\phi(x) \sim a \operatorname{sech}(2\alpha x/l), \quad \text{as} \quad l \to \infty \quad \text{or} \quad \gamma \to \infty, \quad \sim 2a \exp(-2\alpha |x|/l), \quad \text{as} \quad \alpha x \to \infty \quad \text{with } \gamma \text{ and } l \text{ fixed.} \quad (25)$$

The energy of the bound state

$$E = \hbar^2 k^2/2m \simeq -\gamma^2\hbar^2/2m \quad (26)$$

is approximately one-half the well depth $V(0) = -\gamma^2\hbar^2/m$ and is, indeed, exactly one-half in the limit $l \to \infty$ or $\gamma \to \infty$.

The antisymmetric wavefunctions which result from $\delta = \frac{1}{2}\pi$ must have k values which are roots of

$$\tan K = -(\alpha/K) \tanh \alpha, \quad (27)$$

if both (18) and (19) are to be satisfied. Every point which corresponds to an intersection of the function on the left with that on the right yields a possible K value and, hence, a possible energy level; such an intersection appears in each of the intervals $(\frac{1}{2}\pi, \frac{3}{2}\pi)$, $(\frac{3}{2}\pi, \frac{5}{2}\pi)$, $(\frac{5}{2}\pi, \frac{7}{2}\pi), \cdots$, and $(-\frac{3}{2}\pi, -\frac{1}{2}\pi)$, $(-\frac{5}{2}\pi, -\frac{3}{2}\pi)$, etc. The energy associated with each intersection is positive. If one lets $\kappa = i\mu$, then (27) becomes

$$\mu \tanh \mu = \alpha \tanh \alpha,$$

so that $\mu = \alpha$ and $E \propto -\mu^2$, which means that the state with this energy would be a bound state. However, this solution is physically impossible, since the antisymmetrical wavefunction (20b) vanishes identically when $k = 2i\alpha/l \equiv i\gamma$.

V. NETWORK WITH SINGLE BOUND STATE AT EACH NODE

Consider again the Schrödinger equation (15a) with the potential (15c) which leads to a bound state. Also, consider a network such that along each bond (15) is valid when x is the distance from a node of interest in the direction of another node which is connected to it by the bond. Then,

$$\phi(x) = a[\cos(kx + \delta) - (\gamma/k)\sin(kx + \delta)\tanh \gamma x] = a\{\cos[\delta + kx + \theta(x)]\}/\cos\theta(x), \quad (28)$$

where we define $\theta(x)$ by

$$\tan\theta(x) = (\gamma/k)\tanh\gamma x, \quad \text{with} \quad -\pi < \theta(x) < \pi. \quad (29)$$

Also,

$$\frac{d\theta(x)}{dx} = \left(\frac{\gamma^2}{k^2}\right)\cos^2\theta(x)\,\mathrm{sech}^2\,\gamma x, \quad (30)$$

so that

$$\phi'(x) = -ak\left[\frac{\sin[\delta + kx + \theta(x)]}{\cos\theta(x)} + \left(\frac{\gamma}{k}\right)^2\mathrm{sech}^2\,\gamma x\sin(\delta + kx)\right]. \quad (31)$$

We find from these equations that, since $\theta(0) = 0$,

$$\phi(0) = a\cos\delta, \quad (32a)$$

$$\phi'(0)/\phi(0) = -k[1 + (\gamma/k)^2]\tan\delta. \quad (32b)$$

If l is our bond length and we define $\theta(\frac{1}{2}l) \equiv \theta$, then we have $\theta(-\frac{1}{2}l) = -\theta$. Also,

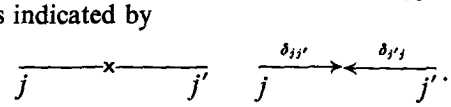
$$\phi(\pm\tfrac{1}{2}l) = \phi(0)[\cos(\theta + \tfrac{1}{2}kl \pm \delta)]/\cos\delta, \quad (33)$$

$$\phi'(\pm\tfrac{1}{2}l) = -k\phi(0)\left(\frac{\sin(\delta \pm \frac{1}{2}lk \pm \theta)}{\cos\theta\cos\delta} + \left(\frac{\gamma}{k}\right)^2\sec\delta\,\mathrm{sech}^2(\tfrac{1}{2}\gamma l)\sin(\delta \pm \tfrac{1}{2}kl)\right). \quad (34)$$

We must now work on the network aspects of our problem. Let j_α be the node points connected to node point j. Then, from (4) and (32b),

$$\sum_\alpha \tan\delta_{jj_\alpha} = 0, \quad (35)$$

where the sum extends over all α which represent nodes connected to j. If we take a typical pair (j, j'), we specify the center of the bond connecting j and j' by an × as indicated by



Here $\delta_{jj'}$ represents the phase factor going from j to j' and $\delta_{j'j}$, then from j' to j. The wavefunction associated with point j must connect at × in a continuous manner with that associated with j'. The two derivatives must also be continuous at that point. We use the notation $\phi[j]$ to represent the wavefunction at node point j. Then at × we have

$$\phi(\tfrac{1}{2}l) = \phi[j]\cos(\theta + \tfrac{1}{2}kl + \delta_{jj'})/\cos\delta_{jj'}, \quad (36a)$$

$$\phi(-\tfrac{1}{2}l) = \phi[j']\cos(\theta + \tfrac{1}{2}kl - \delta_{j'j})/\cos\delta_{j'j}. \quad (36b)$$

Equating these two, we derive

$$\phi[j](c - s\tan\delta_{jj'}) - \phi[j'](c + s\tan\delta_{j'j}) = 0, \quad (37a)$$

where

$$c = \cos(\tfrac{1}{2}kl + \theta) \quad \text{and} \quad s = \sin(\tfrac{1}{2}kl + \theta). \quad (37b)$$

Similarly, if one equates the two derivatives at ×, he finds

$$\phi[j][(c + u)\tan\delta_{jj'} + s + v] - \phi[j'][(c + u)\tan\delta_{j'j} - s - v] = 0. \quad (38)$$

The phase factor $\delta_{j'j}$ can be eliminated from (37) and (38) to obtain the following relationship between $\phi[j]$ and $\phi[j']$:

$$[2s(c + u)\tan\delta_{jj'} + s^2 - c^2 + sv - cu]\phi[j] + (1 + cu + sv)\phi[j'] = 0, \quad (38')$$

where

$$u = (\gamma/k)^2\cos\theta\,\mathrm{sech}^2\,\tfrac{1}{2}\gamma l\cos\tfrac{1}{2}lk, \quad (39a)$$

$$v = (\gamma/k)^2\cos\theta\,\mathrm{sech}^2\,\tfrac{1}{2}\gamma l\sin\tfrac{1}{2}lk. \quad (39b)$$

If we set $j' \equiv j_\alpha$, sum over all j_α points which are attached by bonds to j, and employ (35), we find

$$n_j\left(\frac{c^2 - s^2 + cu - sv}{1 + cu + sv}\right)\phi[j] = \sum_\alpha \phi[j_\alpha], \quad (40)$$

where n_j is the number of bonds connected to node point j. This is the generalization of (14).

After some algebra we find that

$$(c^2 - s^2 + cu - sv)/(1 + cu - sv) = [1 + (\gamma/k)^2]^{-1}\{[1 + (\gamma/k)^2(1 - 2\tanh^2\tfrac{1}{2}\gamma l)]\cos kl - (\gamma/k)[2 + (\gamma/k)^2\,\mathrm{sech}^2\,\tfrac{1}{2}\gamma l]\sin kl\tanh\tfrac{1}{2}\gamma l\} \equiv F. \quad (41)$$

If $\gamma = 0$, this reduces to the free-electron case, $\cos kl$ [see Eq. (14)]. Bound states correspond to a purely imaginary k. Let

$$\epsilon = ik/\gamma \quad \text{and} \quad \alpha = \tfrac{1}{2}\gamma l. \quad (42)$$

Then,

$$F(\epsilon, \alpha) = (c^2 - s^2 + cu - sv)/(1 + cu + sv) = [\epsilon(\epsilon^2 - 1)]^{-1}[\epsilon(\epsilon^2 - 1 + 2\tanh^2\alpha)\cosh 2\alpha\epsilon - (2\epsilon^2 - \mathrm{sech}^2\,\alpha)\tanh\alpha\sinh 2\alpha\epsilon]. \quad (43)$$

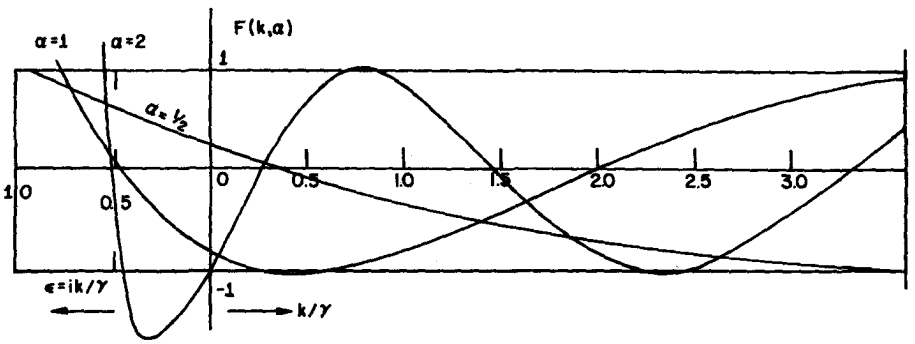


FIG. 6. Form factor F. The region to the right of the origin corresponds to $k/\gamma \geq 0$ [Eq. (52)]. The region to the left corresponds to purely imaginary k/γ with $\epsilon = ik/\gamma$ [Eq. (43)].

While there are apparent singularities at $\epsilon = 0$ and $\epsilon = 1$, a more careful limit analysis shows that our ratio has the limit values

$$F(0, \alpha) = \operatorname{sech}^2 \alpha - \tanh^2 \alpha - 2\alpha \tanh \alpha \operatorname{sech}^2 \alpha, \quad (44a)$$

$$F(1, \alpha) = \operatorname{sech}^2 \alpha - \alpha \operatorname{sech}^2 \alpha \tanh \alpha. \quad (44b)$$

Now that the *form factor* $F(\epsilon, \alpha)$ is known (see Fig. 6), we can discuss the energy levels in our network as well as the density of states. We start with a linear chain to see how the analysis proceeds in a simple case.

VI. ONE-DIMENSIONAL RING

In the case of a 1-dimensional ring of N particles, the jth is connected to $(j-1)$ and $(j+1)$, while $n_j = 2$ for all j. Then, Eq. (40) becomes

$$2F(\epsilon, \alpha)\phi[j] = \phi[j+1] + \phi[j-1], \quad (45a)$$

$$\phi[j+N] \equiv \phi[j], \quad (45b)$$

from ring arrangement. A possible solution of these equations is

$$\phi[j] = A \exp 2\pi ijs/N, \quad (46)$$

where A is a normalization constant. Then, ϵ is a solution of

$$F(\epsilon, \alpha) = \cos(2\pi s/N), \quad (47)$$

while

$$(s/N) = (2\pi)^{-1} \cos^{-1}[F(\epsilon, \alpha)]. \quad (48)$$

The function $F(\epsilon, \alpha)$ is plotted in Fig. 6 as a function of ϵ and $E = -\frac{1}{2}\epsilon^2$ (dimensionless energy) for various values of α. Since $-1 \leq \cos(2\pi s/N) \leq 1$, the possible range of the $F(\epsilon, \alpha)$ for bound states is $-1 < F(\epsilon, \alpha) < 1$. The band structure is clear from the figure. The band edges correspond to the value of ϵ which makes $F = \pm 1$. These values of ϵ and $-\frac{1}{2}\epsilon^2$ are plotted in Fig. 7. If $\phi[j]$ is known, one can find the value of $\phi(ja + x)$ from (28) for points along our network such that $0 < x < a$.

The density of states is, by definition,

$$G(E) = \frac{d(s/N)}{dE} = \frac{d(s/N)}{d\epsilon}\frac{d\epsilon}{dE} = -\frac{1}{\epsilon}\frac{d(s/N)}{dF(\epsilon, \alpha)}\frac{dF(\epsilon, \alpha)}{d\epsilon}$$

$$= \frac{1}{2\pi}\frac{1}{\epsilon}\frac{1}{[1 - F^2(\epsilon, \alpha)]^{\frac{1}{2}}}\frac{dF(\epsilon, \alpha)}{d\epsilon}, \quad (49)$$

where

$$\frac{dF(\epsilon, \alpha)}{d\epsilon} = -\frac{2 \tanh \alpha \cosh 2\alpha\epsilon}{\epsilon(\epsilon^2 - 1)^2} \times [2\epsilon^2 \tanh \alpha + \alpha(\epsilon^2 - 1)(2\epsilon^2 - \operatorname{sech}^2 \alpha)]$$

$$+ \frac{\sinh 2\alpha\epsilon}{\epsilon^2(\epsilon^2 - 1)^2} \times \{[2\epsilon^2(\epsilon^2 + 1) - (3\epsilon^2 - 1)\operatorname{sech}^2 \alpha] \tanh \alpha + 2\alpha\epsilon^2(\epsilon^2 - 1)(\epsilon^2 - 1 + 2\tanh^2 \alpha)\}. \quad (50)$$

All ϵ^2 in (49) and (50) are to be replaced by $-2E$ in the final formula so that $G(E)$ is a function of E rather than of K. The function $dF/d\epsilon$ is plotted in Fig. 8.

In the conduction band, we let

$$\kappa = k/\gamma \quad \text{and} \quad \alpha = \tfrac{1}{2}\gamma l. \quad (51)$$

Then our structure function is

$$F(\kappa, \alpha) = [\kappa(1 + \kappa^2)]^{-1}[\kappa(k^2 + 1 - 2\tanh^2 \alpha)\cos 2\kappa\alpha - (2\kappa^2 + \operatorname{sech}^2 \alpha)\sin 2\kappa\alpha \tanh \alpha]. \quad (52)$$

The density of states in the conduction band, since $E = \frac{1}{2}\kappa^2$, is

$$G(E) = \frac{d(s/N)}{dE} = \frac{d(s/N)}{d\kappa}\frac{d\kappa}{dE} = \frac{1}{\kappa}\frac{d(s/N)}{dF(\kappa, \alpha)}\frac{dF(\kappa, \alpha)}{d\kappa}$$

$$= \frac{1}{2\pi\kappa}[1 - F^2(\kappa, \alpha)]^{-\frac{1}{2}}\frac{dF(\kappa, \alpha)}{d\kappa}, \quad \kappa > 0, \quad (53)$$

where now

$$\frac{dF(\kappa, \alpha)}{d\kappa} = 2[2\kappa^2 \tanh \alpha - \alpha(1 + \kappa^2)(2\kappa^2 + \operatorname{sech}^2 \alpha)] \times \frac{\tanh \alpha \cos 2\kappa\alpha}{\kappa(1 + \kappa^2)^2} + \frac{\sin 2\alpha\kappa}{\kappa^2(1 + \kappa^2)^2} \times \{[2\kappa^2(\kappa^2 - 1) + (3\kappa^2 + 1)\operatorname{sech}^2 \alpha] \tanh \alpha - 2\alpha\kappa^2(1 + \kappa^2)(\kappa^2 + 1 - 2\tanh^2 \alpha)\}. \quad (54)$$

This function is plotted in Fig. 8.

We can now use Eqs. (49) and (53) to obtain the density of states at all energies. This is plotted in Fig. 10 for several values of α.

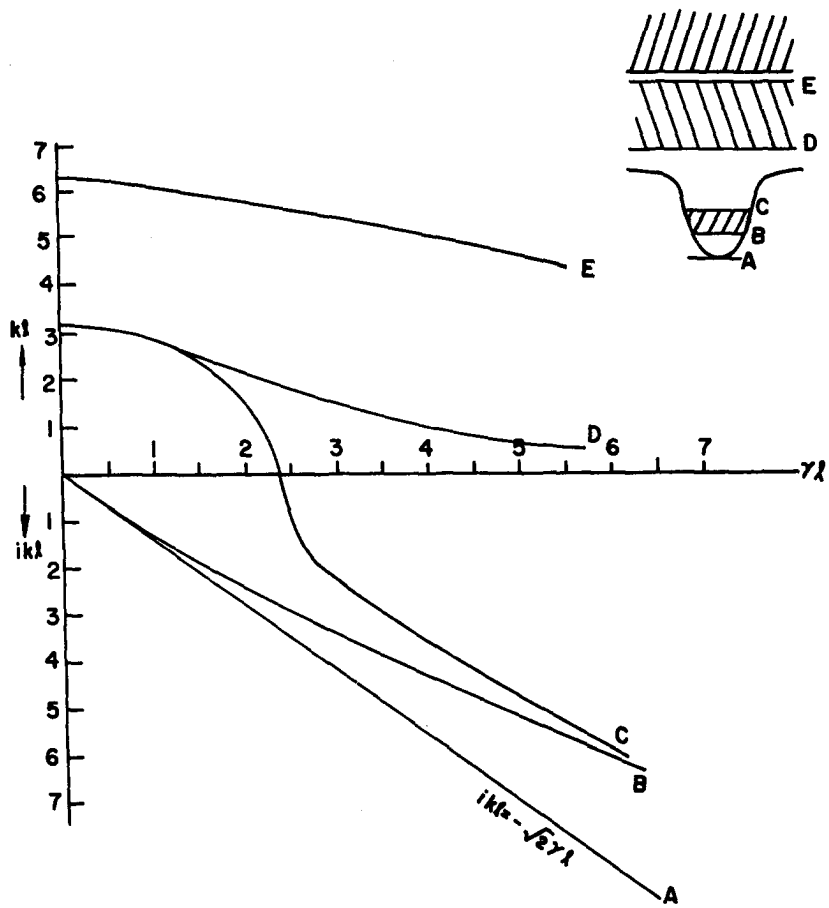


FIG. 7. Location of band edges as a function of potential-well depth.

Curves of E as a function of $\theta = (2\pi s/N)$ are plotted in Fig. 9. Such curves are usually known as E vs k curves. However, since we have used k for another purpose, we call them E vs θ curves.

Now let us refer back to Fig. 6. The points ϵ where the graphs of $F(\epsilon, \alpha)$ intersect the lines $F = \pm 1$ represent the location of the band edges. In Fig. 7, curve B represents the intersection with $F = 1$ and represents the bottom of the band, while C is the intersection with -1 and represents the top of the band. When $\gamma l < 2.4$, there is no intersection with $F = -1$, and there is no gap between C and the energy $E = 0$.

VII. TWO-DIMENSIONAL SQUARE LATTICE

We characterize a lattice point on a 2-dimensional square lattice by $j \equiv (j_1, j_2)$, so that (40) becomes

$$4F(\epsilon, \alpha)\phi(j_1, j_2) = \phi[j_1 + 1, j_2] + \phi[j_1 - 1, j_2] + \phi[j_1, j_2 + 1] + \phi[j_1, j_2 - 1], \quad (55a)$$

while periodic boundary conditions yield

$$\phi[j_1 + N, j_2] = \phi[j_1, j_2 + N] \equiv \phi[j_1, j_2]. \quad (55b)$$

A possible solution of these equations is

$$\phi[j_1, j_2] = A \exp [2\pi i(j_1 s_1 + j_2 s_2)/N], \quad (56)$$

where A is a normalization constant. Then ϵ is a solution of

$$2F(\epsilon, \alpha) = \cos (2\pi s_1/N) + \cos (2\pi s_2/N), \quad s_1, s_2 = 0, 1, 2, \cdots, N - 1. \quad (57)$$

The first step in the determination of the density of states is to find the distribution function of the sum of the two cosines on the right as s_1 and s_2 range from 0 to $N - 1$. Since it is just as easy to formulate the problem with n cosines as with 2, we do so.

Let

$$x_n = \cos (2\pi s_1/N) + \cos (2\pi s_2/N) + \cdots + \cos (2\pi s_n/N) \quad (58)$$

and

$$P_n(x)\, dx = \text{Prob}\,(x < x_n < x + dx). \quad (59)$$

Then the distribution function of ϵ is

$$F(\epsilon) = P_n(nF(\epsilon, \alpha)) \frac{dF}{d\epsilon}, \quad (60)$$

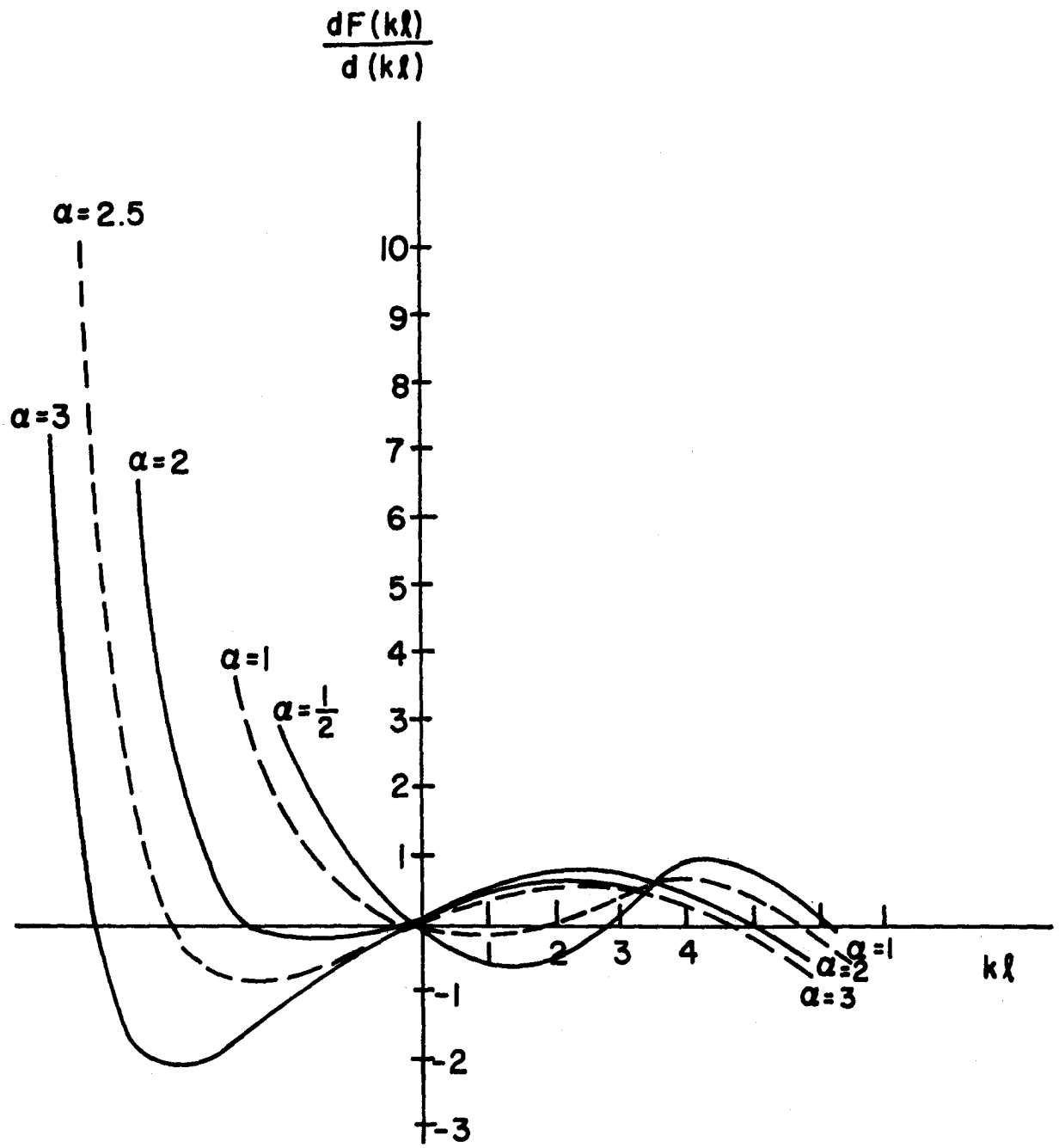


FIG. 8. Derivative of form factor as a function of lk.

so that that of E (the density of states) is, in the band of bound states, $d\epsilon/dE = 1/\epsilon$,

$$G_n(E) = -(1/\epsilon)P_n(nF(\epsilon, \alpha))\frac{dF}{d\epsilon}, \tag{61a}$$

where Eqs. (59) and (60) are to be employed and ϵ^2 is to be replaced by $-2E$ after all formulas are inserted in $G(E)$. In the conduction band,

$$G_n(E) = (1/k)P_n(nF(k, \alpha))\frac{dF}{dk}. \tag{61b}$$

It remains to determine $P_2(x)$, which we do by first finding the formula for $P_n(x)$. Let

$$f_n(\alpha) = \langle \exp i\alpha X_n \rangle_{\text{av}} \tag{62}$$

be the characteristic function of X_n; i.e., the function whose Fourier transform is $P_n(x)$:

$$P_n(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f_n(\alpha)e^{-i\alpha x}\, d\alpha. \tag{63}$$

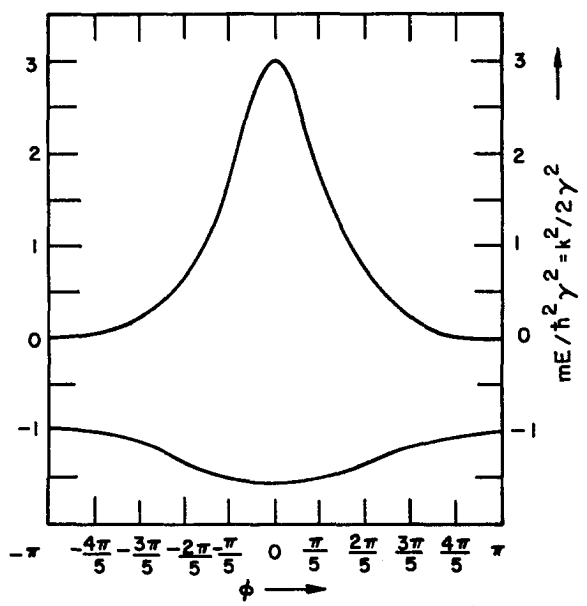


FIG. 9. Variation of dimensionless energy $mE/\hbar^2\gamma^2 = k^2/2\gamma^2$, with ϕ for 1-dimensional chain. The curves correspond to $\alpha = \frac{1}{2}\gamma l = 2$.

Then,

$$f_n(\alpha) = \frac{1}{N^n}\sum_{s_1=0}^{N-1}\cdots\sum_{s_n=0}^{N-1}\exp i\alpha\left[\cos\left(\frac{2\pi s_1}{N}\right) + \cdots + \cos\left(\frac{2\pi s_n}{N}\right)\right] = \left\{\frac{1}{N}\sum_{s=0}^{N-1}\exp\left[i\alpha\cos\left(\frac{2\pi s}{N}\right)\right]\right\}^n. \quad (64)$$

But

$$\lim_{N\to\infty}\frac{1}{N}\sum_{s=0}^{N-1}\exp\left[i\alpha\cos\left(\frac{2\pi s}{N}\right)\right] = \frac{1}{2\pi}\int_0^{2\pi} e^{i\alpha\cos\theta}\,d\theta = J_0(\alpha), \quad (65)$$

where $J_0(\alpha)$ is the zero-order Bessel function. Hence,[13]

$$P_n(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\alpha x}[J_0(\alpha)]^n\,d\alpha. \quad (66)$$

The value of this integral is, for $n = 1, 2$,

$$P_1(x) = (1-x^2)^{-\frac{1}{2}}/\pi, \quad \text{if } |x|<1; \quad = 0, \quad \text{if } |x|>1, \quad (67a)$$

$$P_2(x) = \pi^{-2}K\{[1-\tfrac{1}{2}x^2]^{\frac{1}{2}}\}, \quad \text{if } |x|<2; \quad = 0, \quad \text{if } |x|>2, \quad (67b)$$

where $K(z)$ is the complete elliptic integral of the second kind. The explicit formula for $P_3(x)$ involves generalized hypergeometric functions which are not very helpful for calculations. However, various useful expansions are given in the next section. The functions $P_1(x)$, $P_2(x)$, and $P_3(x)$ are plotted in Fig. 11. It is easily seen that, by combining (61) and (66), one obtains the 1-dimensional result (49).

The 2-dimensional density of states is obtained by combining (61) and (67b). Then, in the band of bound states, $\epsilon^2 = -2E$ and

$$G_2(E) = -\left(\frac{1}{2\epsilon\pi^2}\right)K\{1-[F(\epsilon,\alpha)]^2\}^{\frac{1}{2}}\frac{dF(\epsilon,\alpha)}{d\epsilon}, \quad \text{if } |F(\epsilon,\alpha)|<1. \quad (68a)$$

In the conduction band (with $E \geq 0$),

$$G_2(E) = \left(\frac{1}{2\kappa\pi^2}\right)K\{1-[F(\kappa,\alpha)]^2\}^{\frac{1}{2}}\frac{dF(\kappa,\alpha)}{d\kappa}, \quad \text{if } \kappa>0. \quad (68b)$$

The F-functions are given by (43), (50), (52), and (54). The 2-dimensional density of states is plotted in Fig. 12.

The curves of constant energy in (θ_1, θ_2) space with

$$(\theta_1, \theta_2) = (2\pi s_1/N, 2\pi s_2/N) \quad (69)$$

are obtained from examining the curves

$$\tfrac{1}{2}[\cos\theta_1 + \cos\theta_2] = \text{const} = C. \quad (70)$$

The values of the constant range from -1 to $+1$. $C = -1$ corresponds to the origin in Fig. 13 and $+1$ corresponds to the corners $(\pm\pi, \pm\pi)$, $(\pm\pi, \mp\pi)$. The identification of the various values of C with the dimensionless energy $mE/\hbar^2\gamma^2 = \frac{1}{2}k^2/\gamma^2$ is to be made with graphs such as those in Fig. 6. For example, when $\alpha = \frac{1}{2}l\gamma = 2$, $C = 1$ corresponds to $ik/\gamma = 0.56$ so that $\frac{1}{2}k^2/\gamma^2 = -0.16$.

VIII. SIMPLE CUBIC LATTICE

The lattice points of a 3-dimensional simple cubic lattice are expressed as $j \equiv (j_1, j_2, j_3)$, and (40) becomes

$$6F(\epsilon,\alpha)\phi[j_1,j_2,j_3] = \phi[j_1+1,j_2,j_3] + \phi[j_1-1,j_2,j_3] + \phi[j_1,j_2+1,j_3] + \phi[j_1,j_2-1,j_3] + \phi[j_1,j_2j_3+1] + \phi[j_1,j_2,j_3-1], \quad (71)$$

which can be solved with periodic boundary conditions so that

$$\phi[j_1,j_2,j_3] = A\exp 2\pi i(j_1s_1 + j_2s_2 + j_3s_3)/N. \quad (72)$$

Then ϵ is a solution of

$$3F(\epsilon,\alpha) = [\cos(2\pi s_1/N) + \cos(2\pi s_2/N) + \cos(2\pi s_3/N)]. \quad (73)$$

[13] E. W. Montroll, Proc. Berkeley Symp. Math. Stat. Prob. 3, 209 (1956).

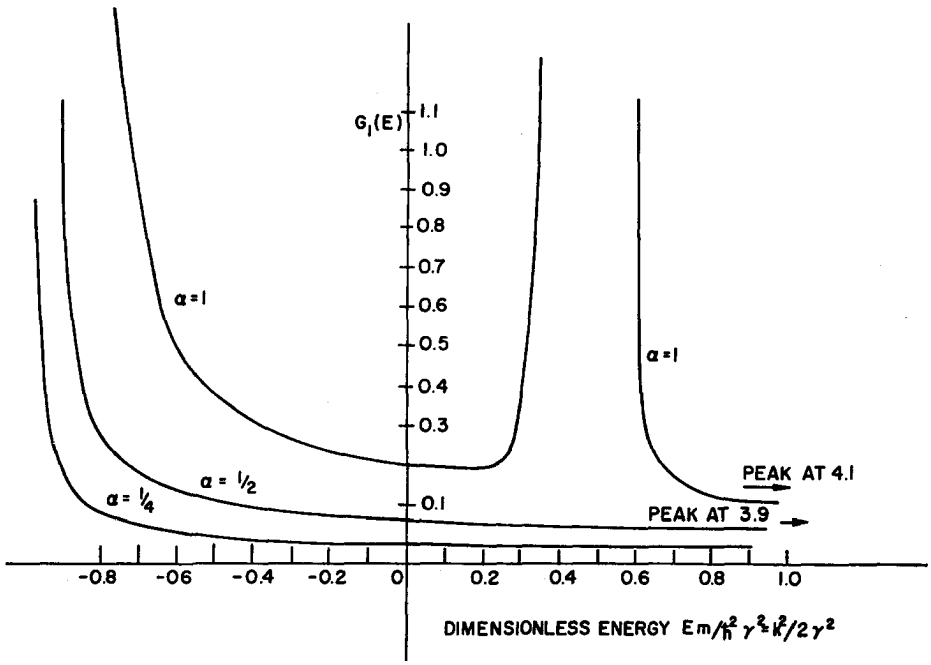


FIG. 10. One-dimensional density of states $G(E)$. Here $\alpha = -\frac{1}{2}l$.

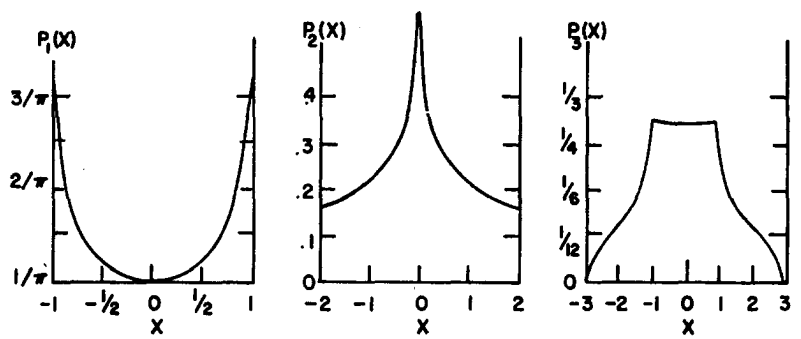


FIG. 11. Distribution of the variable $x_n = \cos\theta_1 + \cos\theta_2 + \cdots + \cos\theta_n$, with all θ uniformly distributed in range $(-\pi, \pi)$.

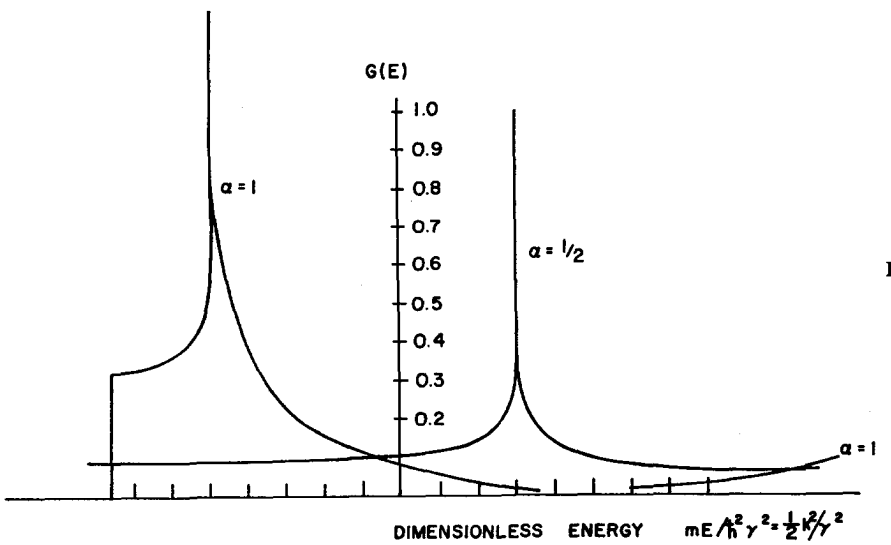


FIG. 12. Two-dimensional density of states.

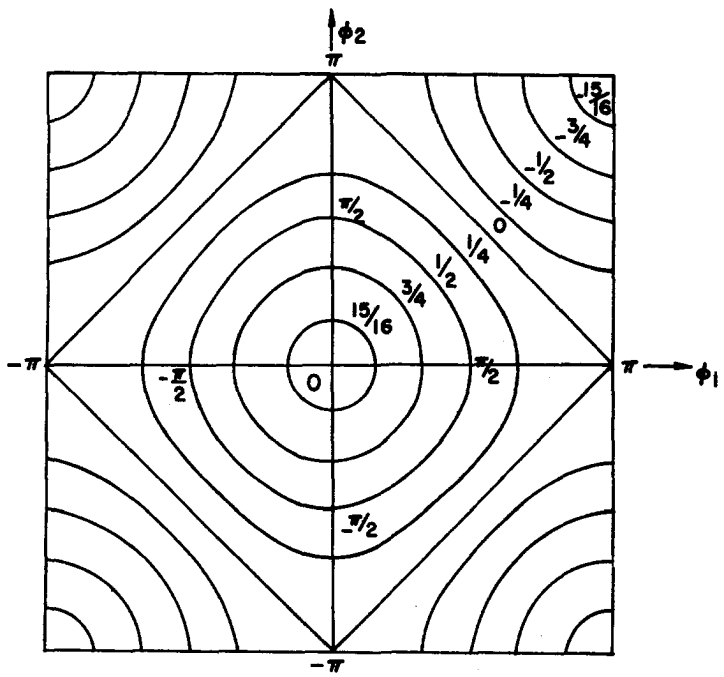


FIG. 13. Curves of constant $\frac{1}{2}(\cos\theta_1 + \cos\theta_2) = C$. The numbers given along the diagonal represent the values of C associated with the curve nearest that number.

The identification of a given value of C can be obtained from curves such as those in Fig. 6. The dimensionless energy is $mE/\hbar^2\gamma^2 = \frac{1}{2}k^2/\gamma^2$. If, for example, we choose $\alpha \equiv \frac{1}{2}\gamma l$ to be 1, then we use that curve in Fig. 6. The value of $\epsilon = 1k/\gamma = 0.46$ is associated with $C = F(\epsilon, \alpha) = 0$. Hence, the curve above with $C = 0$ corresponds to the energy $mE/\hbar^2\gamma^2 = -0.106$.

The distribution $P_3(x)$ does not have a simple analytical expression such as (67). However, its graph is given in Fig. 11. Certain series expansions allow one to calculate $P_3(x)$ with an accuracy of about one part per thousand. First,

$$P_3(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty}[J_0(\alpha)]^3\,d\alpha = \tfrac{2}{3}[\Gamma(\tfrac{5}{6})\Gamma(\tfrac{2}{3})]^{-1} = 0.2853. \quad (74)$$

The function $P_3(x)$ has an infinite slope at the points $x = \pm 1$ as well as $x = \pm 3$. $P_3(x)$ vanishes if $|x| > 3$.

In the neighborhood of $x = 1$ [and similar results exist for $x = -1$ since $P_3(x)$ is an even function],

$$P_3(x) = P_3(1) - \frac{1}{2\pi}\int_{-\infty}^{\infty}(e^{-i\alpha} - e^{-i\alpha x})[J_0(\alpha)]^3\,d\alpha. \quad (75)$$

Since the quantity in the parenthesis has the form

$$i(x-1)\alpha + \tfrac{1}{2}\alpha^2(x^2-1) + \cdots$$

when α is small, we see that the small-α range does not contribute significantly to the integral. However, when α is large, the two exponentials do not tend to cancel each other, so that the range of large α can be expected to contribute more to the integral. On this basis, we follow Baroody[14] and introduce the expansion

$$J_0(\alpha) = (2/\pi\alpha)^{\frac{1}{2}}[\cos(\alpha - \tfrac{1}{4}\pi) + (8\alpha)^{-1}\sin(\alpha - \tfrac{1}{4}\pi) + \cdots] \quad (76)$$

into (75). Then a number of useful expansions for $P_3(x)$ result. As $x \to 1$ from above,

$$P_3(x) = P_3(1) - \frac{3}{\pi^2}[\tfrac{1}{2}(x-1)]^{\frac{1}{2}} + \frac{1}{8\pi^2}(2^{\frac{5}{2}} - 5)(x-1) + \frac{1}{2\pi^2}[\tfrac{1}{2}(x-1)]^{\frac{3}{2}} + \cdots, \quad (77a)$$

while, as $x \to -1$ from below,

$$P_3(x) = P_3(-1) - \frac{3}{\pi^2}[-\tfrac{1}{2}(1+x)]^{\frac{1}{2}} - \frac{1}{8\pi^2}(2^{\frac{5}{2}} - 5)(1+x) + \frac{1}{2\pi^2}[-\tfrac{1}{2}(1+x)]^{\frac{3}{2}} + \cdots. \quad (77b)$$

In the interval $-1 < x < 1$,

$$P_3(x) = P_3(-1) - \frac{1}{2\pi^2}(3 + 2^{\frac{5}{2}}) + \frac{3^{\frac{1}{2}}}{4\pi^2}\{[\tfrac{1}{2}(1+\tfrac{1}{3}x)]^{\frac{1}{2}}(7+x) + [\tfrac{1}{2}(1-\tfrac{1}{3}x)]^{\frac{1}{2}}(7-x)\} + \cdots \quad (77c)$$

[14] E. M. Baroody, J. Math. Phys. 10, 475 (1969).

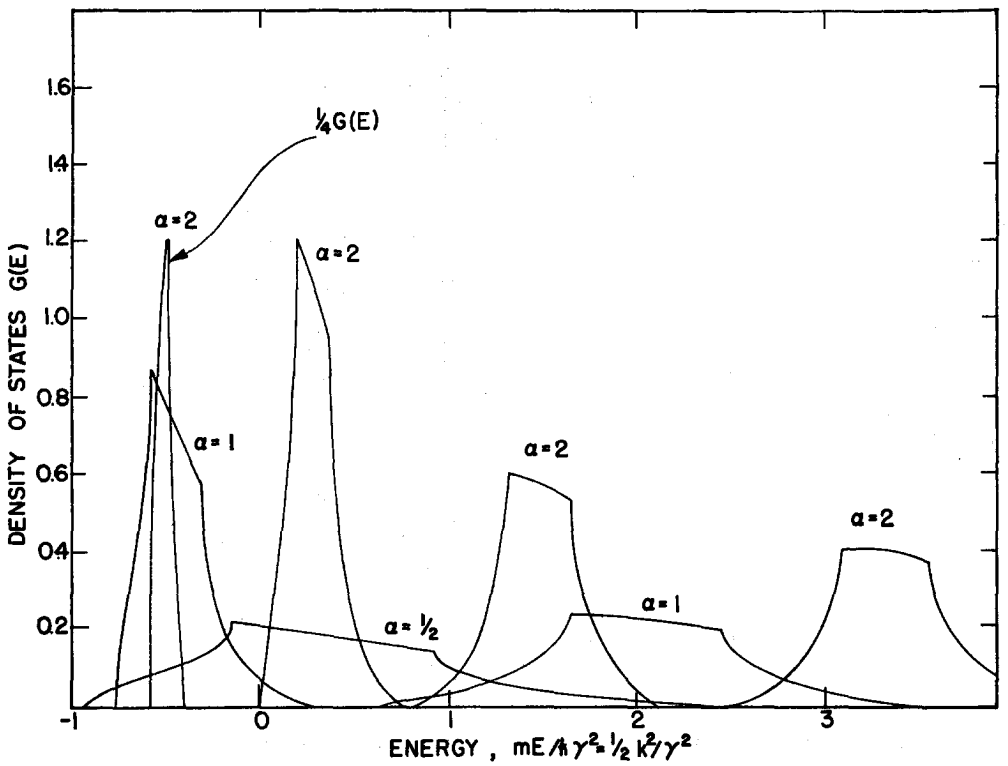


FIG. 14. Density of states of simple cubic lattice for several values of $\alpha = \frac{1}{2}\gamma l$. Note that for the lowest energy band of the case $\sigma = 2$, we have plotted $\frac{1}{4}G(E)$ rather than $G(E)$.

and, as $x \to -3$,

$$P_3(x) = \frac{1}{\pi^2}[\tfrac{1}{2}(3+x)]^{\frac{1}{2}} \times \left\{4\left[1 - \tfrac{1}{2}d_1\left(1+\frac{x}{3}\right) - \tfrac{1}{4}d_2\left(1+\frac{x}{3}\right)^2 - \tfrac{1}{8}d_3\left(1+\frac{x}{3}\right)^3 \cdots\right] - 3[-\tfrac{1}{2}(1+x)]^{\frac{1}{2}}\right\}. \tag{77d}$$

The coefficients $d_1, d_2, \cdots$ are exhibited in Table I.

TABLE I. Coefficient d_j of expansion[14] of $P_3(x)$ as given in Eq. (77d).

j	d_j	j	d_j
1	$\frac{3}{4}$	7	0.90
2	$\frac{9}{40}$	8	2.20
3	$\frac{153}{1120}$	9	5.5
4	$\frac{81}{640}$	10	13.7
5	0.1861	11	34.3
6	0.382	12	87.0

The quantity $P_3(1) \equiv P_3(-1)$ which appears above can be obtained from (77c) by setting $x = 0$ and using (74). As a check of consistency it can also be found from (77d). The two values check very well and, indeed,

$$P_3(1) = P_3(-1) = 0.289. \tag{78}$$

To obtain an expression for $P_3(x)$ as $x \to +3$, we replace x by $-x$ in (77d).

By referring back to (61), we can obtain the density of states of a simple cubic lattice. This is plotted in Fig. 14. The variables analogous to (58) for face-centered and body-centered cubic lattices are, respectively,

$$x_{Fc} = c_1c_2 + c_2c_3 + c_3c_1 \quad \text{and} \quad x_{Bc} = c_1c_2c_3.$$

The distribution functions of these variables has been discussed by Jelitto.[15] Their use in the construction of the density of states will be presented elsewhere.

IX. DIATOMIC LATTICES

The above discussion can easily be generalized to diatomic lattices. Let us suppose that there is a set of lattice sites which we identify as sites of kind "1" and that the remaining ones are of kind "2." Furthermore, let the value of the parameter γ associated with a site of kind ν be γ_ν.

The first step in the analysis is to relate the wave-function at node point j to that of a nearest-neighbor node point j'. As was done in Sec. V, we must choose the various integration constants so that the wave-function and its derivative along the line connecting j and j' are continuous at the point $\frac{1}{2}l$. We define the

[15] R. J. Jelitto, J. Phys. Chem. Solids 30, 609 (1969).

wavefunction at j to be $\phi_\nu[j]$, if j is a site of type ν, and $\phi_\nu(x)$ to be the wavefunction at a distance x from node j on the way to node j'.

One finds the generalization of (36) to be

$$\phi_1(\tfrac{1}{2}l) = \phi_1[j] \sec \delta_{jj'} \cos [\theta_1(\tfrac{1}{2}l) + \tfrac{1}{2}kl + \delta_{jj'}], \tag{79a}$$

$$\phi_2(-\tfrac{1}{2}l) = \phi_2[j'] \sec \delta_{j'j} \cos [\theta_2(\tfrac{1}{2}l) + \tfrac{1}{2}kl - \delta_{j'j}]. \tag{79b}$$

Since the network wavefunction is continuous at the midpoint between nodes j and j', we find the generalization of (37) to be

$$\phi_1[j](c_1 - s_1 \tan \delta_{jj'}) - \phi_2[j'](c_2 + s_2 \tan \delta_{j'j}) = 0, \tag{80a}$$

where

$$c_\nu = \cos [\tfrac{1}{2}kl + \theta_\nu(\tfrac{1}{2}l)], \quad s_\nu = \sin [\tfrac{1}{2}kl + \theta_\nu(\tfrac{1}{2}l)], \tag{80b}$$

$$\tan \theta_\nu(x) = (\gamma_\nu/k) \tanh x\gamma_\nu. \tag{80c}$$

Furthermore, since

$$\phi_\alpha'(x) = -k\phi_2(0)\left[\frac{\sin [\delta + kx + \theta_\alpha(x)]}{\cos \theta_\alpha(x)} + \left(\frac{\gamma_\alpha}{k}\right)^2 \operatorname{sech}^2 x\gamma_\alpha \sin (\delta + kx)\right], \tag{81}$$

we find the following generalization of (38) after equating $\phi_1'(\frac{1}{2}l)$ to $\phi_2'(-\frac{1}{2}l)$:

$$(\sec \theta_1)\phi_1[j][(c_1 + u_1) \tan \delta_{jj'} + s_1 + v_1] - (\sec \theta_2)\phi_2[j'][(c_2 + u_2) \tan \delta_{j'j} - s_2v_2] = 0, \tag{82a}$$

where

$$u_\alpha = (\gamma_\alpha/k)^2 \cos \theta_\alpha \operatorname{sech}^2 \tfrac{1}{2}l\gamma \cos \tfrac{1}{2}kl,$$
$$v_\alpha = (\gamma_\alpha/k)^2 \cos \theta_\alpha \operatorname{sech}^2 \tfrac{1}{2}l\gamma \sin \tfrac{1}{2}kl, \tag{82b}$$

$$\theta_\nu \equiv \theta_\nu(\tfrac{1}{2}l). \tag{82c}$$

The quantity $\tan \delta_{j'j}$ can then be eliminated from (80a) and (82a) to yield the generalization of (38′):

$$\phi_1[j]\left[\frac{c_1(c_2 + u_2)}{\cos \theta_2} - \frac{s_2(s_1 + v_1)}{\cos \theta_1} - (\cdots) \tan \delta_{jj'}\right] = \frac{1 + c_2u_2 + s_2v_2}{\cos \theta_2} \phi_2[j'].$$

The quantity $(\cdots)$ is irrelevant because Eq. (4) implies that, after summing j' over all nearest neighbors to j,

$$F_1\phi_1[j] = \sum_\alpha \phi_2[j_\alpha], \quad \text{if } j \text{ is a type-1 site,} \tag{83a}$$

$$F_2\phi_2[j] = \sum_\alpha \phi_1[j_\alpha], \quad \text{if } j \text{ is a type-2 site,} \tag{83b}$$

where

$$F_1 = \frac{n_j \cos \theta_2}{1 + c_2u_2 + s_2u_2}\left(\frac{c_1(c_2 + u_2)}{\cos \theta_2} - \frac{s_2(s_1 + v_1)}{\cos \theta_1}\right) \tag{84}$$

and F_2 has the same form with subscripts 1 and 2 interchanged.

Now define[16]

$$F_1^{\frac{1}{2}}\phi_1[j] \equiv \psi[j], \quad \text{if } j \text{ is a type-1 site,} \tag{85a}$$

$$F_2^{\frac{1}{2}}\phi_2[j] \equiv \psi[j], \quad \text{if } j \text{ is a type-2 site.} \tag{85b}$$

Then, for *all* node points j,

$$(F_1F_2)^{\frac{1}{2}}\psi[j] = \sum_\alpha \psi[j_\alpha]. \tag{86}$$

This has the same mathematical form as (40) if we identify (F_1F_2) with F.

X. SOME EXTENSIONS OF THE MODEL

There are various ways in which our model can be made more realistic. While these will be discussed in detail elsewhere, we wish to identify a few of them here.

First, we introduce the influence of more distant neighbors. Let us suppose that the node points to which a given node point are connected can be divided into a number of classes: those a distance l_1 away, those l_2 away, etc. If node j is a distance l_μ from j', then the generalization of (38′) is

$$(\tan \delta_{jj'} + G_\mu)\phi[j] = H_\mu\phi[j'], \tag{87}$$

where

$$G_\mu = (c_\mu^2 - s_\mu^2 + c_\mu u_\mu - s_\mu v_\mu)/2s_\mu(c_\mu + u_\mu), \tag{88a}$$

$$H_\mu = (1 + c_\mu u_\mu + s_\mu v_\mu)/2s_\mu(c_\mu + u_\mu), \tag{88b}$$

$$c_\mu = \cos (\tfrac{1}{2}kl_\mu + \theta_\mu), \quad \text{with} \quad \theta_\mu = \theta(\tfrac{1}{4}l_\mu), \text{ etc.} \tag{88c}$$

Then, from (35) after summing over all neighbors j' of j,

$$\phi[j] \sum_\mu n_\mu G_\mu = \sum_\mu H_\mu \sum_{\alpha\mu} \phi[j_{\alpha\mu}^\mu], \tag{89}$$

where j^μ represents all neighbors to j of class μ. The summation extends over all members α_μ of class and over all classes of neighbors. Also, n_μ is the number of neighbors of j which are of class μ.

Let us apply (89) to a square lattice (see Fig. 14). Then

$$\begin{aligned} &f(G_1 + G_2)\phi[j_1, j_2] \\ &= H_1(\phi[j_1, j_2 - 1] + \phi[j_1, j_2 + 1] + \phi[j_1 - 1, j_2] \\ &\quad + \phi[j_1 + 1, j_2]) + H_2(\phi[j_1 + 1, j_2 + 1] \\ &\quad + \phi[j_1 + 1, j_2 - 1] + \phi[j_1 - 1, j_2 + 1] \\ &\quad + \phi[j_1 - 1, j_2 - 1]). \end{aligned} \tag{90}$$

[16] A. A. Maradudin, P. Mazur, E. W. Montroll, and G. H. Weiss, Rev. Mod. Phys. 30, 175 (1958).

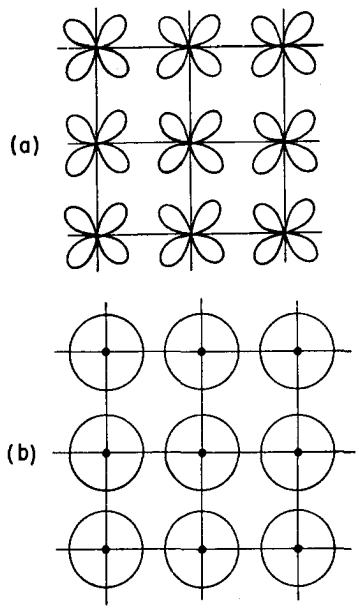


FIG. 15. Examples of decorated lattices.

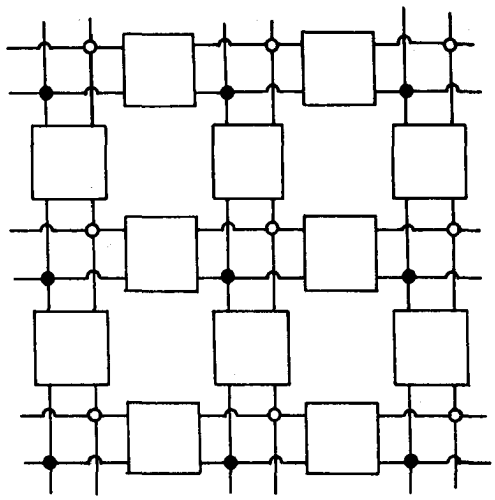


FIG. 16. Lattice on which two end states are possible and on which spin flipping can occur between lattice points.

Following (72), the energy levels are to be found from

$$2(G_1 + G_2) = H_1(\cos\phi_1 + \cos\phi_2) + 2H_2 \cos\phi_1 \cos\phi_2. \quad (91)$$

The corresponding expression for a simple cubic lattice is

$$3(G_1 + 2B_2) = H_1(\cos\phi_1 + \cos\phi_2 + \cos\phi_3) + 2H_2(\cos\phi_1 \cos\phi_2 + \cos\phi_2 \cos\phi_3 + \cos\phi_3 \cos\phi_1). \quad (92)$$

If we set $l_1 \equiv l$, then $l_2 = 2^{\frac{1}{2}}l$.

The main difference between Eq. (40) and Eqs. (91) and (92) is that in (40) the geometrical character of the lattice is reflected on the right-hand side of the equation and the force law on the left. This separation is no longer possible when second neighbors are included. As will be discussed elsewhere, the determination of the density of states is more difficult in this case.

Now, suppose we wish to return to the nearest-neighbor model, but also wish to give more variety to the electron density distribution in the neighborhood of each node point. This can be accomplished by decorating the lattice with bonds which start and end at the same node point, as exhibited in Fig. 15(a) or by changing the local connectivity as shown in Fig. 15(b).

Finally, if one wishes to give each electron a spin, it can be done through the introduction of two node points at each atomic site, the spin-up node point and the spin-down node point with spin-up and spin-down tracks connected to the appropriate node point, as in Fig. 16. A "junction box" can be placed between each neighboring pair of atoms such that switching from a spin-up to a spin-down track can be achieved in the junction box. Since the junction box relates two inputs to two outputs, it can be described by a 2×2 matrix. The wavefunctions at each node point has two components, the spin-up and spin-down components.

ACKNOWLEDGMENTS

The author wishes to thank Dr. R. G. Mills and Mr. C. H. Wu for performing the calculations on which some of the figures are based, as well as for numerous discussions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Variational Methods in Potential Scattering

HANS MORAWITZ
IBM Research Laboratory, San José, California 95114

(Received 26 September 1968; Revised Manuscript Received 9 May 1969)

A comparison of a variety of standard variational methods with a new method due to Harris is made with a view towards selection of the optimal method to be used in realistic many-body scattering calculations. Numerical results for two short-range potentials, the attractive exponential and the attractive Yukawa potential, are given and compared with exact results obtained analytically or by direct numerical integration. It is demonstrated by calculation that the source of an anomaly observed in earlier studies by Schwartz is not due to the attributed reason and, furthermore, we find the Kohn method significantly more accurate at the Harris eigenvalues than any of the other methods. Therefore, we propose the use of the Kohn method at the eigenvalues of the matrix Hamiltonian in the trial function subspace as the optimal method to be used in variational scattering calculations.

I. INTRODUCTION

Although scattering theory has been one of the most active areas in physics for many years and, consequently, is a well-mined field of research,[1] high precision computations of scattering parameters (phase shifts) are rare in atomic and molecular physics when compared with bound-state computations. The reason is twofold: the many-body nature of a problem like electron-atom scattering presents formidable obstacles in the positive energy region (we note that apart from $e^{\pm}$-H and e–He^{+} scattering, few other problems have been tackled quantitatively, i.e., beyond tentative applications of various theoretical schemes) and, secondly, the scattering data have only recently started to reach accuracies warranting a more detailed push for numerical precision on the theoretical and computational front.

In view of the existence of large scale computers and groups expert in their use, it is to be expected and, in fact, is already partly a fact, that massive computations in the scattering region (elastic as well as inelastic) are going to supplement the great amount of computational work done in bound-state applications, which will be of great use and value in that task. In view of this trend and in analogy to the bound-state problem, it is interesting to consider and compare the status of existing variational procedures, so as to get an idea about optimal methods as well as accuracy, speed of convergence, and size of basis sets necessary for high precision results, which we arbitrarily take to mean agreement to one part in 10^6 with exact results in the variational determination of the phaseshift.

In the following article we study potential scattering, i.e., the one-body problem, and, in particular, choose two well-behaved short-range potentials, the attractive exponential $V(r) = -e^{-r}$ and the attractive Yukawa potential $V(r) = -e^{-r}/r$, to calculate s-wave scattering from them. Because of the dearth of exactly soluble potentials, much effort has gone towards finding approximate solutions for the phase shifts.[2–7] Our interest lies in the usefulness of variational methods and the development of optimal methods for the determination of the phase shifts by variational means. In particular, a difficulty with the standard Hulthén–Kohn formulation,[2,3] noted by C. Schwartz,[8] is discussed, and it is shown how a recent approach proposed by Harris[9] points a way toward a new look at the problem, closely related to the bound-state problem[10] and resonance scattering.[11,12] For potential theory, of course, we can always calculate the phase shifts directly by numerical integration,[7] which is a straightforward task with available high-speed computers. The motivation of this study is, however, the exploration of an optimal numerical approach to be used on the much more formidable problem of doing good calculations on genuine many-body systems, such as the scattering of electrons from atoms and molecules, as well as nucleon–nucleus scattering.

The choice of the two potentials considered here is to be attributed to their importance in atomic and

[1] Any textbook on scattering theory, e.g., N. F. Mott and H. S. Massey, *The Theory of Atomic Collisions, International Series of Monographs in Physics* (Oxford University Press, London, 1965), 3rd ed.

[2] L. Hulthén, Kgli. Fysiograf. Sallskap. Lund, Forh. **14**, 1 (1944).
[3] W. Kohn, Phys. Rev. **74**, 1763 (1948).
[4] J. Schwinger (unpublished lectures, 1947).
[5] T. Kato, Progr. Theoret. Phys. (Kyoto) **6**, 295 (1951).
[6] L. Spruch, in *Lectures in Theoretical Physics*, W. E. Brittin *et al.*, Eds. (Interscience Publishers Inc., New York, 1961), Vol. IV; Y. Hahn, T. O. Malley, L. Spruch, Phys. Rev. **118**, 184 (1960); **130**, 381 (1963); **134**, B911 (1964).
[7] F. Calogero, *Variable Phase Approach to Potential Scattering* (Academic Press Inc., New York, 1967).
[8] C. Schwartz, Ann. Phys. (N.Y.) **16**, 36 (1961).
[9] F. Harris, Phys. Rev. Letters **19**, 173 (1967).
[10] J. S. Slater, *Quantum Theory of Atomic Structures* (McGraw-Hill Book Co., Inc., New York, 1960).
[11] E. P. Wigner, Phys. Rev. **70**, 15, 606 (1946).
[12] H. Feshbach, Ann. Phys. (N.Y.) **19**, 287 (1962).

nuclear physics as well as their analytic simplicity. Furthermore, one of them, the exponential potential, possesses an analytic solution for the s-wave scattering amplitude; i.e., the Fredholm determinant for the s-wave radial function can be given in closed form.[13,14] This fact has also been used in a study of various approximate methods in perturbation theory.[15] The plan of this paper is first to construct exact solutions for the s-wave scattering from the closed expression (for the exponential potential) or by direct numerical integration, utilizing the integral equation for the radial s-wave function, and then to calculate successively the same parameters by using standard variational techniques (Hulthén–Kohn approach) as a function of the size of the set of trial functions.

Finally, we compare these with the results obtained by directly looking at the eigenvalues and eigenvectors of the finite $n \times n$ matrix problem $(H_{ij} - EF_{ij})\alpha_j = 0$ (where $H_{ij} = \langle \chi_i H \chi_j \rangle$ is the matrix element of the Hamiltonian H and $F_{ij} = \langle \chi_i \chi_j \rangle$ the overlap integral in the space of expansion functions χ_i and α_j are the n variational determined expansion coefficients of the internal part of the wavefunction for scattering energy E). The phase shift for the n eigenvalues of this problem is then found according to a suggestion by Harris,[9] which requires $(\phi^{(m)}, (H - \epsilon_m)(S + \lambda C)) = 0$ where $\phi^{(m)} = D_i^{(m)} \chi_i$ is the appropriate eigenvector corresponding to eigenvalue ϵ_m of the matrix problem and S and C are the two linearly independent sinelike and cosinelike continuum solutions at energy $E = \epsilon_m$, while λ is the Harris value of the tangent of the phase shift.

II. CALCULATION OF s-WAVE PHASE SHIFT BY DIRECT ANALYTICAL AND NUMERICAL METHODS

For the exponential potential $V = -e^{-r}$, Bethe and Bacher[13] have long ago given an analytic solution for the s-wave phase shift:

$$k \cot \delta_0 = -\frac{1 + \sum_{n=1}^{\infty} \frac{(-2)^n}{n!} \sum_{m=1}^{n} \frac{(-1)^{m+1}}{(n-m)!\,(m-1)!} \frac{m}{m^2 + 4k^2}}{2 \sum_{n=1}^{\infty} \frac{(-2)^n}{n!} \sum_{m=1}^{n} \frac{(-1)^{m+1}}{(n-m)!\,(m-1)!} \frac{1}{m^2 + 4k^2}} = k \frac{\operatorname{Re} D(k)}{\operatorname{Im} D(k)}. \quad (1)$$

For our choice of range and strength parameters, $k = (2E)^{\frac{1}{2}}$, where E is the scattering energy. The Jost function[14] $D(k)$ is given in terms of standard functions

$$D(k) = J_{2ik}(2)/\Gamma(1 - 2ik). \quad (2)$$

Taking the limit $k \to 0$, we find from (1) directly, for the scattering length a_0,

$$\lim_{k \to 0} k \cot \delta_0(k) = -\frac{1}{a_0} = -\frac{1 + \sum_{n=1}^{\infty} \frac{(-2)^n}{(n!)^2}}{2 \sum_{n=1}^{\infty} \frac{(-2)^n}{(n!)^2} \sum_{m=1}^{n} \frac{1}{m}}. \quad (3)$$

The sums in (1) and (3) converge rapidly, and a few terms suffice to evaluate them. The results for the scattering amplitude $|T_0| = \sin \delta_0 / k$ for the exponential potential are shown in Table I. Atomic units are used throughout.

TABLE I. Scattering length a_0 and scattering amplitude for $V = -e^{-r}$.

$$a_0 = -\lim_{k \to 0} \frac{\tan \delta_0(k)}{k} = 8.693254$$

$k = (2E)^{\frac{1}{2}}$	$\lvert T_0 \rvert = (\sin \delta_0)/k$
0.1	6.998826
0.3	3.333307
0.5	1.870210
0.7	1.193395
0.9	0.8266603
1.0	0.7033727
2.0	0.2196898
4.0	0.06014139

For the Yukawa potential we compensate for our lack of an analytical solution by using the integral-equation[16] version of the Schrödinger equation for the radial s-wave function $u_0(r)$,

$$u_0(r) = \frac{\sin kr}{k} + \frac{2}{k} \int_0^r (\sin kr \cos kr' - \sin kr' \cos kr) \times V(r') u_0(r')\, dr, \quad (4)$$

and read off the asymptotic form ($r \to \infty$). The tangent of the phase shift $\delta_0(k)$ is given directly as the ratio of the coefficients of the cosine and sine terms of (4) as $r \to \infty$:

$$\tan \delta_0(k) = -\frac{2 \int_0^{\infty} \sin kr' V(r') u_0(r')\, dr'}{1 + 2 \int_0^{\infty} \cos kr' V(r') u_0(r')\, dr'}. \quad (5)$$

Equation (4) is readily integrated numerically, and we also find the scattering length a_0 by taking $k \to 0$

[13] H. A. Bethe and R. Bacher, Rev. Mod. Phys. **8**, 111 (1936).
[14] H. Jost, Helv. Phys. Acta **20**, 756 (1949).
[15] J. D. Bjorken and A. Goldberg, Nuovo Cimento **16**, 539 (1960).
[16] B. A. Lippmann and J. Schwinger, Phys. Rev. **79**, 469 (1950).

TABLE II. Scattering length a_0 and scattering amplitude for $V = -e^{-r}$ and $V = -e^{-r}/r$ from numerical integration of integral equation, $V(r) = -e^{-r}$, $V(r) = -e^{-r}/r$,

$$a_0 = -\lim_{k\to 0}\frac{\tan\delta_0(k)}{k} = 8.693260,$$

$$a_0 = -\lim_{k\to 0}\frac{\tan\delta_0(k)}{k} = 7.911394.$$

$k = (2E)^{\frac{1}{2}}$	$\lvert T_0\rvert = (\sin\delta_0)/k$	$\lvert T_0\rvert = (\sin\delta_0)/k$
0.1	6.998845	6.457029
0.3	3.333317	3.277175
0.5	1.870210	1.986129
0.7	1.193395	1.364755
0.9	0.8266603	1.011231
1.0	0.7033732	0.8877534
2.0	0.2196900	0.3545005
4.0	0.06014158	0.1284343

in (5):

$$a_0 = \lim_{k\to 0} -\frac{\tan\delta_0(k)}{k} = \frac{2\int_0^\infty V(r')u_0(r')r'\,dr'}{1 + 2\int_0^\infty V(r')u_0(r')\,dr'}. \quad (6)$$

We repeat the calculation for the exponential potential as well to determine the accuracy of our numerical integration. The computing time for evaluating (4), (5), and (6) is negligible, and we exhibit in Table II the scattering amplitude for both potentials for the energy range $0 \le E_k \le 227$ eV [$E_k = \frac{1}{2}k^2$ and $0 \le k \le 4(a_B^{-1})$]. We note the familiar general result that, despite the $(-1/r)$-type singularity for the Yukawa potential at the origin not shared by the exponential potential, the scattering is rather similar as we picked the same two parameters ($g = 1$, $r_0 = 1$) for potential strength and range, respectively. That is,

$$V(r) = gf(r); \quad r_0 = \int_0^\infty V(r')r'\,dr'.$$

III. STANDARD VARIATIONAL APPROACHES (HULTHÉN-KOHN)

Armed now with reliable data, we are prepared to enter the domain of variational approaches. We will first repeat the standard treatment of the variational type to define our notation and then proceed to use the familiar results in order to study their rate of convergence, efficiency, and accuracy. We will also come to grips with the source of previous difficulties[8] and show how these can be exploited to lead to a new insight and approach to the calculation of scattering parameters. The study of the functional $I(E) = \int u(H - E)u\,dr$ under variations of u is equivalent to the solution of the partial differential equation $(H - E)u = 0$, if u is expanded in a complete set of eigenstates of H. The idea of the variational approach is to work with a finite basis set and try to get a good approximation to the wavefunction or the energy by taking a sufficiently large number of terms in a linear expansion of u in terms of suitable functions. In bound-state calculations the Hylleras–Undheim principle provides the basis for a monotonic approach to the correct (negative) energy which is a lower bound, so that an increase in the number of trial functions necessarily leads to an improvement in the calculated energy.

For positive energies the energy spectrum of H becomes continuous, and, although Spruch, Rosenberg, and coworkers[6] have in a large number of papers established minimum principles for single as well as multichannel scattering, the use of these requires considerable numerical work on the static approximation wavefunction as well as information concerning the number of discrete eigenvalues of the modified Hamiltonian below the scattering energy.

From the point of view of a purely variational (Hulthén–Kohn) approach, it has been shown[17] that if one uses the standard Hulthén–Kohn variational principle and performs enough calculations with different size basis sets and a variable nonlinear parameter, one sees the emergence of some convergence towards a reasonable solution for the scattering amplitude. Since this approach is rather prohibitive in view of the large computational labor in actual situations of interest, we will, after rapidly using the Hulthén–Kohn approach for our case, go on and point out how to get around its main defect.

We define a trial wavefunction for the s-wave function:

$$u_t(k, r; \lambda_0, \lambda; \alpha_1, \cdots, \alpha_n) = \lambda_0\left(\frac{\sin kr}{k} + \lambda\frac{\cos kr}{k}(1 - e^{-r}) + \alpha_i\chi_i\right), \quad (7)$$

where the χ_i,

$$\chi_i = \{(2\alpha)^{i+\frac{1}{2}}/[(2i)!]^{\frac{1}{2}}\}r^i e^{-\alpha r}, \quad (8)$$

are our normalized basis functions chosen to describe the wavefunction inside the potential (α is a nonlinear scale parameter and we use the Einstein summation convention). If we assume that our trial wavefunction u_t is reasonably good, such that $\delta u = u_t - u$ is a small quantity (in some sense), we can construct the functional

$$I_t(\lambda_0, \lambda; \alpha_1, \cdots, \alpha_n) = \int u_t(k, r; \lambda_0, \lambda; \alpha_1, \cdots, \alpha_n) \times (H - E)u_t(k, r; \lambda_0, \lambda; \alpha_1, \cdots, \alpha_n)\,dr \quad (9)$$

[17] R. Armstead, thesis, University of California, Berkeley, Calif., 1965.

and study it under variations of the $n + 2$ parameters $\lambda_0, \lambda, \alpha_1, \cdots, \alpha_n$. Because of the nonvanishing of the continuum solutions as $r \to \infty$, we find by partial integration

$$\frac{\delta I}{\delta \lambda} = -\tfrac{1}{2}k, \tag{10}$$

since

$$I = \int u(r)(H - E)u(r)\, dr = 0 \tag{9'}$$

for the exact solution $u(r)$, which asymptotically tends to

$$(\sin kr)/k + \tan \delta_0(k)(\cos kr)/k.$$

Variation with respect to the remaining n coefficients α_i gives

$$\frac{\partial I_t}{\partial \alpha_i} = 0, \quad i = 1, \cdots, n, \tag{11}$$

and variation with respect to λ_0 gives

$$I_t = 0 \tag{9''}$$

for the correct trial function which we want to determine. We now face the problem of overdetermination, since we have $n + 2$ conditions for $n + 1$ unknowns α_i, $i = 1, \cdots, n$, and λ as the over-all normalization is arbitrary (nonnormalizable solutions, since the continuum solutions are nonzero throughout space). The Hulthén and Kohn methods differ by dropping (10) or (9''), respectively, to determine λ. The Kohn approach goes beyond solving (11) and (10) simultaneously by further extrapolation

$$\lambda_K = \lambda^{(1)} + (2/k)I_t(\lambda^{(1)}), \tag{12}$$

which is stationary to second order in $\delta\lambda$ because of (10). Equation (11) is kept in both approaches to eliminate the α_i. Let us write out Eqs. (9)–(11) to get a feeling for the difficulties.

Write u_t schematically

$$u_t = S + \lambda C + \phi, \tag{7'}$$

and rewrite

$$I_t = ((S + \lambda C + \phi), (H - E)(S + \lambda C + \phi)). \tag{9'''}$$

Then (11) can be written as an inhomogeneous matrix equation, using the hermiticity of $(H - E)$ with respect to the basis functions χ_i because of their vanishing, as $r \to \infty$,

$$(\chi_i, (H - E)\chi_j)\alpha_j = -(\chi_i, (H - E)S) - \lambda(\chi_i, (H - E)C), \tag{11'}$$

from which we determine the α_i by inverting [if $E \neq \epsilon_i$, the ith eigenvalue of the left-hand side of (11'), so that the inverse exists]:

$$\alpha_j = \alpha_j^{(s)} + \lambda\alpha_j^{(c)}, \tag{13}$$

with

$$\alpha_j^{(s)} = -G_{jm}^{-1}S_m = -(D_j^{(l)}D_l^{(m)}S_m/\delta_m),$$
$$\alpha_j^{(c)} = -G_{jm}^{-1}C_m = -(D_j^{(l)}D_l^{(m)}C_m/\delta_m) \tag{13'}$$

in terms of the eigenvectors and eigenvalues of G, where

$$G_{jk}D_k^{(m)} = \delta_m D_j^{(m)}, \tag{14}$$

and the remaining symbols are defined as

$$\begin{aligned} G_{jk} &= (\chi_j, (H - E)\chi_k) = H_{ij} - EF_{ij}, \\ S_m &= (\chi_m, (H - E)S), \\ C_m &= (\chi_m, (H - E)C), \\ \delta_m &= \epsilon_m - E. \end{aligned} \tag{15}$$

All these quantities are well defined for our choice of $V(r)$ and χ_m. Using (13), we rewrite

$$\phi(r) = (\alpha_j^{(s)} + \lambda\alpha_j^{(c)})\chi_j(r) = \phi_s(r) + \lambda\phi_c(r), \tag{16}$$

and, utilizing (16), we obtain

$$(\phi, (H - E)(S + \lambda C + \phi)) = 0. \tag{17}$$

We can, therefore, simplify (9) to

$$\begin{aligned} I_t &= (S, (H - E)(S + \phi_s)) + \lambda(S, (H - E)(C + \phi_c)) \\ &\quad + \lambda[(C, (H - E)(S + \phi_s)) \\ &\quad + \lambda(C, (H - E)(C + \phi_c))] \\ &= \bar{G}_{00} + \lambda\bar{G}_{01} + \lambda(\bar{G}_{10} + \lambda\bar{G}_{11}) = G_0(\lambda) + \lambda G_1(\lambda) \end{aligned} \tag{18}$$

in self-explanatory notation and, further,

$$\bar{G}_{01} = \bar{G}_{10} - \tfrac{1}{2}k. \tag{19}$$

We observe that all this is possible only if $\det |G| \neq 0$, i.e., if we are not encountering any eigenvalues $\delta_j = 0$. [The scattering energy has to be kept away from the eigenvalues of the finite matrix equation $(H_{ij} - EF_{ij})\alpha_j = 0$.]

The expressions for the Hulthén and Kohn variational estimates for the tangent of the phase shift follow directly by solving the following:

(A) the Hulthén expression $[I_t(\lambda) = 0]$:

$$\lambda_H = 2\bar{G}_{11}^{-1} \times \{[-(\bar{G}_{10} + \bar{G}_{01}) - [(\bar{G}_{10} + \bar{G}_{01})^2 - 4\bar{G}_{00}\bar{G}_{11}]^{\frac{1}{2}}\}; \tag{20}$$

(B) the Kohn expression:

$$\frac{\partial I_t}{\partial \lambda} = -\tfrac{1}{2}k,$$
$$\bar{G}_{01} + \bar{G}_{01} + 2\lambda^{(1)}\bar{G}_{11} = -\tfrac{1}{2}k, \tag{21}$$

which, using (19), gives

$$\lambda^{(1)} = -(\bar{G}_{10}/\bar{G}_{11}). \tag{21'}$$

This also corresponds to taking $G_1(\lambda) = 0$. Extrapolating to obtain the Kohn expression, we form

$$\lambda_K = \lambda^{(1)} + \frac{2}{k}[I_t(\lambda^{(1)})]$$
$$= \lambda^{(1)} + \frac{2}{k} G_0(\lambda^{(1)}) = -\frac{\bar{G}_{10}}{\bar{G}_{11}} + \frac{2}{k}\left(\bar{G}_{00} - \frac{\bar{G}_{10}\bar{G}_{01}}{\bar{G}_{11}}\right). \tag{22}$$

If

$$\det|\bar{G}| = \bar{G}_{00}\bar{G}_{11} - \bar{G}_{10}\bar{G}_{01} = 0,$$

the Hulthén and Kohn expressions agree, i.e.,

$$\lambda_H = \lambda_K = -\bar{G}_{10}/\bar{G}_{11} = -\bar{G}_{00}/\bar{G}_{01}, \tag{23}$$

as we verify by substitution of $\bar{G}_{00}\bar{G}_{11} = \bar{G}_{10}\bar{G}_{01}$ in the square root in (20). We can get to additional variational estimates for λ, which we will label as I and II, by taking

$$G_0(\lambda) = 0, \quad \lambda_I = -\bar{G}_{00}/\bar{G}_{01}, \tag{24}$$

$$G_1(\lambda) = 0, \quad \lambda_{II} = -\bar{G}_{10}/\bar{G}_{11}, \tag{25}$$

which reduces to (23) for $\bar{G}_{00}\bar{G}_{11} - \bar{G}_{10}\bar{G}_{01} = 0$. These two methods always give worse results than the Kohn formula, but approach the Hulthén formula for large basis sets. Most of this is covered in textbooks, and we give in Table III the results for both potentials for the s-wave scattering amplitude. We have used sets of basis functions up to ten functions, and point out that the results have converged to six decimals for $n > 6$ except at the lower limit. The relatively rapid convergence is no doubt due to the simplicity of the situation we are studying and the smooth nature of the potentials, a feature not expected to be present in electron–atom scattering calculations.

IV. THE HARRIS METHOD

We now shift our viewpoint and ask whether there is another way to choose the internal wavefunction $\phi(r)$ defined in (7'), suitably expanded in a variable size set of basis functions χ_m, so as to get a good estimate of the phase shift. In particular, it would be preferable to have a way of choosing $\phi(r)$ without having to study the scattering amplitude at a particular energy under variation of the size of the basis set and nonlinear parameters, in order to find a stationary value of the scattering amplitude. We go back to Eq. (11') and obtain

$$G_{ij}\alpha_j = -S_i - \lambda C_i. \tag{11''}$$

The root of the difficulties with the Hulthén and Kohn methods was the appearance of eigenvalues $\delta_j = 0$ corresponding to eigenvalues of the Hamiltonian matrix (in our finite basis set) equal to the scattering energy. Suppose we take the opposite approach[9] and look directly at the eigenvalues and eigenvectors of

$$G_{mn}\alpha_n = 0. \tag{26}$$

This is equivalent to transforming our basis set of functions χ_m to a new orthogonal basis $\phi^{(m)}(r)$, such that

$$(H_{mn} - EF_{mn})\alpha_n = 0, \tag{26'}$$

where we have written out

$$G_{mn} = (\chi_m, (H - E)\chi_n) = H_{mn} - EF_{mn} \tag{15'}$$

and

$$\phi^{(m)}(r) = D_n^{(m)}\chi_n(r) \tag{27}$$

is the mth eigenfunction of H in the finite function space of the χ_m, corresponding to the eigenvalue ϵ_m. Writing

$$\alpha_j = R_{jl}\alpha'_l \tag{28}$$

and multiplying (11') by R^T_{ki}, where R_{jl} is the orthogonal matrix with columns $D_l^{(j)}$ (H_{mn} and F_{mn} are symmetric matrices) which diagonalizes G_{ij}, we obtain

$$R^T_{ki}G_{ij}R_{jl}c'_l = -R^T_{ki}S_i - \lambda R^T_{ki}C_i \tag{11'''}$$

and, using

$$R^T_{ki}G_{ij}R_{jl} = (\epsilon_k - E)\delta_{kl}, \tag{29}$$

we find, at $E = \epsilon_k$ for λ, the tangent of the approximate phase shift

$$\lambda(\epsilon_k) = -\frac{R^T_{ki}S_i}{R^T_{ki}C_i} = \frac{(\phi^{(k)}, (H - \epsilon_k)S)}{(\phi^{(k)}, (H - \epsilon_k)C)} \tag{30}$$

corresponding to the use of the internal wavefunction

$$\phi^{(k)}(r) = D_n^{(k)}\chi_n(r). \tag{31}$$

We obtain, therefore, n values of the approximate phase shift $\delta_0(\epsilon_k)$ at scattering energies equal to the eigenvalues of (26'). This is the prescription for determining the phase shift given by Harris.[9] An alternate way of interpretation[18] is the use of Feshbach's projection operator technique

$$(Q = |\phi^{(\alpha)}\rangle\langle\phi^{(\alpha)}|)$$

to construct the scattering amplitude at $E = \epsilon_\alpha$. We find a simple pole at $E = \epsilon_\alpha$ (on the real axis in the complex energy plane), and the correct choice of the tangent of the phase shift in the matrix element of the numerator (32) cancels the singularity and gives,

[18] H. Morawitz, Ann. Phys. (N.Y.) 50, 1 (1968).

TABLE III. Hulthén, Kohn, I, and II methods scattering amplitude for $V = -e^{-r}$ and $V = -e^{-r}/r$ with n trial functions, with $\chi_n = \{(2\alpha)^{n+\frac{1}{2}}/[(2N)!]^{\frac{1}{2}}\}r^n e^{-\alpha r}$ as a function of n and for $\alpha = 1.0$ and $\alpha = 2.5$, respectively, $V = -e^{-r}$, $\alpha = 1.0$.

$k = (2E)^{\frac{1}{2}}$		$n = 1$	$n = 3$	$n = 5$	$n = 7$	$n = 9$	$n = 10$	exact
	H	7.07560	7.00162	6.99885	6.99883			
	K	7.09828	6.99914	6.99885				6.998845
0.1	I	7.63197	7.01928	6.99739	7.00793	7.00867	7.00859	
	II	7.30054	7.02005	6.99504	7.03638	7.03919	7.03921	
	H	3.33333	3.33331					
	K	3.33333	3.33331					
0.3	I	3.32289	3.33223	3.33291	3.33291	3.33291	3.33293	3.333307
	II	3.33333	3.33331	3.33331	3.33331	3.33331	3.33331	
	H	1.89096	1.87020					
	K	1.85069	1.87021					1.870210
0.5	I	1.81599	1.87716	1.86933	1.87051	1.87039	1.87006	
	II	1.72899	1.86844	1.86992	1.86996	1.87023	1.87024	
	H	1.17661	1.19354					
	K	1.18561	1.19302					1.193395
0.7	I	1.13430	1.19959	1.19465	1.19246	1.19329	1.19361	
	II	1.14726	1.21308	1.19192	1.19362	1.19336	1.19332	
	H	0.821441	0.820293	0.820623	0.820594			
	K	0.822736	0.826561	0.826645	0.826660			0.8266603
0.9	I	0.703994	0.848774	0.829425	0.826190	0.826691	0.827511	
	II	0.801599	0.834332	0.828389	0.826193	0.826723	0.826749	
				$V = [(-e^{-r})/r]$,	$\alpha = 2.5$			
	H	6.56755	6.46426	6.45702	6.45693	6.45696		
	K	6.60752	6.47009	6.45740	6.45697			
0.1	I	6.11879	6.75237	6.49450	6.46193	6.45769	6.45725	6.457029
	II	6.32152	6.58055	6.48024	6.46134	6.45784	6.45736	
	H	3.26713	3.27723	3.27717				
	K	3.27973	3.27739	3.27717				3.277175
0.3	I	3.22745	3.28970	3.27740	3.27681	3.27700	3.27704	
	II	3.24371	3.28079	3.27747	3.27736	3.27720	3.27719	
	H	1.97142	1.98612	1.98612				
	K	1.95355	1.98380	1.98612				1.986129
0.5	I	1.99999	1.98655	1.98805	1.96519	1.98611	1.98614	
	II	1.98046	1.99954	1.98635	1.98617	1.98613	1.98613	
	H	1.13148	1.36373	1.36457	1.36476			
	K	1.33476	1.36433	1.36474				1.364754
0.7	I	1.42856	1.37679	1.34242	1.36495	1.36499	1.36496	
	II	1.36734	1.36996	1.36565	1.36459	1.36484	1.36479	
	H	0.966966	1.00999	1.01122	1.01123			
	K	0.982351	1.01059	1.01122	1.01123			1.011231
0.9	I	0.532526	1.05665	1.01069	1.01277	1.01118	1.01133	
	II	1.02310	1.01922	1.01161	1.01189	1.01125	1.01115	

as approximate (finite) value of $T_0(\epsilon_\alpha)$,

$$|T_0(\epsilon_\alpha)| = C \lim_{E\to\epsilon_\alpha} \frac{(\sin kr + [\tan\delta_0 \cos kr](1-e^{-r}), (H-\epsilon_\alpha)\phi^{(\alpha)})}{E-\epsilon_\alpha} \times \left(\phi^{(\alpha)}, (H-\epsilon_\alpha)\frac{\sin kr}{k}\right), \quad (32)$$

$$C = \{(\phi^{(\alpha)}, (H-\epsilon_\alpha)\sin kr)^2 + (\phi^{(\alpha)}, (H-\epsilon_\alpha)\cos kr(1-e^{-r}))^2\}^{-\frac{1}{2}}.$$

Analytically, this corresponds to replacing the usual right-hand branch cut in the scattering amplitude by a simple pole at $E = \epsilon_\alpha$. Variation of the size of the basis set and the nonlinear parameter α allows one to reach almost any scattering energy as experience with reasonably sized matrices and scale parameters shows. The limit of going to a complete set of basis functions implies convergence to the exact result. The merit of the method lies in its rapid convergence to the exact result by optimal choice of the internal wavefunction. The frontal assault on the main difficulty (working directly at $E = \epsilon_\alpha$) thus overcomes the previous source of convergence problems due to the appearance of an increasing number of poles in (22), because $G_{11} = 0$. By considering the homogeneous

form of (11′), we arrive at a single equation determining the phase shift, which is physically equivalent to the choice of an internal wavefunction $\phi^{(m)}(r)$ at scattering energy ϵ_m, such that the continuum parts of the wavefunction [$\sin kr$, $\cos kr(1 - e^{-r})$] exactly cancel inside the range of the potential, when integrated over $\phi^{(m)}(r)$. This condition fixes the asymptotic relation between the two linearly independent solutions of the free-particle equation and thus determines the phase shift. The only drawback is that once one has chosen a certain size basis set and specific nonlinear parameters, the eigenvalues ϵ_m, at which the scattering amplitude is to be calculated, are fixed. In contrast, one chooses an arbitrary energy in the standard variational methods and works at that energy until numerical convergence emerges. Therefore, it is suggested to consider the various methods as complementary.

V. RESULTS AND CONCLUSIONS

We have used (30) for both potentials and calculated the eigenvalues and eigenvectors of G_{lm} for $0.5 \leq \alpha \leq 3.0$ and basis sets χ_m from 2 to 20 functions. The numerical results agree very well with our earlier results. In practice, the various steps are part of a simple Fortran computer program which calculates the matrix elements in the particular basis set chosen, diagonalizes the matrix, and computes the scattering amplitude. In Table IV we exhibit some typical eigenvalues and scattering amplitudes $T_0(\epsilon_\alpha)$ for various values of n and α. Even for very small basis sets ($n = 2$) the numerical accuracy is high. To compare the variational Hulthén, Kohn and I, II methods with the Harris method directly, we choose, for different values of α and different size basis sets n, eigenvalues of the latter method and compare the result of the variational program with these input data with the value of $|T_0| = (\sin \delta_0)/k$ obtained from the Harris method. The results are exhibited in Table V, and we note that numerical agreement is excellent, particularly for larger n, where it again reaches the numerical limit of accuracy of our calculation, i.e., one part in 10^6. Additionally, we note that $\alpha = 2.5$ gives somewhat better results for all methods than $\alpha = 1.0$ for the potential parameters chosen. In general, the values of T_0 calculated from the Hulthén, I, and II methods are very close to those obtained from the Harris method, while the Kohn formula is noticeably closer to the exact number for $|T_0|$. We consider this a significant result of our calculations, supporting the preference of the Kohn method over the other studied variational formulas. We suggest, therefore, that the optimal procedure for finding phase shifts and scattering amplitudes is to select scattering energies determined by diagonalizing the matrix Hamiltonian in the subspace of expansion functions χ_i (as in the Harris method). The use of the Kohn formula then leads to the best variational estimate of the phase shift.

TABLE IV. Some values of the scattering amplitude obtained from the Harris method for $V = -e^{-r}$ and $V = -e^{-r}/r$ for different size basis sets n and different values of α in the n-dimensional trial function space of

$$\chi_{(n)} = \{(2\alpha)^{n+\frac{1}{2}}/[(2N)!]^{\frac{1}{2}}\}r^n e^{-\alpha r}.$$

k_α	α	n	$\lvert T_0\rvert = (\sin \delta_0)/k_\alpha$
	$V = -e^{-r}$		
0.590159	0.5	2	1.46680
5.06620	3.0	2	0.03692
0.310458	0.6	4	3.21778
2.00004	3.0	4	0.217182
0.317372	0.9	6	3.14871
3.00940	2.5	6	0.10303
0.592386	0.8	8	1.50127
1.99404	2.5	8	0.220917
0.403675	0.7	10	2.42147
0.184817	0.9	10	5.05029
0.140357	3.0	10	6.00879
0.114078	0.5	20	6.64135
0.487047	1.5	20	1.93271
0.406544	3.0	20	2.40170
	$V = -e^{-r}/r$		
0.114551	0.8	2	6.40158
1.11738	3.0	2	0.79707
0.284122	0.6	4	3.41342
0.519956	3.0	4	1.90546
0.504376	0.5	6	1.97307
1.01550	1.5	6	0.871354
0.615976	0.6	8	1.57858
0.998357	2.0	8	0.889832
0.727807	0.5	10	1.30557
0.592412	1.5	10	1.65045
0.0905956	3.0	10	6.66360
0.309523	1.0	20	3.14956
0.585738	1.5	20	1.69651
4.20888	2.0	20	0.157604

We note that the direct use of the Kohn method at arbitrary scattering energies is faster, if one is interested in only a few scattering energies, since one can solve Eq. (18) directly without constructing G_{ij}^{-1} or finding its eigenvectors and eigenvalues. Variation of the nonlinear scale parameter α, as done in Refs. 8 and 7, forces one to repeatedly solve these equations the nonlinear scale parameter α, as done in references 8 and 7, forces one to repeatedly solve these equations at the same energy to avoid possible singularities. These are due to the vanishing of G_{11} [Eq. (22)] in the Kohn formula as noted by Nesbet.[19] If one has to solve Eq. (13) many times over, it becomes simpler

[19] R. Nesbet, Phys. Rev. 175, 134 (1968).

TABLE V. Comparison of the scattering amplitude for $V = -e^{-r}$ and $V = -e^{-r}/r$ obtained from the Harris method with the Hulthén, Kohn, I, and II variational methods at a Harris eigenvalue for different size basis sets and different values of α.

n	k	T_0^{Harr}	T_0^{H}	T_0^{K}	T_0^{I}	T_0^{II}	T_0^{exact}
			$\alpha = 2.5, \quad V = -e^{-r}$				
2	1.12907	0.5603282	0.5602522	0.5798384	0.560382	0.5603870	0.5808257
4	0.492075	1.89945	1.899398	1.907729	1.899453	1.899453	1.908088
6	0.264369	3.76536	3.765392	3.764224	3.765362	3.765362	3.764254
8	0.766193	1.04872	1.048632	1.048950	1.048725	1.048725	1.048950
10	0.620347	1.41033	1.410399	1.410395	1.410331	1.410330	1.410395
			$\alpha = 2.5, \quad V = -e^{-r}/r$				
2	0.864666	1.07859	1.078646	1.060304	1.078585	1.078585	1.061822
4	0.389493	2.56653	2.566524	2.566571	2.566525	2.566525	2.566571
6	0.205068	4.44529	4.445375	4.443515	4.445292	4.445292	4.443532
8	0.716891	1.32771	1.327591	1.327388	1.327711	1.327711	1.327386
10	0.582987	1.68025	1.680205	1.680200	1.680250	1.680250	1.680198
			$\alpha = 1.0, \quad V = -e^{-r}/r$				
2	0.192450	4.69922	4.699406	4.647494	4.699217	4.699217	4.643500
4	0.525406	1.88405	1.884033	1.883143	1.884051	1.884051	1.883171
6	0.341028	2.91579	2.915783	2.916023	2.915788	2.915788	2.916027
8	0.472234	2.10982	2.109837	2.109847	2.109847	2.109820	2.109846
10	0.2021	4.48943	4.489427	4.489262	4.489449	4.489449	4.489282
			$\alpha = 1.0, \quad V = -e^{-r}$				
2	0.214283	4.49304	4.493255	4.514291	4.493052	4.493052	4.514167
4	0.565136	1.59780	1.597772	1.597459	1.597804	1.597804	1.597465
6	0.360080	2.75384	2.753819	2.753638	2.753837	2.753837	2.753639
8	0.490367	1.91625	1.916270	1.916401	1.916248	1.916248	1.916401
10	0.2100	4.58747	4.587366	4.587385	4.587466	4.587465	4.587387

to find the eigenvalues and eigenvectors of G_{ij} once and for all, and to obtain the relevant quantities by multiplication [see Eq. (13′)]. We feel, therefore, that our prescription leads to the best variational result in an efficient manner, provided one is interested in a reasonable energy range. As the size of the basis sets tends towards completeness ($n \to \infty$), we expect all approximate results to converge to the exact T_0. This, of course, corresponds to the internal wavefunction approaching an exact solution of the Schrödinger equation over the range of the potential.

ACKNOWLEDGMENTS

The author wishes to express his appreciation to R. K. Nesbet and H. Taylor for interesting comments and discussion.

Functional Random-Walk Model of the Many-Particle System

IWAO HOSOKAWA*
Laboratory for Theoretical Studies, NASA-Goddard Space Flight Center, Greenbelt, Maryland 20771

(Received 17 June 1969)

By Fourier-transforming the author's recently proposed state functional formalism for the BBGKY hierarchy, a new perspective of nonequilibrium statistical mechanics is given: the basic equation is formally very close to the Fokker–Planck equation and may readily be modified to a universal master equation (with irreversibility) by a slight change. Hence, the problem reduces to one of a generalized random-walk such that the stochastic quantity to be considered is the particle-number density in the 1-body phase space. A general solution is formulated for the weak interaction case.

1. INTRODUCTION

Recently the functional formalism of classical nonequilibrium statistical mechanics was proposed by the author[1] to search for general closed-form solutions for distribution functions governed by the Bogoliubov–Born–Green–Kirkwood–Yvon (BBGKY) hierarchy for the infinitely-many-particle system with a finite average density. There the state functional, which embraces all distribution functions in itself as the coefficients of the *functional* Taylor expansion (except for the numerical factors), was introduced. It is governed by a single second-order functional-differential equation. This formalism originates with Bogoliubov,[2] but a basic modification was necessary for a meaningful functional calculus to be performed within the formalism. Be that as it may, it would be quite inconvenient if the state functional were too *fine grained* to represent a real irreversible process, even if the solution could be numerically pursued. In fact, on the basis of the exact dynamics, i.e., the Liouville equation or its equivalent, we can hardly expect all possible arbitrary states to develop into a unique, steady (e.g., equilibrium) state. This is obvious, if we remember the time-reversibility of the equation as well as the fact that the entropy in Gibbs's sense is conserved so that the maximum entropy is never achieved, starting from a state of lower entropy. In order to establish an entropy-productive nonequilibrium statistical mechanics, some modification of the basic equation, comparable with coarse-graining procedure, also seems to be unavoidable in our state functional formalism.

The first aim of this paper is to find the best irreversible perspective for the evolution of the state functional.[3] We first notice that the Fourier-transformed basic equation in the *cylinder functional approach* (established in Sec. 2) has a structure very close to the Fokker–Planck equation which describes a Markov *stochastic* process, but retaining the time-reversibility as a natural consequence of its equivalence to the BBGKY hierarchy (in the limit when the number of particles tends to infinity). The best method of modifying the equation into a real Fokker–Planck equation at the expense of losing the time-reversibility is described in Sec. 3. As a result, our problem reduces to nothing but the problem of a generalized random walk played within the special function space equivalent to a certain Riemannian space by the so-called *stochastic* particle-number density function in the 1-body phase space. The mathematical procedure of such a modification is unique in the sense that any other Markov process expected to simulate nonequilibrium statistical mechanics is not closer to the dynamical process than the present one; but it seems somewhat difficult to re-express the procedure in simple conventional physical words, such as coarse graining.

As the paper's second object, the general method is applied to the case of a gas or a weakly interacting plasma in Sec. 4. In this case, the equivalent Riemannian space reduces approximately to a Euclidean space, so that the complexity in calculation is reduced. The closed-form general solution of the initial-value problem is formulated in terms of a repeated multiple integral, which can, in practice, be conveniently solved by the Monte Carlo quadrature. Since in our theory the accuracy of calculation is increased simply by increasing the multiplicity M of cylinder functional but not by changing (or deepening) the formulation in a complicated way and, moreover, since there is no difficulty in principle in applying the theory to the strong interaction case, the proposed approach seems

* NRC-NASA Resident Research Associate.

[1] I. Hosokawa, Progr. Theoret. Phys. (Kyoto) **36**, 420 (1966); J. Math. Phys. **8**, 221 (1967); Progr. Theoret. Phys. (Kyoto) **39**, 242 (1968).

[2] N.N.Bogoliubov, in *Studies in Statistical Mechanics*, J. deBoer and G. E. Uhlenbeck, Eds. (North-Holland Publ. Co., Amsterdam, 1962), Vol. 1.

[3] A short explanation of the idea was presented at the International Conference on Statistical Mechanics (Kyoto, September, 1968); see J. Phys. Soc. Japan Suppl. **26**, 224 (1969).

to have the proper advantage which is not found in other analytical approaches to the BBGKY hierarchy or the Liouville equation which exploit the more or less sophisticated series expansion in a small parameter,[2,4,5] which may not necessarily be convergent.

Another type of statistical-dynamical research, which may be also called a functional approach, has been developed by considering the evolution of the probability on an ensemble of the events, a member of which evolves strictly according to some dynamics such as the Vlasov equation, the Langevin equation, etc.[6,7] Of course, this research is different in principle from ours, except for the case of taking the Klimontovich equation[8] as the dynamics, when the characteristic functional equation is formally equivalent to the BBGKY hierarchy, as was verified by Nakayama and Dawson.[6] However, it will be seen (in Sec. 3) that, if the Vlasov equation is taken as the dynamics, there is a somewhat close relation between this ensemble mechanics and the presented perspective of our formalism.

For simplicity the discussion is restricted to a single-component system. Extension to a multicomponent system is straightforward.[1]

2. FOURIER-TRANSFORMED STATE FUNCTIONAL FORMALISM

According to the previous paper,[1] if the s-body generic distribution function is denoted by F_s, the state functional is defined as

$$\psi(y,t) = 1 + \sum_{s=1}^{\infty} \frac{i^s}{s!} \int_X \cdots \int_X F_s(x_1, \cdots, x_s, t) \times y(x_1) \cdots y(x_s)\, dx_1 \cdots dx_s, \quad (2.1)$$

where $i = \sqrt{-1}$, X is the entire 1-body phase space, $x_1, \cdots, x_s \in X$, t is the time variable, and $y(x)$ is a real-valued function. Once the state functional is introduced, the BBGKY hierarchy for the infinitely-many-particle system with a finite average density is completely replaced by the following single basic equation with functional differentiation:

$$\frac{\partial \psi}{\partial t} = i \int_X y(x) \left[H_1(x); \frac{\delta \psi}{i \delta y(x)} \right] dx - \tfrac{1}{2} \int_X \int_X [y(x)y(x') - iny(x) - iny(x')] \times \left[\phi(|q - q'|); \frac{\delta^2 \psi}{i^2 \delta y(x) \delta y(x')} \right] dx\, dx', \quad (2.2)$$

where $H_1(x)$ is the 1-body Hamiltonian, [;] the Poisson bracket, n the average number density, and $\phi(|q - q'|)$ the interaction potential between particles in which q denotes the displacement vector in the physical space. The operator $\delta/\delta y(x)$ denotes a functional derivative.[9,10] ψ should be subject to the following two associative conditions[1]:

$$\psi(0, t) = 1, \quad (2.3)$$

$$\lim_{V \to \infty} \int_X \frac{\delta \psi}{i \delta y(x)} \frac{dx}{V} = \psi, \quad (2.4)$$

where V is the volume containing the entire system. Both conditions emerge from the form of (2.1) on using the definition of F_s.

Here we introduce a kind of cylinder functional approach to functionals. If $\{s_j(x)\}$ is an orthonormal function set in X, we can calculate $a_j = \int_X s_j(x) y(x)\, dx$ to define a new function

$$y^M(x) = \sum_{j=1}^{M} a_j s_j(x). \quad (2.5)$$

We call $\psi(y^M)$ the cylinder functional $\psi^M(y)$, which is still a functional of y through the definition of a_j, though it can essentially be considered a function of M variables $\{a_j\}$. Then, ψ is understood as ψ^M in the limit $M \to \infty$ if it is convergent (for example, with respect to the "maximum" norm). If ψ is replaced by ψ^M, the basic equations (2.2)–(2.4) become more feasible to treat, since then we have the relation

$$\frac{\delta}{\delta y(x)} = \sum_{j=1}^{M} \frac{\delta a_j}{\delta y(x)} \frac{\partial}{\delta a_j} = \sum_{j=1}^{M} s_j(x) \frac{\partial}{\partial a_j} \quad (2.6)$$

and, thus, (2.2) becomes just a partial differential equation. From a practical point of view we always consider ψ^M first, assuming that the converging sequence $\{\psi^M\}$ of state cylinder functionals exists. Then, we can introduce a new functional $\rho^M(z, t)$ as the Fourier component of ψ^M; namely, we have

$$\psi^M(y, t) = \int_{A^M} \rho^M(z, t) \exp\left[i \int y^M(x) z^M(x)\, dx \right] \delta z^M, \quad (2.7)$$

[4] E. A. Frieman, J. Math. Phys. **4**, 410 (1963); J. E. McCune, G. Sandri, and E. A. Frieman, in *Rarefied Gas Dynamics (Advances in Applied Mechanics, Suppl. 2; Third International Symposium on Advances in Applied Mechanics, Paris*, 1962), J. A. Laurmann, Ed. (Academic Press Inc., New York, 1963), Pt. 1. A classification of kinetic equation approaches is given in T.-Y. Wu, *Kinetic Equations of Gases and Plasmas* (Addison-Wesley Publ. Co., Inc., Reading, Mass., 1966).

[5] I. Prigogine and R. Balascu, Physica **25**, 281, 302 (1959); **26**, 145 (1960); I. Prigogine, *Non-Equilibrium Statistical Mechanics* (Interscience Publishers, Inc., New York, 1963).

[6] T. Nakayama and J. M. Dawson, J. Math. Phys. **8**, 553 (1967).

[7] G. Vojta, J. Phys. Soc. Japan Suppl. **26**, 221 (1969); W. E. Brittin and W. R. Chappell, J. Math. Phys. **10**, 661 (1969).

[8] Iu. L. Klimontovich, Zh. Eksp. Teor. Fiz. **33**, 982 (1957) [Sov. Phys.—JETP **6**, 753 (1958)]; *Statisticheskaya Teoriya Neravnovesnyikh protsessov v plazme* (Moscow State University, 1964) [English transl.: *The Statistical Theory of Non-equilibrium Processes in a Plasma* (Pergamon Press, Inc., New York, 1967)].

[9] V. Volterra, *Theory of Functionals and of Integral and Integro-differential Equations* (Dover Publications, New York, 1959).

[10] E. Hopf, J. Ratl. Mech. Anal. **1**, 87 (1952).

where

$$\int_{A^M} \delta z^M \equiv \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{j=1}^{M} \frac{db_j}{(2\pi)^{\frac{1}{2}}} \tag{2.8}$$

and $z^M(x) = \sum b_j s_j(x)$. For simplicity, we omit all the superscript M hereafter, unless necessary for clarity, keeping in mind that every functional is a cylinder functional.

As a result of the transformation (2.7), the basic equation (2.2) transforms to the following equation for ρ:

$$\frac{\partial \rho}{\partial t} = -\int_X \frac{\delta}{\delta z(x)} \{Qz(x)\rho\}\, dx + \frac{1}{2}\int_X \int_X \frac{\delta^2}{\delta z(x)\delta z(x')} \times \{[\phi(|q-q'|); z(x)z(x')]\rho\}\, dx\, dx', \tag{2.9}$$

where Q denotes the nonlinear operator such that

$$Qz(x) = [H_1(x); z(x)] + n\int_X dx'[\phi(|q-q'|); z(x)z(x')]. \tag{2.10}$$

It is noted[1] that $\partial z/\partial t = Qz$ is nothing but the self-consistent Vlasov equation and, therefore, Q may be called the Vlasov operator. Corresponding to (2.3) and (2.4), we have the conditions

$$\int_A \rho \delta z = 1, \tag{2.11}$$

$$\int_A \rho \left[\lim_{V\to\infty} \int_X z(x)\frac{dx}{V} - 1\right] \times \exp\left[i\int_X y(x)z(x)\, dx\right]\delta z = 0. \tag{2.12}$$

From (2.12), ρ should vanish for all $z(x)$ unless $z(x)$ is such that

$$\lim_{V\to\infty} \int_X z(x)\frac{dx}{V} = 1, \tag{2.13}$$

which defines a hyperplane in R^M. We may call this the hyperplane A_1. The restriction (2.12) is conservative in time, since the conservation of (2.4) was proved by the previous paper.[1] Therefore, without any inconsistency, we have only to consider the basic equation (2.9) always in this hyperplane rather than R^M. Accordingly, A in (2.7) and (2.11) may be read as A_1 on the understanding that $\int_{A_1} \delta z$ indicates the volume integral over A_1. The time invariance of the condition (2.11) is evident by virtue of (2.9).

Since our basic equations obtained so far are equivalent to the BBGKY hierarchy or the Liouville equation in the limit $N = nV \to \infty$ (n finite), any dynamical property of the particle system is preserved in our formalism. It is easy to verify the time-reversibility of the basic equation (2.9): if $\rho(z(q, p), t)$ is a solution, then $\rho(z(q, -p), -t)$ becomes another solution, p denoting the momentum part of the vector x. Also, it is possible to derive the conservation laws of mass, momentum, angular momentum, and energy directly on the basis of (2.9). In this case, on comparing (2.1) with the series expression of (2.7) in powers of y, we first have

$$F_1(x) = \int_{A_1} \rho(z, t)z(x)\delta z, \tag{2.14}$$

$$F_2(x, x') = \int_{A_1} \rho(z, t)z(x)z(x')\delta z; \tag{2.15}$$

we then can proceed with examining the time evolution of each physical quantity related with ρ through the above relations.

As a representative example, let us examine the momentum case. From (2.14) and (2.9), we calculate

$$\frac{\partial}{\partial t}\int_X npF_1(x)\, dx = n\int_{A_1} \frac{\partial \rho}{\partial t}\int_X pz(x)\, dx\delta z = n\int_{A_1}\int_X pQz(x)\rho\, dx\delta z,$$

where use was made of partial integration, keeping in mind that $|\rho| \to 0$ at infinity ($\int_X z^2\, dx = \infty$) in A_1 in order for $\int_{A_1} |\rho|\, \delta z$ to exist. As is easily seen, there is no effect of the term with a second-order functional derivative in (2.9) in this case, since $\int_X pz(x)\, dx$ is a functional of no higher order than linear in z. A further calculation gives

$$\begin{aligned}
&\frac{\partial}{\partial t}\int_X npF_1\, dx \\
&= n\int_{A_1} \rho\left\{\int_X p[H_1; z(x)]\, dx + n\int_X\int_X p[\phi; z(x)z(x')]\, dx\, dx'\right\}\delta z \\
&= n\int_{A_1} \rho\left[-\int_X \frac{\partial H_1}{\partial q} z(x)\, dx - n\int_X\int_X \left(\frac{\partial \phi}{\partial q} + \frac{\partial \phi}{\partial q'}\right) z(x)z(x')\, dx\, dx'\right]\delta z.
\end{aligned} \tag{2.16}$$

Here the boundary condition on $z(x)$ was taken into account: $z(x) \to 0$ for $|p| \to \infty$ and on the boundary of the configuration space. The second term in the curly bracket vanishes because $\partial\phi/\partial q = -\partial\phi/\partial q'$. (This illustrates that the sum of internal forces vanishes because of the action–reaction law.) The first term also vanishes if there is no external force ($\partial H_1/\partial q = 0$). Otherwise, it expresses the total external force. Therefore, we can see in (2.16) the generalized conservation law of momentum. It is needless to say that total

momentum $\int_X npF_1\,dx$ is infinite in the limit $N = nV \to \infty$, but it can be meaningful if we consider the average momentum per particle or per unit volume. The other conservation laws are quite similarly examined, except for the energy case in which care should be taken of the fact that the effect of the term with a second-order functional derivative in (2.9) appears for the part of the interaction energy including the expression (2.15), which has the factor quadratic in z explicitly.

Finally, we add a comment on the other functional formalism equivalent to the BBGKY hierarchy by Nakayama and Dawson.[6] In this formalism the characteristic functional plays a main role of describing the state according to the theory of Hopf[10] for turbulence mechanics. There the evolution of an ensemble is considered, all members of which are microscopic states evolving according to the Klimontovich equation[8] starting with various initial positions and momenta of particles given in the 1-body phase space. The basic equation, i.e., the Hopf equation for this ensemble, is somewhat different from (2.2); in particular, it is remarkable that it has no term with a coefficient quadratic in $y(x)$. Correspondingly, the characteristic functional itself is different from the state functional (2.1). A difficulty with this formalism (occurring if we want to develop the theory in a functional-analytical way), however, stems from the fact that every solution of the Klimontovich equation does not represent a microscopic state. Hence, all solutions of the Hopf equation are not realistic, even if they satisfy the general conditions on the characteristic functional. For the solution to be realistic, the members of the ensemble should be restricted to Klimontovich's special class of singular functions in terms of superposed δ functions.[8] This limits the class of allowed characteristic functionals severely, and it would not be easy to formulate this restriction in a simple manner. Another disadvantage may be seen in the difficulty in finding an appropriate simple-form initial characteristic functional which can be used in practice; for the state functional it is easy, as described in Ref. 1 and shown in the next section.

3. MASTER-EQUATION APPROACH

The basic equation (2.9) is very close in form to the Fokker–Planck equation which describes a Markov stochastic process. The first term of the right-hand side is the analog of the so-called friction term and the second resembles the diffusion term. However, there is a difference in principle between (2.9) and the Fokker–Planck equation; the former is time reversible, as was mentioned already, but the latter is time irreversible. This reflects the fact that the coefficient function $[\phi; z(x)z(x')]$ inside the second-order derivative is not positive-definite. This similarity and difference may contain a key to solving the noted historical question: how to bridge the microscopic reversibility and the macroscopic irreversibility in the system evolution in a general way.

Now, if the coefficient function is considered as the sum of the positive-definite and negative-definite parts at each local point in A_1, the former plays a role of creating an irreversible process, while the latter that of destroying it or creating an anti-irreversible process so that both roles may offset each other to cause the reversible process exactly. So, in order to extract a purely irreversible process from the dynamical process, evidently it is necessary and sufficient to retain only the former part of the coefficient function, so that (2.9) reduces to a real Fokker–Planck equation. This is the simplest universal way of introducing the positive time arrow in the system evolution. There is no further assumption in this procedure, such as weak interaction, diluteness, etc. Therefore, its application will be free from any such restriction on physical conditions. Let the Fokker–Planck equation thus obtained be written as

$$\frac{\partial \tilde{\rho}}{\partial t} = -\int_X \frac{\delta}{\delta z(x)}\,[Qz(x)\tilde{\rho}]\,dx + \frac{1}{2}\int_X \int_X \frac{\delta^2}{\delta z(x)\delta z(x')}\,\{\mathcal{F}[\phi; z(x)z(x')]\tilde{\rho}\}\,dx\,dx', \tag{3.1}$$

where the symbol ρ has been replaced by $\tilde{\rho}$ to distinguish its approximate nature because of the above procedure and $\mathcal{F}$ is the operator on the coefficient function to make it positive definite in the way just described. There can still be many ways of constructing a Markov process starting from (2.9); for example, $\mathcal{F}[\phi; z(x)z(x')]$ plus any nonnegative function can also be positive definite. But it is natural that (3.1) is the closest approximation to (2.9) among them, since in this modification the first term is invariant and the second term was *minimumly* changed to make a real diffusion term. Then, the error bound $\sup|\rho - \tilde{\rho}|$ would be minimum. Thus, (3.1) may be called diffusion term. Then, the difference $|\rho - \tilde{\rho}|$ on an average would be minimum. Thus, (3.1) may be called the master equation in the sense that it masters the irreversible system evolution which is the most loyal to the basic dynamical rule.

By the property of the Fokker–Planck equation it is possible to interpret $\tilde{\rho}$ as the probability density in the space A_1. The condition (2.11) plays an important

role for this interpretation; that is,

$$\int_{A_1} \tilde{\rho}\delta z = 1. \tag{3.2}$$

If the diffusion term is neglected in (3.1), it can be found that the equation governs the time evolution of the probability on an ensemble of the trajectories in the space A_1, which develop from various initial values according to the Vlasov equation, $\partial z/\partial t = Qz$. (Consider the characteristic curves in A_1.) This fact is obvious from the perfect analogy with the Hopf equation in turbulence mechanics,[10] if we go back to the equation for the characteristic functional $\tilde{\psi}$ which corresponds to $\tilde{\rho}$ [defined in the same way as (2.7)]. In this case, we really deal with nothing but a so-called turbulent field which is basically governed by the Vlasov equation instead of the Navier–Stokes equation. Here we find a complete coincidence with the other statistical-dynamical research based on the Vlasov equation mentioned in the introduction.[6,7] If the diffusion term in (3.1) is included, the coincidence breaks down and the motion (expressed by z) is affected by some *internal* random force implied by $\mathfrak{F}[\phi; z(x)z(x')]$, as is well expected from the Langevin equation for Brownian motion; thus, *what we may call a turbulence* will be more irregular and more random. For this case, there is the analogy with the Novikov equation in the turbulence mechanics with random force action[11] (though there is a slight difference in situation in that the random force in the Novikov equation is not internal but externally given independently of the field z). From all these facts, it is quite reasonable to interpret $z(x)$ as the *stochastic* particle-number density in X, normalized in the sense of (2.13). Obviously, $z(x)$ is not a macroscopic observable [as is also known from the relations (2.14) and (2.15)], nor is it a microscopic density such as considered by Klimontovich,[8] since it can be a regular function; but it may be understood as a fictitious, mathematical working field with an intermediate property. [See Appendix for the essential nonnegativeness of $z(x)$.]

Let us see if (3.1) is consistent with the conservation laws in spite of its modification by $\mathfrak{F}$. It is known from the discussion of the previous section that the term with the second-order functional derivative in the basic equation has no effect on the conservation laws except for conservation of energy. Accordingly, it is evident that the conservation laws of mass, momentum, and angular momentum hold also for (3.1). For the energy case, however, we have

$$\frac{\partial}{\partial t}\Big[\int_X nH_1(x)\tilde{F}_1(x)\,dx + \int_X\int_X \tfrac{1}{2}n^2\phi(q-q')\tilde{F}_2(x,x')\,dx\,dx'\Big] = \tfrac{1}{2}n^2\int_{A_1}\int_X\int_X \phi\mathfrak{F}[\phi; z(x)z(x')]\,dx\,dx'\tilde{\rho}\delta z, \tag{3.3}$$

which cannot vanish in general. The right-hand side is $O(n^2\lambda^2)$ (λ being the representative order of magnitude of ϕ), but the error will accumulate with time. Therefore, it is necessary to correct this error in our master equation some way. This can be made by confining the space for z into a manifold in which the conservation of energy is strictly insured. Indeed, since the total energy found on the left-hand side of (3.3) is rewritten in terms of (2.14) and (2.15) as

$$\int_{A_1}\tilde{\rho}\Big[\int_X nH_1 z(x)\,dx + \int_X\int_X \tfrac{1}{2}n^2\phi z(x)z(x')\,dx\,dx'\Big]\delta z,$$

we can impose the condition on z,

$$\int_X nH_1 z(x)\,dx + \int_X\int_X \tfrac{1}{2}n^2\phi z(x)z(x')\,dx\,dx' = \text{const}, \tag{3.4}$$

to guarantee the constancy of the total energy. Equation (3.4) prescribes a subset of A_1, which may be called a quadratic hypersurface. [The difficulty arising from the singularity of ϕ at the particle-core region ($|q-q'| \simeq 0$) is excluded by limiting the domain of the double integration to the outside of the core.]

It is thus essential that our basic equation (3.1) should be redefined in such a Riemannian space as prescribed by (3.4). There is no special difficulty in dealing with the Fokker–Planck equation in a Riemannian space.[12] As is well known, a Fokker–Planck equation governs a generalized Brownian motion and, hence, we may imagine from our equation a generalized Brownian motion or random walk in the Riemannian space.

It is known that there is a unique steady state of $\tilde{\rho}$, and $\tilde{\rho}$ tends asymptotically to this state irrespectively of any initial condition.[13] It is unfortunate that we cannot at present have any explicit form of the steady-state solution to be compared with the equilibrium canonical distribution for the Gibbs ensemble. However, some similarity in situation may be expected. Namely, the final asymptotic state is the state in

[11] E. A. Novikov, Zh. Eksp. Teor. Fiz. **47**, 1919 (1964) [Soviet Phys.—JETP **20**, 1290 (1965)]; I. Hosokawa, J. Phys. Soc. Japan **25**, 271 (1968).

[12] A. Kolmogoroff, Math. Ann. **108**, 149 (1933).

[13] A. H. Gray, Jr., J. Math. Phys. **6**, 644 (1965).

which all Brownian particles have fully diffused over the whole Riemannian space, i.e., the quadratic hypersurface (3.4) that corresponds to the constant-energy shell in the *grand* ($6N$-dimensional) phase space; and then every point in that space is realizable with some measure (possibly including zero) irrespective of an initial state, just like a *grand* phase point in the energy shell in the equilibrium state. Entropy may be introduced as $S^M = \int \tilde{\rho}^M \log \tilde{\rho}^M \delta z^M / M$,[14] and then it can be identified that the diffusion term of (3.1) takes charge of entropy-productive action. But we note that, in our theory, the concept of entropy is not particularly necessary to conclude the one-directional evolution of the system, since we have already the uniquely given asymptotic state.

The initial condition imposed on $\tilde{\rho}$ is easily formulated, if the initial correlations among particles are neglected. Namely, on starting from the initial state functional[1]

$$\psi = \exp\left[i \int_X y(x) F_0(x)\, dx \right], \tag{3.5}$$

where $F_0(x)$ is the given initial 1-body distribution function, we have

$$\rho = \delta[z(x) - F_0(x)] \tag{3.6}$$

by means of the relation (2.7). (See Ref. 1 for the δ functional $\delta[\ ]$.) Naturally, here we rewrite ρ as $\tilde{\rho}$. This means that our random walk begins from the single point in the Riemannian space which is given as $F_0(x)$. Then, the total of the Brownian trajectories beginning from this point provides all the physical information of the irreversible system evolution. Indeed, as is known from comparing (2.1) and (2.7), all $\tilde{F}_s$ (the symbol $\tilde{F}_s$ means the approximation to F_s in the sense of our master-equation approach) are given as the sth-order correlation function of the stochastic field $z(x)$:

$$\tilde{F}_s(x_1, \cdots, x_s, t) = \int z(x_1) \cdots z(x_s) \tilde{\rho}(z, t) \delta z, \tag{3.7}$$

where the integration is taken over the Riemannian space in A_1; and all physical quantities are related to $\tilde{F}_s$. It may be further noted that, since the fluctuation of a macroscopic quantity is closely related to the behavior of F_2,[2] an observable macroscopic turbulence should be related through $\tilde{F}_2$ [in (3.7)] with the turbulence in z which is strong enough to have a correlation length over a macroscopic scale. This is an interesting situation which opens a new way of statistically pursuing a turbulence in plasma.

In order to compute $\tilde{\rho}$ in (3.7) for a given M, it is necessary to solve the Fokker–Planck equation with many variables. A practically useful formula for solving the Fokker–Planck equation is presented in the next section for the simple case where the Riemannian space can be approximated by a Euclidean space. This formula tells us how to follow the generalized random walk with the aid of a high-speed computer to constitute the solution of our initial-value problem. It will turn out that the principle of the calculation is nothing but the Monte Carlo quadrature. This method of solution is, in principle, available also to the (not simple but) general case, for which we should consider the random walk on the Riemannian space. Indeed, since such a Riemannian space is always approximated by the tangential hyperplane locally, the infinitesimal random-walk motion around the tangential point can be followed by quite the same way as given for the simple case, and continuation of this process would complete a chain of random walk on the hypersurface.

Finally, it is interesting to point out that a new hierarchy similar to the BBGKY hierarchy can be obtained from our master equation. Let us multiply (3.1) by $z(x)$ and integrate it with respect to δz over the Riemannian space, taking into account (2.10), (2.14), and (2.15). Then, we have

$$\frac{\partial \tilde{F}_1}{\partial t} = [H_1; \tilde{F}_1] + n \int_X dx' [\phi; \tilde{F}_2] \tag{3.8}$$

in the limit $M \to \infty$. Here use was made of partial integration with respect to δz and the closure property of $\{s_j(x)\}$. This is formally the same as the first equation of the BBGKY hierarchy. However, if we multiply (3.1) by $z(x)z(x')$ and integrate it to obtain the second equation which governs $\tilde{F}_2$, we easily find that the equation obtained is not the same as the second equation of the BBGKY hierarchy because now the second term of (3.1) comes into play to give an effect of the modification caused by the operator $\mathfrak{F}$. The same is true for the higher-order equations of the hierarchy. Thus, $\tilde{F}_2, \tilde{F}_3, \cdots$ cannot be time reversible, so that $\tilde{F}_1$ is also time irreversible. This is an essential difference between (3.8) and the corresponding equation for F_1. From the viewpoint of the new hierarchy, it is no puzzle that the proper irreversible approximation to $\tilde{F}_2$ in (3.8) as seen in Born and Green's[15] work gives rise to the Boltzmann equation with irreversibility. One may see a coarse-grained aspect[16] of $\tilde{F}_1$ from this

[14] According to Shannon, in C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, Ill., 1964).

[15] M. Born and H. S. Green, Proc. Roy. Soc. (London) **A188**, 10 (1946).

[16] J. G. Kirkwood, J. Chem. Phys. **15**, 72 (1947); J. G. Kirkwood and J. Ross, in *Transport Processes in Statistical Mechanics*, I. Prigogine, Ed. (Interscience Publishers, Inc., New York, 1958), Vol. I; N. G. Van Kampen, in *Fundamental Problems in Statistical Mechanics*, E. G. D. Cohen, Ed. (North-Holland Publ. Co., Amsterdam, 1962).

fact, but we will come back to this matter in the final subsection of Sec. 4.

4. APPLICATION TO THE WEAK-INTERACTION CASE

By the weak-interaction case, we mean the case where the total interaction energy between particles is far less than the whole energy of the system, such as the case of a gas or sometimes a plasma. In this case, the Riemannian space reduces to the fixed hyperplane in A_1 which satisfies

$$n\int_X H_1(x)z(x)\frac{dx}{V} = e, \quad e = \text{const.} \tag{4.1}$$

This constitutes a subspace of A_1, which we may call B. Our problem thus reduces to solving, in general, the master equation (3.1) in B. Here we give the explicit formulation of the operator $\mathfrak{T}$ and next describe the practical program of how to calculate a physical quantity on the basis of our functional random-walk model with the use of a high-speed computer in mind.

A. Euclidean Coordinates in B

To represent a point in B in terms of a set of the Euclidean coordinates, we derive an orthonormal function set in B, starting from $\{s_j(x)\}$. The limiting process $V \to \infty$ is put off until we arrive at the final formula.

Now in the representation $z(x) = \sum b_j s_j(x)$, Eq. (2.13), defining A_1, is given as

$$\sum_j b_j \int_X s_j(x)\frac{dx}{V} = 1. \tag{4.2}$$

To find a rotation of R^M such that one of the basic vectors becomes normal to the hyperplane A_1, the transformation matrix (t_{ij}), such that

$$b_j = \sum_k b_k^{(1)} t_{kj} \quad \text{or} \quad b_k^{(1)} = \sum_j t_{kj} b_j \tag{4.3}$$

is introduced. $\{b_k^{(1)}\}$ are the new coordinates, and the normal coordinate to A_1 is taken as $b_1^{(1)}$. If we put $b_k^{(1)} = 0$, except for $k = 1$ and $b_1^{(1)} = 1$, then (4.3) gives

$$t_{1j} = \frac{\int_X s_j(x)\,dx/V}{\left\{\sum_k \left[\int_X s_k(x)\,dx/V\right]^2\right\}^{\frac{1}{2}}}, \tag{4.4}$$

since $\{b_j\}$ in this case are nothing but the direction cosines of a normal to A_1. This fixes a part of elements of the matrix (t_{ij}). The other elements, however, can be arbitrarily given except under the condition that (t_{ij}) should be an orthogonal matrix; this arbitrariness corresponds to the freedom of space rotation within A_1.

In the new coordinate system, the hyperplane can be expressed by $b_1^{(1)} = b_{10}^{(1)}$ (a certain const). This constant, which is the distance between the hyperplane and the origin of R^M, is found by the substitution of (4.3) into (4.2), together with $b_k^{(1)} = 0$ except for $k = 1$. Namely,

$$b_{10}^{(1)} = \left(\sum_j t_{1j}\int_X s_j(x)\frac{dx}{V}\right)^{-1} \quad \text{or} \quad \left[\sum_j \left(\int_X s_j(x)\frac{dx}{V}\right)^2\right]^{-\frac{1}{2}}. \tag{4.5}$$

Thus, a function in A_1 can be expressed as

$$z(x) = \sum_{j\geq 2} b_j^{(1)} s_j^{(1)}(x) + c_1(x), \tag{4.6}$$

where

$$s_j^{(1)}(x) = \sum_k t_{jk} s_k(x), \tag{4.7}$$

$$c_1(x) = b_{10}^{(1)} s_1^{(1)}(x). \tag{4.8}$$

In quite a similar way, we can find a rotation of A_1 such that one of the basic vectors becomes a normal to the hyperplane B. This time we note that the transformation matrix $(t_{ij}^{(1)})$ is of $(M-1)$ dimensions with the suffix conveniently beginning with 2. Corresponding to (4.4), we have

$$t_{2j}^{(1)} = \frac{\int H_1 s_j^{(1)}(x)\,dx/V}{\left\{\sum_{k\geq 2}\left[\int_X H_1 s_k^{(1)}(x)\,dx/V\right]^2\right\}^{\frac{1}{2}}}, \quad j \geq 2, \tag{4.9}$$

on account of (4.1). The distance between B and the origin of A_1 is given as

$$b_{20}^{(2)} = \frac{e - n\int H_1 c_1(x)\,dx/V}{n\left\{\sum_{j\geq 2}\left[\int_X H_1 s_j^{(1)}(x)\,dx/V\right]^2\right\}^{\frac{1}{2}}}. \tag{4.10}$$

As a result, a function in B is expressed as

$$z(x) = \sum_{j\geq 3} b_j^{(2)} s_j^{(2)}(x) + c_2(x) + c_1(x), \tag{4.11}$$

where

$$s_j^{(2)}(x) = \sum_{k\geq 1} t_{jk}^{(1)} s_k^{(1)}(x), \tag{4.12}$$

$$c_2(x) = b_{20}^{(2)} s_2^{(2)}(x). \tag{4.13}$$

Hence, we know that an orthonormal function set in B is given by (4.12), i.e.,

$$s_j^{(2)}(x) = \sum_{k\geq 2} t_{jk}^{(1)} \sum_{l\geq 1} t_{kl} s_l(x), \quad j \geq 3, \tag{4.14}$$

and $\{b_j^{(2)}\}$ is a set of the Euclidean coordinates based on it; but note that the space R^{M-2} spanned by (4.14) is not equal by itself but parallel to B, as is seen from (4.11).

B. The Explicit Form of $\mathfrak{F}$

In the frame of R^{M-2} now obtained, $z(x)$ in the master equation (3.1) should be replaced by

$$z(x) = \bar{z}(x) + c_1(x) + c_2(x); \tag{4.15}$$

here and hereafter a function with a bar on top indicates that it belongs to $\bar{B} \equiv R^{M-2}$ [spanned by (4.14)]. Differentiation is invariant to this replacement:

$$\delta/\delta z(x) = \delta/\delta \bar{z}(x).$$

In this frame, let us consider the functional quadratic form with the coefficient kernel function $[\phi; z(x)z(x')]$. After the orthogonal transformation in $\bar{B}$ to make it diagonal, it can be written as

$$\int_X \int_X \bar{y}(x)\bar{y}(x')[\phi; z(x)z(x')]\,dx\,dx' = \sum_{j\geq 3} a_j^{*2} D_j(z), \tag{4.16}$$

where a_j^* are the Euclidean coordinates of $\bar{y}(x)$; the (real) eigenvalues D_j depend on ϕ and z. Thus, the minimum modification of (4.16) to make it positive definite is simply to neglect $D_j(z)$ when $D_j(z) < 0$. This leads to the formula

$$\int_X \int_X \bar{y}(x)\bar{y}(x')\mathfrak{F}[\phi; z(x)z(x')]\,dx\,dx' = \sum_{j\geq 3} a_j^{*2} \tilde{D}_j(z), \tag{4.17}$$

where

$$\tilde{D}_j(z) = D_j(z), \quad D_j > 0,$$
$$= 0, \qquad D_j \leq 0. \tag{4.18}$$

Although this means just "completely prohibiting the anti-irreversible process from happening," as already described, an interesting relation with the coarse-graining idea will be explained in the final subsection.

C. General Solution for $\tilde{\rho}$

A general solution for $\tilde{\rho}$ can formally be constructed in the form of a repeated multiple integral by the propagation kernel method.[1] That is,

$$\tilde{\rho}(z, t) = \lim_{\Delta t \to 0} \int_B \cdots \int_B P_{\Delta t}(z^L/z^{L-1}) \cdots P_{\Delta t}(z^2/z^1) \times \tilde{\rho}(z^1, 0) \prod_{k=1}^{L-1} \delta \bar{z}^k, \tag{4.19}$$

where $z = z^L$, $\Delta t = t/L$, and the superscripts indicate the order of the time subintervals. The infinitesimal kernel $P_{\Delta t}$ is explicitly obtained from (3.1) as

$$P_{\Delta t}(z^{k+1}/z^k) = \int_B \exp\left\{ i \int_X \bar{y}^k(x)[\bar{z}^k(x) - \bar{z}^{k+1}(x) + \Delta t Q z^k(x)]\,dx - \tfrac{1}{2}\Delta t \int_X \int_X \bar{y}^k(x)\bar{y}^k(x') \times \mathfrak{F}[\phi; z^k(x)z^k(x')]\,dx\,dx' \right\} \delta \bar{y}^k, \tag{4.20}$$

and $\tilde{\rho}(z^1, 0)$ is the initial condition on $\tilde{\rho}$. It should be noted that the arguments in $\tilde{\rho}$ and $P_{\Delta t}$ are actually $\bar{z}^k$ rather than z^k, through (4.15) in the present frame of $\bar{B}$. The present notation of arguments is entirely for convenience.

The expression (4.20) may be rewritten in terms of the Euclidean coordinates for $\bar{y}^k(x)$:

$$\bar{y}^k(x) = \sum a_j^{k*} s_j^{k*}(x), \tag{4.21}$$

where $\{s_j^{k*}(x); j \geq 3\}$ is the orthonormal function set in $\bar{B}$ which is related to (4.14) by the orthogonal transformation described in the preceding subsection. By noting (4.17), we have

$$P_{\Delta t}(z^{k+1}/z^k) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left\{ i \sum_{j\geq 3} a_j^{k*} \times \left[\beta_j^{k*} - b_j^{k+1*} + \Delta t \int_X s_j^{k*}(x) Q z^k(x)\,dx \right] - \tfrac{1}{2}\Delta t \sum_{j\geq 3} (a_j^{k*})^2 \tilde{D}_j(z^k) \right\} \prod_{j\geq 3} \frac{da_j^{k*}}{(2\pi)^{\frac{1}{2}}}, \tag{4.22}$$

where

$$\beta_j^{k*} = \int_X s_j^{k*}(x)\bar{z}^k(x)\,dx \tag{4.23}$$

and

$$b_j^{k+1*} = \int_X s_j^{k*}(x)\bar{z}^{k+1}(x)\,dx, \tag{4.24}$$

which is further calculated as

$$P_{\Delta t}(z^{k+1}/z^k) = \prod_j [\Delta t \tilde{D}_j(z^k)]^{-\frac{1}{2}} \times \exp\left\{ -\left[\beta_j^{k*} + \Delta t \int_X s_j^{k*}(x) Q z^k(x)\,dx - b_j^{k+1*} \right]^2 [2\Delta t \tilde{D}_j(z^k)]^{-1} \right\}. \tag{4.25}$$

Here we note that the jth factor with $\tilde{D}_j = 0$, if any, reduces to the δ function

$$\delta\left(\beta_j^{k*} + \Delta t \int_X s_j^{k*}(x) Q z^k(x)\,dx - b_j^{k+1*} \right)$$

times $(2\pi)^{\frac{1}{2}}$.

$P_{\Delta t}$ in (4.25) is a generalized Gaussian measure factor with respect to $\{b_j^{k+1*}\}$, so that it may be interpreted as the probability density for the stochastic quantity $\{b_j^{k+1*}\}$. Since this probability depends on z^k and since $\{b_j^{k+1*}\}$ represents z^{k+1} in the manner of (4.24), it may be understood as the transition probability that a random walker in the space B (or $\bar{B}$) moves from the point z^k (or $\{b_j^{k*}\}$) to the point z^{k+1} (or $\{b_j^{k+1*}\}$) in a small time Δt. A great difference in our random walk from the simple Brownian motion is seen in the complicated dependence of $P_{\Delta t}$ on z^k, i.e., the starting point at each instant. As a consequence, the measure $\tilde{\rho}\delta\bar{z}$ established by (4.19) together with

(4.25) is far more complicated in shape than the Wiener measure. However, it is worth noting that the present master-equation approach has the great mathematical advantage over the original state-functional approach[1] in that the existence of the integral

$$\tilde{\psi} = \int_B \exp\left(i\int yz\,dx\right)\tilde{\rho}\delta\bar{z}$$

is guaranteed by establishment of the measure $\tilde{\rho}\delta\bar{z}$ even in the limit $M \to \infty$.[17] Hence, it is concluded that all $\tilde{F}_s$ converge as $M \to \infty$.

D. The Monte Carlo Quadrature

It is now obvious that, if (3.6) is adopted for $\tilde{\rho}(z_1, 0)$, (4.19) expresses the probability of finding the random walker, which started from $F_0(x)$ in B, at $z(x)$ after a continuous chain of random walk during the time t. The whole integration in (4.19), however, is analytically impossible. It is rather fortunate that the first integral with respect to z^1 can be calculated as

$$\begin{aligned}\int_B P_{\Delta t}(z^2/z^1)\tilde{\rho}(z^1, 0)\delta\bar{z}^1 &= \int_B P_{\Delta t}(z^2/z^1)\delta[\bar{z}^1(x) + c_1(x) + c_2(x) - F_0(x)]\delta\bar{z}^1 \\ &= P_{\Delta t}(z^2/F_0)\,. \qquad (4.26)\end{aligned}$$

But the next successive integrals with respect to $\bar{z}^2$, $\bar{z}^3, \cdots$, i.e., in the Euclidean coordinates $\{b_j^{2*}\}$, $\{b_j^{3*}\}, \cdots$, are involved enough to recommend the Monte Carlo quadrature with the use of a high-speed computer, although then we should be satisfied with an approximate value to the true integral.

It is nice for an importance sampling to be applied that the integrand itself constitutes a product of conditional probability densities. Then, the first task to estimate the second integral with respect to $\{b_j^{2*}\}$ is just to sample values for $\{b_j^{2*}\}$ out of the ensemble with the probability distribution $P_{\Delta t}(z^2/F_0)$ and to insert them through $z^2(x)$ into $P_{\Delta t}(z^3/z^2)$. [$\{s_j^{1*}(x)\}$ is known as a set of eigenvectors with the eigenvalues $\{\tilde{D}_j(F_0)\}$. Hence, $\bar{z}^2(x)$ is solved for by (4.24). Together with $\{\tilde{D}_j(z^2)\}$, a new set of eigenvectors $\{s_j^{2*}(x)\}$ is found and, hence, $\{\beta_j^{2*}\}$ is solved for.] Thus, the probability for $\{b_j^{3*}\}$ is decided, and so the same sampling process can be done for $\{b_j^{3*}\}$ and further for $\{b_j^{4*}\}, \cdots$ until we arrive at $\{b_j^{L-1*}\}$, when one round of the importance sampling for the whole integral (4.19) finishes. The next task is the average over many rounds of sampling.

If we wish to estimate $\tilde{F}_s$ by the formula (3.7), another sampling is necessary for the integration with respect to $\{b_j^{L*}\}$. In this case, we only have as the estimator a function $z(x_1)\cdots z(x_s)$, made of the sampled values for $\{b_j^{L*}\}$, so that the average over many such sample functions may give $\tilde{F}_s(x_1, \cdots, x_s, t)$. This is the principle of the Monte Carlo quadrature for calculating $\tilde{F}_s$. Since (4.25) is a Gaussian probability distribution for $\{b_j^{k+1*}\}$, we can sample the values for $\{b_j^{k+1*}\}$ out of normal random numbers with proper variances and averages. It is interesting to note that each round of the importance sampling simulates nothing but each chain of random walk during the time t in a segmented way. In the limit when $\Delta t \to 0$ or $L \to \infty$, the simulation becomes perfect. Also, it is expected that the other approximations included in the present procedure are improved with the values of M and V increasing independently.

E. The Relation with the Coarse-Graining Idea

We may point out a slight relation of our master-equation approach with the coarse-graining idea. The prescription of (4.17) and (4.18) may also be understood to limit the space for $\bar{y}(x)$ into a smaller subspace than R^{M-2} by neglecting all the jth eigenvectors of the space when $D_j \leq 0$. This means that $\bar{y}(x)$ has been coarse-grained by losing many orders of orthonormal functions which are not desirable for the irreversible description of the system. However, this coarse-graining procedure is not so fixed an operation as the conventionally understood one,[16] but is very flexible because it depends on z. If we proceed so thoroughly with the physically motivated coarse-graining idea that we artificially change the domain of integration in (4.22) from R^{M-2} to the subspace described above, the transition probability in (4.25) will be the product of only the jth factors with $D_j(z^k) > 0$. Then, b_j^{k+1} with the same j as that of the neglected eigenvector in $\bar{y}^k(x)$ does not appear in $P_{\Delta t}$, so that $\bar{z}^{k+1}(x)$ and, hence, $z^{k+1}(x)$ would be coarse-grained in the same way (depending on z^k) as $\bar{y}^k(x)$. Hence, we can coarse-grain the distribution function $\tilde{F}_s$ (in a flexible sense) through the relation (3.7). The above discussion, based on the artificial change of the domain of integration, is interesting in suggesting the physical meaning of our master-equation approach to some degree, but obviously such an artifice is strictly inexact. Therefore, we can only see by this discussion a roughly sketched, probable physical picture of our theory. All that we may say is that some very flexible kind of averaging process must have been introduced as a result of the operation $\mathfrak{F}$.

5. CONCLUSION

A minimum modification of the functional formalism of classical nonequilibrium statistical mechanics

[17] P. R. Halmos, *Measure Theory* (D. Van Nostrand Co., Princeton, N.J., 1950).

was performed (on the basis of the cylinder functional approach) such that the basic equation turns into the master equation, which includes the irreversible system evolution. The general theory was applied to the case with a gas or a weakly interacting plasma. The method of solution proposed in Sec. 4 has no great technical difficulty but for the use of a high-speed computer. All calculations with the Monte Carlo quadrature will be rather simple for the machine. It may be said that the possibility suggested in Ref. 1 of approaching nonequilibrium statistical mechanics by the Monte Carlo quadrature has been clarified in this paper. It is expected that this method will clarify many unsolved problems in relation with the irreversible process of the many-particle system, even for the case with spatial inhomogeneity. If we are interested in a steady state, it can be studied by examining an asymptotic behavior of the solution as $t \to \infty$ for an arbitrary type of initial condition. Together with the tangential approximation to the Riemannian space, the method of solution can extensively be used for the general strong-interaction case with an additional analytical-geometric care. In these respects, the practical value of the present theory depends to a large extent on future works employing a computer. It is interesting to note that a similar type of functional integral (with the Gaussian measure), which corresponds to a solution for the characteristic functional for the Burgers model turbulence, was recently calculated by the Monte Carlo quadrature with promising success.[18]

ACKNOWLEDGMENTS

The author would like to express his hearty thanks to Dr. J. Herring for kindly reading the manuscript and improving the author's original English. He is also indebted to Dr. T. G. Northrop and other members of Laboratory for Theoretical Studies, Goddard Space Flight Center, for their helpful discussions and comments. The work was supported by the National Academy of Sciences.

APPENDIX

Since $z(x)$ has been interpreted as the stochastic particle-number density, it is undesirable on a physical basis that $\tilde{\rho}$ does not vanish in the domain of A_1 where $z(x)$ is not nonnegative. First, it is natural that we restrict an initial value of $\tilde{\rho}$ to the *desirable* class in which all $\tilde{\rho}$ may be positive only for an nonnegative $z(x)$. The initial condition (3.6) belongs to this class. Second, what we need is to prove that $\tilde{\rho}$ in the desirable class never evolves outside the class. This is formally possible as follows.

As is seen from (4.19), $\tilde{\rho}$ at time t may be given as the population density of all possible random-walker locations at t in the subspace of A_1 with the condition (3.4). If there is no diffusion term in (3.1), which means all $\tilde{D}_j = 0$ in (4.17) and (4.25), the so-called random-walk trajectories are no longer random and their differential path in Δt becomes the projection on the subspace of A_1 [with (3.4)] of the differential evolution of $z(x)$ subject to the Vlasov equation. Since the Vlasov equation does not lead to any negative value in solution for a nonnegative initial value, our trajectories remain in the *desirable* domain of A_1 where $z(x)$ is nonnegative, if they were initially inside. Therefore, the possibility of a trajectory's migrating outside may only arise through the diffusive action because of a nonvanishing value of (4.17).

Here, let us take the more direct representation of functions y and z which has been explained in the second paper of Ref. 1; i.e., any functional of y and z can be considered as a function of $\{y(x_k)(\Delta x_k)^{\frac{1}{2}}\}$ and $\{z(x_k)(\Delta x_k)^{\frac{1}{2}}\}$, respectively. [We divide X into many cells and denote the representative value and the volume of the kth cell by x_k and Δx_k, respectively. The representation $\{a_j\}$ is obtained from $\{y(x_k)(\Delta x_k)^{\frac{1}{2}}\}$ by a special transformation.] Then, (4.17) is expressed in the form

$$\sum_{k,m} \bar{y}(x_k)(\Delta x_k)^{\frac{1}{2}} \bar{y}(x_m)(\Delta x_m)^{\frac{1}{2}} R(z)_{km}, \qquad \text{(A1)}$$

where $R(z)_{km}$ is the matrix element which depends on z and its (first-order) derivative at $x = x_k$ and x_m. The most interesting diffusive action is that which occurs in the neighborhood of points in X where $z(x) = 0$. If the eigenvalues of $R(z)_{km}$ just like $\tilde{D}_j$ do not vanish for these points, our random walk may at the next instant yield a negative $z(x)$ there. [See (4.25).] However, it is obvious that

$$R(z)_{km} = 0, \qquad \text{(A2)}$$

if either $z(x_k)$ or $z(x_m) = 0$, because the first-order derivative of $z(x)$ with respect to x should also be zero at the points where $z(x) = 0$ so long as $z(x)$ is analytic enough and belongs to the desirable domain of A_1. Equation (A2) just shows that the eigenvalue for the coordinate $\bar{y}(x_k)(\Delta x_k)^{\frac{1}{2}}$ should vanish if $z(x_\alpha) = 0$, so that we have no diffusive action around the points in X where $z(x) = 0$. Namely, it never happens that a trajectory starting from within the desirable domain of A_1 migrates to a $z(x)$ with a negative value somewhere in X. This concludes the proof.

[18] I. Hosokawa, Phys. Fluids **11**, 2052 (1968).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Solution of the Differential Equation

$$\left(\frac{\partial^2}{\partial x \partial y} + ax\frac{\partial}{\partial x} + by\frac{\partial}{\partial y} + cxy + \frac{\partial}{\partial t}\right)P = 0$$

Subject to the Initial Condition

$$P(x, y, 0) = \Phi(x, y)$$

M. E. GOLDSTEIN
Lewis Research Center, Cleveland, Ohio

(Received 2 July 1969)

A simple form of the solution to the differential equation

$$\left(\frac{\partial^2}{\partial x \partial y} + ax\frac{\partial}{\partial x} + by\frac{\partial}{\partial y} + cxy + \frac{\partial}{\partial t}\right)P = 0,$$

subject to the initial condition

$$P(x, y, 0) = \Phi(x, y),$$

is given.

1. INTRODUCTION

Lambropoulos[1] derived a solution to the initial-value problem described above, which is most conveniently applied when the initial data $\Phi(x, y)$ can easily be expressed as a Taylor series about the point $(x = 0,\ y = 0)$. However, the solution which he obtained is rather cumbersome. This situation was recently alleviated by Neuringer[2] who obtained a solution to this initial-value problem in a reasonably compact closed form. Neuringer's solution, however, only converges if $e^{-\lambda xy}\Phi(x, y)$ has a 2-dimensional Fourier transform [λ is defined by Eq. (4) below]. It will be shown herein that a solution which is useful when $e^{-\lambda xy}\Phi(x, y)$ is expressible as a Taylor series and not necessarily as a Fourier integral can also be obtained in a simple form. Both the method developed herein, and that developed by Neuringer,[2] can be readily adapted to obtain solutions to combined initial and boundary-value problems associated with the equation. Since the form of the solution obtained herein is quite different from that given by Neuringer, it is likely that this form of the solution may be more useful than that given by Neuringer for certain combined initial and boundary-value problems.

2. SOLUTION

Let a, b, and c be constants. A formal solution to the differential equation

$$\frac{\partial^2 P}{\partial x \partial y} + ax\frac{\partial P}{\partial x} + by\frac{\partial P}{\partial y} + cxyP + \frac{\partial P}{\partial t} = 0, \quad (1)$$

subject to the initial condition

$$P(x, y, 0) = \Phi(x, y), \quad (2)$$

will be sought.

To this end, introduce a new dependent variable Q defined by

$$P(x, y, t) = e^{\lambda(xy-t)}Q(x, y, t), \quad (3)$$

where λ is given by Neuringer's[2] equation (6), i.e.,

$$\lambda = \tfrac{1}{2}\{-(a + b) \pm [(a + b)^2 - 4c]^{\frac{1}{2}}\}, \quad (4)$$

and either sign may be used.

Then Q is a solution of the differential equation

$$\frac{\partial Q}{\partial t} + (a + \lambda)x\frac{\partial Q}{\partial x} + (b + \lambda)y\frac{\partial Q}{\partial y} + \frac{\partial^2 Q}{\partial x \partial y} = 0, \quad (5)$$

subject to the initial condition

$$Q(x, y, 0) = e^{-\lambda xy}\Phi(x, y). \quad (6)$$

In order to obtain a solution to Eq. (5) by separation of variables, introduce the new dependent variables ξ, η, and τ defined by

$$\xi = xe^{-(a+\lambda)t}, \quad (7)$$

$$\eta = xy\beta/(e^{\beta t} - 1), \quad (8)$$

$$\tau = -\beta t + \ln(e^{\beta t} - 1), \quad (9)$$

where

$$\beta = a + b + 2\lambda. \quad (10)$$

Upon introducing (7), (8), and (9) into Eq. (5), we find

$$\xi Q_{\xi\eta} + \eta Q_{\eta\eta} + (1 - \eta)Q_\eta + Q_\tau = 0. \quad (11)$$

[1] P. Lambropoulos, J. Math. Phys. **8**, 2167 (1967).
[2] J. Neuringer, J. Math. Phys. **10**, 250 (1969).

This equation is now separable; hence, we seek a solution in the form

$$Q(\xi, \eta, \tau) = T(\tau)N(\eta)Z(\xi). \tag{12}$$

Upon substituting this into Eq. (11) and separating variables, we find

$$\frac{dT}{d\tau} = -\sigma T, \tag{13}$$

$$\xi \frac{dZ}{d\xi} = -\gamma Z, \tag{14}$$

$$\eta \frac{d^2N}{d\eta^2} + \frac{dN}{d\eta}(1 - \gamma - \eta) - \sigma N = 0, \tag{15}$$

where σ and γ are the separation constants. The solutions to these ordinary differential equations are

$$T(\tau) = e^{-\sigma\tau}, \tag{16}$$

$$Z(\xi) = \xi^{-\gamma}, \tag{17}$$

$$N(\eta) = A_{\sigma\gamma}\,{}_1F_1(\sigma; 1 - \gamma; \eta) + B_{\sigma\gamma}\eta^{\gamma}\,{}_1F_1(\sigma + \gamma; 1 + \gamma; \eta), \tag{18}$$

where $A_{\sigma\gamma}$ and $B_{\sigma\gamma}$ are arbitrary constants and ${}_1F_1$ is the Pochhammer–Barnes confluent hypergeometric function.

Hence, reverting to the original independent variables, the elementary solutions to Eq. (5) can be written as

$$Q = (\bar{A}_{s\gamma}x^{\gamma}e^{-\gamma(a+\lambda)t} + \bar{B}_{s\gamma}y^{\gamma}e^{-\gamma(b+\lambda)t}) \times e^{-s\beta t}(e^{\beta t} - 1)^s\,{}_1F_1\left(-s; 1 + \gamma; \frac{xy\beta}{e^{\beta t} - 1}\right), \tag{19}$$

where we have put $s = -\sigma$ and $\gamma = -\gamma$ to obtain the first term and $s = -(\sigma + \gamma)$, keeping γ fixed, to obtain the second term. The most general solution to any initial and boundary value problem for Eq. (5) can presumably be obtained by superimposing the elementary solutions (19) for suitable values of γ and s. However, we are interested here in the solution to an initial-value problem, and for this purpose let the values of s and γ be the nonnegative integers. Now for $s = 0, 1, 2, \cdots$ the series representing ${}_1F_1(-s; 1 + \gamma; \eta)$ terminates after a finite number of terms, and we have, in fact,

$${}_1F_1(-s; 1 + \gamma; \eta) = [s!/(1 + \gamma)_s]L_s^{(\gamma)}(\eta), \tag{20}$$

where $L_s^{(\gamma)}$ is the generalized Laguerre or Sonine polynomial, and

$$(1 + \gamma)_s = \prod_{k=1}^{s}(\gamma + k - 1). \tag{21}$$

Hence, Eq. (19) shows that the solution to Eq. (5) is in this case

$$Q = \sum_{s,k=0}^{\infty}[C_{s,k}^{(1)}x^k e^{-k(a+\lambda)t} + C_{s,k}^{(2)}y^k e^{-k(b+\lambda)t}] \times (1 - e^{-\beta t})^s L_s^{(k)}\left(\frac{xy\beta}{e^{\beta t} - 1}\right), \tag{22}$$

where $C_{s,k}^{(1)}$ and $C_{s,k}^{(2)}$ are arbitrary constants. Now, for $t = 0$,

$$(e^{\beta t} - 1)^s L_s^{(k)}\left(\frac{xy\beta}{e^{\beta t} - 1}\right) = \frac{(xy\beta)^s}{s!}.$$

Hence,

$$Q(x, y, 0) = \sum_{s,k=0}^{\infty}\frac{\beta^s}{s!}(C_{s,k}^{(1)}x^{(k+s)}y^s + C_{s,k}^{(2)}y^{(k+s)}x^s). \tag{23}$$

Upon setting $C_{s,0}^{(1)} = C_{s,0}^{(2)}$ for $s = 0, 1, 2, \cdots$, and defining $p_{r,n}$ by

$$p_{(k+s),s} = (k + s)!\,\beta^s C_{s,k}^{(1)},$$
$$p_{s,(k+s)} = (k + s)!\,\beta^s C_{s,k}^{(2)},$$
$$p_{s,s} = 2C_{s,0}^{(1)}\beta^s s! = 2s!\,\beta^s C_{s,0}^{(2)},$$

where $k = 1, 2, 3, \cdots$ and $s = 0, 1, 2, \cdots$, we find after some manipulation that

$$Q(x, y, 0) = \sum_{s,r=0}^{\infty}\frac{p_{r,s}}{s!\,r!}x^r y^s. \tag{24}$$

Hence,

$$p_{r,s} = \left(\frac{\partial^{(r+s)}Q(x, y, 0)}{\partial x^r \partial y^s}\right)_{x=0,y=0} = \left(\frac{\partial^{(r+s)}e^{-\lambda xy}\Phi}{\partial x^r \partial y^s}\right)_{x=0,y=0} \tag{25}$$

and

$$P = e^{\lambda(xy-t)}\sum_{s=0}^{\infty}\frac{1}{s!}p_{s,s}\left(\frac{1 - e^{-\beta t}}{\beta}\right)^s L_s\left(\frac{xy\beta}{e^{\beta t} - 1}\right) + e^{\lambda(xy-t)}\sum_{k=1}^{\infty}\sum_{s=0}^{\infty}\frac{1}{(k + s)!}(p_{(k+s),s}x^k e^{-k(a+\lambda)t} + p_{s,(k+s)}y^k e^{-k(b+\lambda)t})\left(\frac{1 - e^{-\beta t}}{\beta}\right)^s L_s^{(k)}\left(\frac{xy\beta}{e^{\beta t} - 1}\right), \tag{26}$$

where $L_s = L_s^{(0)}$ is the simple Laguerre polynomial.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Many-Body Point Transforms*

NORMAN M. WITRIOL

Brandeis University, Waltham, Massachusetts 02154 *and Solid State Physics Branch, Physical Sciences Laboratory, Redstone Arsenal, Alabama* 35809

(Received 9 June 1969)

The method of point transforms is extended to the consideration of the truly many-body terms of many-body systems. The point transformation is taken to be regional (one body isolated, two bodies interacting, three bodies interacting, $\cdots$, N bodies interacting) and continuous in the $3N$-dimensional transformed space. The transformed Hamiltonian and wavefunction of an N-body system can then be easily approximated by one-body, two-body, three-body, $\cdots$, or N-body additivity. As an example of the use of this method, a set of regional point transforms is developed for the short range, strongly repulsive, or hard-core many-body problem. Applying these transforms to the many-body Hamiltonian, we obtain a set of equivalent Hermitian Hamiltonians which are Fourier analyzable and amenable to ordinary perturbation and variational techniques. Assuming two-body additivity, this set is shown to be the set of pairwise point transformations developed by Eger and Gross.

1. INTRODUCTION

In this paper we extend the method of point transforms to include the truly many-body terms of many-body systems. The basis of the method is to perform a classical canonical transformation on the variables of the system and then quantize these transformed variables. The requirement for this transformation to be equivalent to a unitary transformation of the original quantized variables is that the canonical transformation can be generated by a series of infinitesimal canonical transformations from the Cartesian variables.[1] For simplicity we choose time-independent point transforms.

Previous investigations[2-6] using this method have assumed pairwise additivity, wherein the system is assumed to be sufficiently dilute to enable one to neglect configurations in which more than two particles interact. The original Hamiltonian can then be approximated by the sum of pairwise interactions

$$H = \sum_{\substack{i,j=1 \\ i<j}}^{N} H_{ij}, \tag{1.1}$$

where

$$H_{ij} = \frac{1}{2m}\sum_{\alpha=1}^{3}(P_{i\alpha}^2 + P_{j\alpha}^2) + V(\mathbf{R}_{ij}, \mathbf{P}_{ij}). \tag{1.2}$$

Our notation for equal-mass particles shall be

$$X_{ij\alpha} = X_{i\alpha} - X_{j\alpha}, \quad x_{ij\alpha} = x_{i\alpha} - x_{j\alpha}, \tag{1.3}$$

$$P_{ij\alpha} = \tfrac{1}{2}(P_{i\alpha} - P_{j\alpha}), \quad p_{ij\alpha} = \tfrac{1}{2}(p_{i\alpha} - p_{j\alpha}), \tag{1.4}$$

where the $X_{i\alpha}$ (the $\mathbf{R}_i$) are the original coordinates, the $x_{i\alpha}$ (the $\mathbf{r}_i$) are the transformed coordinates, the $P_{i\alpha}$ are the original momenta, the $p_{i\alpha}$ are the transformed momenta, Latin indices stand for particle numbers and run from 1 to N, and Greek indices stand for the Cartesian coordinates with the range 1 to 3. Under the pairwise point transformation

$$X_{ij\alpha} = x_{ij\alpha}(1 + f(\mathbf{r}_{ij})), \tag{1.5}$$

$$P_{ij\alpha} = \frac{1}{2}\sum_{\beta=1}^{3}\left(\frac{\partial x_{ij\beta}}{\partial X_{ij\alpha}}\, p_{ij\beta} + p_{ij\beta}\frac{\partial x_{ij\beta}}{\partial X_{ij\alpha}}\right), \tag{1.6}$$

each H_{ij} is transformed into

$$H'_{ij} = \sum_{\alpha=1}^{3}\left(\frac{1}{2m}(p_{i\alpha}^2 + p_{j\alpha}^2) + \frac{1}{4m}\sum_{\beta=1}^{3} p_{ij\alpha}\mathfrak{L}^{\alpha\beta}(\mathbf{r}_{ij})p_{ij\beta}\right) + W(\mathbf{r}_{ij}) + V(\mathbf{r}_{ij}, \mathbf{p}_{ij}), \tag{1.7}$$

where

$$\mathfrak{L}^{\alpha\beta}(\mathbf{r}_{ij}) = g_{i\alpha j\beta} - \delta_{ij}\delta_{\alpha\beta}, \quad g_{i\alpha j\beta} = \sum_{\gamma=1}^{3}\frac{\partial x_{ij\alpha}}{\partial X_{ij\gamma}}\frac{\partial x_{ij\beta}}{\partial X_{ij\gamma}}, \tag{1.8}$$

$$W(\mathbf{r}_{ij}) = \frac{\hbar^2}{m}\sum_{\alpha,\beta=1}^{3}\tfrac{1}{4}g_{i\alpha j\beta}\frac{\partial \ln B}{\partial x_{ij\alpha}}\frac{\partial \ln B}{\partial x_{ij\beta}} - \frac{1}{2}\frac{\partial}{\partial x_{ij\alpha}}g_{i\alpha j\beta}\frac{\partial \ln B}{\partial x_{ij\beta}}, \tag{1.9}$$

B is the Jacobian $|\partial x_{i\alpha}/\partial X_{j\beta}|$ of the inverse transformation ($\mathbf{r} \to \mathbf{R}$), and $V(\mathbf{r}_{ij}, \mathbf{p}_{ij})$ is $V(\mathbf{R}_{ij}, \mathbf{P}_{ij})$ written in terms of $\mathbf{r}_{ij}$ and $\mathbf{p}_{ij}$. In addition, the original normalized wavefunction $\psi_{ij}(\mathbf{R}_{ij})$ is related to the transformed normalized wavefunction $\psi'_{ij}(\mathbf{r}_{ij})$ by

$$\psi_{ij}(\mathbf{R}_{ij}(\mathbf{r}_{ij})) = [B(\mathbf{r}_{ij})]^{\frac{1}{2}}\psi'_{ij}(\mathbf{r}_{ij}). \tag{1.10}$$

$W(\mathbf{r})$ represents a new coordinate dependent potential and $p\mathfrak{L}p$ represents a new coordinate-and momentum-dependent potential. The range in coordinate space

* This research is based upon a portion of a dissertation submitted to Brandeis University, Waltham, Massachusetts (1968) in partial fulfillment of the requirements for the Ph.D. degree, and supported in part by the U.S. Air Force Office of Scientific Research.

[1] This requirement excludes, for example, the transformation to polar coordinates.

[2] F. M. Eger and E. P. Gross, Ann. Phys. (N.Y.) **24**, 63 (1963).

[3] F. M. Eger and E. P. Gross, Nuovo Cimento **34**, 1225 (1964).

[4] F. M. Eger and E. P. Gross, J. Math. Phys. **6**, 891 (1965).

[5] F. M. Eger and E. P. Gross, J. Math. Phys. **7**, 578 (1966).

[6] N. Witriol, thesis, Brandeis University, Waltham, Mass., 1968.

of W and $\mathfrak{L}$ is determined by the transformation function f. For the pairwise assumption to be reasonable in the transformed space, we must require

$$f \to 0, \quad r \to r_0, \\ = 0, \quad r > r_0, \tag{1.11}$$

where r_0 is the average interparticle separation.[7] The Hamiltonian H is then transformed into the Hermitian Hamiltonian H',

$$H' = \sum_{\substack{i,j=1 \\ i<j}}^{N} H'_{ij} \tag{1.12}$$

$$= \frac{1}{2m}\sum_{i=1}^{N}\sum_{\alpha=1}^{3} p_{i\alpha}^2 + \sum_{\substack{i,j=1 \\ i<j}}^{N}\left(\frac{1}{4m}\sum_{\alpha,\beta=1}^{3} p_{ij\alpha}\mathfrak{L}^{\alpha\beta}(\mathbf{r}_{ij})p_{ij\beta} + W(\mathbf{r}_{ij}) + V(\mathbf{r}_{ij}, \mathbf{p}_{ij})\right). \tag{1.13}$$

The transformed Hamiltonian H' is a function of the transformation function $f(\mathbf{r})$. Therefore, Eq. (1.13) represents a set of Hermitian Hamiltonians. Each element H' is specified by a particular $f(\mathbf{r})$ and is equivalent, to the extent of the validity of the pairwise approximation in the transformed space, to the original Hamiltonian H. By properly choosing $f(\mathbf{r})$, we can include part of the interaction in the transformation and thereby weaken the interaction term dealt with by means of perturbation theory. For example, if the potential $V(\mathbf{R}, \mathbf{P})$ in Eq. (1.2) contains a hard-core (hc) interaction part,

$$V_{\mathrm{hc}}(R) = \infty, \quad R < c, \\ = 0, \quad R > c, \tag{1.14}$$

the class of functions

$$f = c/r\phi(r), \tag{1.15}$$

where

$$\phi(r) = 1, \quad r = 0, \\ \to 0, \quad r \to r_0, \\ = 0, \quad r > r_0, \tag{1.16}$$

removes it[8]—i.e., from Eq. (1.5) we obtain $V_{\mathrm{hc}}(r) = 0$ over all space. The effects of $V_{\mathrm{hc}}(R)$ are now contained in W, $\mathfrak{L}$, and B. The advantage of this method is that W, $\mathfrak{L}$, and $V(r)$, are Fourier analyzable. The transformed Hamiltonian H' is therefore Fourier analyzable, Hermitian, and amenable to ordinary perturbation and variational techniques while being equivalent to the original Hamiltonian. Taking the Fourier transform of H', and introducing the standard second quantization formalism, one can obtain a convenient form of the Hamiltonian for calculating the properties of low-temperature quantum fluids.[5] Previous investigations[3,5,6] have characterized f, and thus ϕ, W, $\mathfrak{L}$, and B by a range λ where $c < \lambda < r_0$. The dilute gas limit is reached by letting λ be a fixed small fraction of r_0 and taking $c/\lambda \to 0$. The ground-state energy and the low-level excitations of the hard-core Bose liquid have been calculated[5] and found to agree in the dilute gas limit with the low-density results of Lee, Huang, and Yang.[8] Similarly, the ground-state energy and the Landau parameters of the hard-core Fermi liquid have been calculated[6] and found to agree in the dilute gas limit with the series expansion in $c\rho^{\frac{1}{3}}$ (ρ being the density of the system) of Huang and Yang,[9] and Lee and Yang[10] for the ground-state energy, and with the series expansions in $c\rho^{\frac{1}{3}}$ of Abrikosov and Khalatnikov[11] for the Landau parameters. The λ-dependent terms, ignored in the dilute gas limit, represent higher-density corrections to these dilute-gas results. In addition, Cooper[12] has applied this method to the classical hard-core problem and obtained the standard results for the virial coefficients.

Let us now consider the pairwise assumption in relation to some physical systems of interest: liquid ^{3}He, ^{4}He, and nuclear matter. The parameters c/r_0 and λ/r_0 for these systems are (with $\frac{4}{3}\pi r_0^3 = \rho^{-1}$)

$$(c/r_0)_{^3\mathrm{He}} = (1.9/2.44) = 0.78 < \lambda/r_0 < 1, \\ (c/r_{0^4\mathrm{He}}) = (1.9/2.22) = 0.86 < \lambda/r_0 < 1, \\ (c/r_0)_{\substack{\mathrm{nuclear} \\ \mathrm{matter}}} = (0.4/1.7) = 0.37 < \lambda/r_0 \lesssim 1.$$

For the pairwise assumption to be good, the probability of three or more particles interacting with one another must be negligible. For the above systems, this assumption is poor.

In this paper we shall extend the point transformation method to the consideration of three-, four-, $\cdots$, and N-body interaction terms. For simplicity, we shall restrict ourselves to central potentials, $V(\mathbf{R}, \mathbf{P}) = V(R)$. We shall chiefly be concerned with strong short-range potentials, in particular with the hard-core potential, Eq. (1.14). The transformation shall be chosen to shrink the system, thereby reducing the range of the strong short-range potential. Since the effects of a potential $V(R)$ are proportional not only to its strength but also to its range, the effects of the

[7] Note, from Eq. (1.5) and Eq. (1.11), that r_0 is the interparticle separation in both the original and transformed spaces.

[8] T. D. Lee, K. Huang, and C. N. Yang, Phys. Rev. **106**, 1135 (1957).

[9] K. Huang and C. N. Yang, Phys. Rev. **105**, 767 (1957).

[10] T. D. Lee and C. N. Yang, Phys. Rev. **105**, 1119 (1957).

[11] A. A. Abrikosov and I. M. Khalatnikov, Zh. Eksp. Teor. Fiz. **33**, 1154 (1957) [Sov. Phys.—JETP **6**, 888 (1958)].

[12] M. J. Cooper, J. Math. Phys. **9**, 571 (1968).

transformed strong short-range potential $V(r)$ can thereby be made small. In particular, the transformed hard-core potential can be eliminated. The large effects of the strong repulsive potential in the original Hamiltonian will be contained in the new coordinate-dependent potential W and in the new coordinate- and momentum – dependent potential $p\mathfrak{L}p$. In Sec. 2, the point transforms are taken to be of the form

$$\begin{aligned} X_{i\alpha} = {} & \prod_{\substack{a=1\\a\neq i}}^{N} \theta(r_{ia}-b)x_{i\alpha} \\ & + \sum_{\substack{j=1\\j\neq i}}^{N}\left(\prod_{\substack{a=1\\a\neq i,j}}^{N}\theta(r_{ia}-b)\theta(r_{ja}-b)\right)\theta(b-r_{ij})\bar{A}_{ij\alpha} \\ & + \sum_{\substack{j,k=1\\k,j\neq i}}^{N}\left(\prod_{\substack{a=1\\a\neq i,j,k}}^{N}\theta(r_{ia}-b)\theta(r_{ja}-b)\theta(r_{ka}-b)\right) \\ & \times [\theta(b-r_{ik})\theta(b-r_{jk})\theta(r_{ij}-b) \\ & + \theta(b-r_{jk})\theta(b-r_{ij})\theta(r_{ik}-b) \\ & + \theta(b-r_{ij})\theta(b-r_{ik})\theta(r_{jk}-b) \\ & + \theta(b-r_{ij})\theta(b-r_{ik})\theta(b-r_{jk})]\bar{A}_{ijk\alpha} \\ & + \cdots + [\theta \text{ functions for the } N\text{-body} \\ & \quad (N, N-1, \cdots, 3, 2, 1) \text{ cluster} \\ & \quad \text{combinations}]\, \bar{A}_{iN(N-1)\cdots(i+1)(i-1)\cdots 321\alpha}, \end{aligned} \tag{1.17}$$

where b is the arbitrary range of the interaction in the transformed coordinates, subject to $c < b$. Equation (1.17) is a general expression for the transformation of $X_{i\alpha}$, differing from the identity transformation only when particle i is within distance b of (i.e., is interacting in the transformed space with) other particles in the system. We can consider only two-body interactions (the first two terms), or three-body interactions (the first three terms), $\cdots$, or N-body interactions (all N terms), according to our preference and the amount of algebraic manipulations we are willing to perform. Initially the functional forms of the $\bar{A}$'s are arbitrary, subject to the requirements that the transformation be continuous and one-to-one. An explicit set of $\bar{A}$'s, which can be specified to reduce the effects of strong short-range potentials, or to eliminate hard-core potentials, is then given. In Sec. 3, we show that in the limit of two-body interactions the resulting set of point transformations reduces to Eq. (1.5) with Eqs. (1.15), (1.16). In Sec. 4, the regional form of the transformed Hamiltonian resulting from Eq. (1.17) is developed. We shall find that, for potentials with range $c < b$, the transformed Hamiltonian H' can be expressed as

$$\begin{aligned} H' = {} & \frac{1}{2m}\sum_{i=1}^{N}\prod_{\substack{a=1\\a\neq i}}^{N}\theta(r_{ia}-b)\sum_{\alpha=1}^{3}p_{i\alpha}^2 \\ & + \sum_{\substack{i,j=1\\j<i}}^{N}\left(\prod_{\substack{a=1\\a\neq i,j}}^{N}\theta(r_{ia}-b)\theta(r_{ja}-b)\right)\theta(b-r_{ij})H'_{ij} \\ & + \sum_{\substack{i,j,k=1\\k<j<i}}^{N}\left(\prod_{\substack{a=1\\a\neq i,j,k}}^{N}\theta(r_{ia}-b)\theta(r_{ja}-b)\theta(r_{ka}-b)\right) \\ & \times [\theta(b-r_{ik})\theta(b-r_{jk})\theta(r_{ij}-b) \\ & + \theta(b-r_{jk})\theta(b-r_{ij})\theta(r_{ik}-b) \\ & + \theta(b-r_{ij})\theta(b-r_{ik})\theta(r_{jk}-b) \\ & + \theta(b-r_{ij})\theta(b-r_{ik})\theta(b-r_{jk})]H'_{ijk} \\ & + \cdots + [\theta \text{ functions for the } N\text{-body} \\ & \quad (N, N-1, \cdots, 3, 2, 1) \text{ cluster} \\ & \quad \text{combinations}]\, H'_{N(N-1)\cdots 321}. \end{aligned} \tag{1.18}$$

Thus, H' can be approximated by one-body additivity, i.e., no interactions (the first term), or two-body additivity (the first two terms), or three-body additivity (the first three terms), $\cdots$, or N-body additivity (all N terms). The number of terms kept in Eq. (1.18) is, of course, equal to or less than the number of terms kept in Eq. (1.17). Each $H'_{i_1\cdots i_n}$ term in Eq. (1.18) is Hermitian and Fourier analyzable, even when the original potential contains hard-core components. Therefore, the total Hamiltonian H' is Hermitian, Fourier analyzable, and amenable to ordinary perturbation and variational techniques. Thus, we shall have demonstrated the extension of the point transform method to the consideration of the truly many-body terms of N-body systems.

2. TRANSFORMATION

In this section we develop a set of point transforms whereby the effects of the transformed strong short-range potential of an N-body system can be made small, assuming one-, or two-, $\cdots$, or N-body interactions as desired. In particular, if the interparticle potentials contain hard-core parts, the set of point transforms can be chosen to eliminate these singular components.

We start by introducing an arbitrary distance b taken to be greater than or equal to the interaction radius in the original space c. We shall, in general,

find that for continuity we must require $b > c$. The transformation is chosen such that, when the interparticle separation R_{ij} in the original space is greater than b, R_{ij} equals the interparticle separation r_{ij} in the transformed space, and vice versa:

$$R_{ij} = r_{ij}, \quad \text{for} \quad (R_{ij}; r_{ij}) > b. \tag{2.1}$$

Therefore, b is the cutoff distance of the transformation of the interparticle separations. Thus, it is also the range of the interactions created by this transformation.

We define an n-body cluster to be a linked set of n particles, a link existing between particles i and j if the interparticle separation $(R_{ij}; r_{ij}) \leq b$. Diagrammatically, the first lower-order n-body clusters, $n = 1, 2, 3$, are

1-body (i): $(R_{ia}; r_{ia}) > b$, all $a \neq i$, $1 \leq a \leq N$,

i

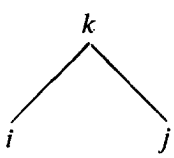
2-body (ij): $(R_{ia}, R_{ja}; r_{ia}, r_{ja}) > b$,

all $a \neq i, j$, $1 \leq a \leq N$,

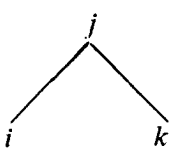
i ——— j $\quad (R_{ij}; r_{ij}) \leq b$,

3-body (ijk): $(R_{ia}, R_{ja}, R_{ka}; r_{ia}, r_{ja}, r_{ka}) > b$,

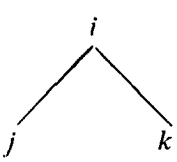
all $a \neq i, j, k$, $1 \leq a \leq N$,



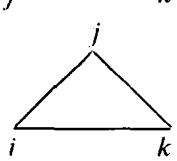
$(R_{ik}, R_{jk}; r_{ik}, r_{jk}) \leq b$, $(R_{ij}; r_{ij}) > b$,



$(R_{ij}, R_{jk}; r_{ij}, r_{jk}) \leq b$, $(R_{ik}; r_{ik}) > b$,



$(R_{ij}, R_{ik}; r_{ij}, r_{ik}) \leq b$, $(R_{jk}; r_{jk}) > b$,



$(R_{ik}, R_{jk}, R_{ij}; r_{ik}, r_{jk}, r_{ij}) \leq b.$

We note, from the three-body cluster, that there will be many combinations in each n-body cluster for $n \geq 3$. The transformation is defined such that it is the identity transformation for a one-body cluster, depends on $(\mathbf{r}_i, \mathbf{r}_j)$ [or $(\mathbf{R}_i, \mathbf{R}_j)$] for a two-body (ij) cluster, depends on $(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k)$ [or $(\mathbf{R}_i, \mathbf{R}_j, \mathbf{R}_k)$] for a three-body (ijk) cluster, etc. Let us now discuss the results we desire from the application of the transformation to the Hamiltonian of our system.

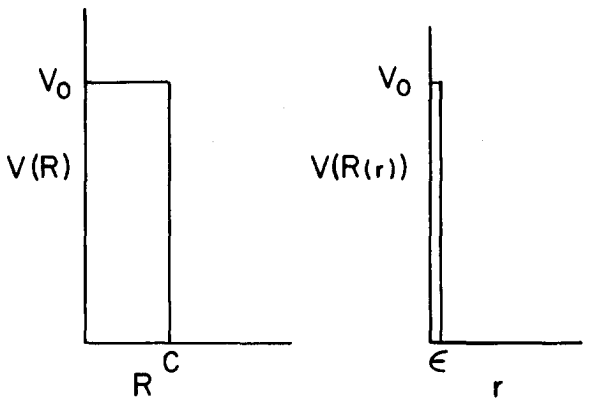


FIG. 1. Reduction (by the transformation) of the range of the strongly repulsive potential V_0.

We wish to reduce the effects of the transformed short-range, strongly repulsive or hard-core potentials. Since the effects of a potential are proportional to both its strength and its range, we can reduce its effects by reducing its range. Diagrammatically, the desired effect of the transformation is shown in Fig. 1. To achieve this goal the transformation for the clusters is defined in a manner such that each n-particle cluster $(i_1 \cdots i_n)$ shrinks about its center of mass $\mathbf{R}_{cm(i_1 \cdots i_n)}$, the shrinkage factor depending on the smallest interparticle distance. This shrinkage factor is so chosen that when the smallest interparticle separation R_{ij} is equal to the strong interaction radius c, the smallest transformed interparticle separation r_{ij} is equal to the transformed strong interaction radius ϵ, where ϵ is an arbitrarily chosen small distance. Applying the transformation to the cluster is akin to having Alice, in Lewis Carroll's *Alice's Adventures in Wonderland*, eat or drink substances that change her size—the shrinkage of the cluster depending upon the smallest interparticle separation vs the shrinkage of Alice, depending upon the amount of the substances she eats or drinks. For example, consider the transformation of the three-body cluster shown diagrammatically in Fig. 2. Take r_{ij} to be the smallest interparticle separation. Then,

$$R_{ij} = \alpha_{ij} r_{ij}, \quad R_{ik} = \alpha_{ij} r_{ik}, \quad R_{jk} = \alpha_{ij} r_{jk},$$

where the shrinkage proportion α_{ij}, a function of r_{ij}, is chosen as follows: when

$R_{ij} = c$, then $r_{ij} = \epsilon$ and $V(R_{ij}) = V(R_{ij}(r_{ij})) = V_0$;

$R_{ij} < c$, $\quad r_{ij} < \epsilon$ $\quad V(R_{ij}) = V(R_{ij}(r_{ij})) = V_0$;

$R_{ij} > c$, $\quad r_{ij} > \epsilon$ $\quad V(R_{ij}) = V(R_{ij}(r_{ij})) = 0$.

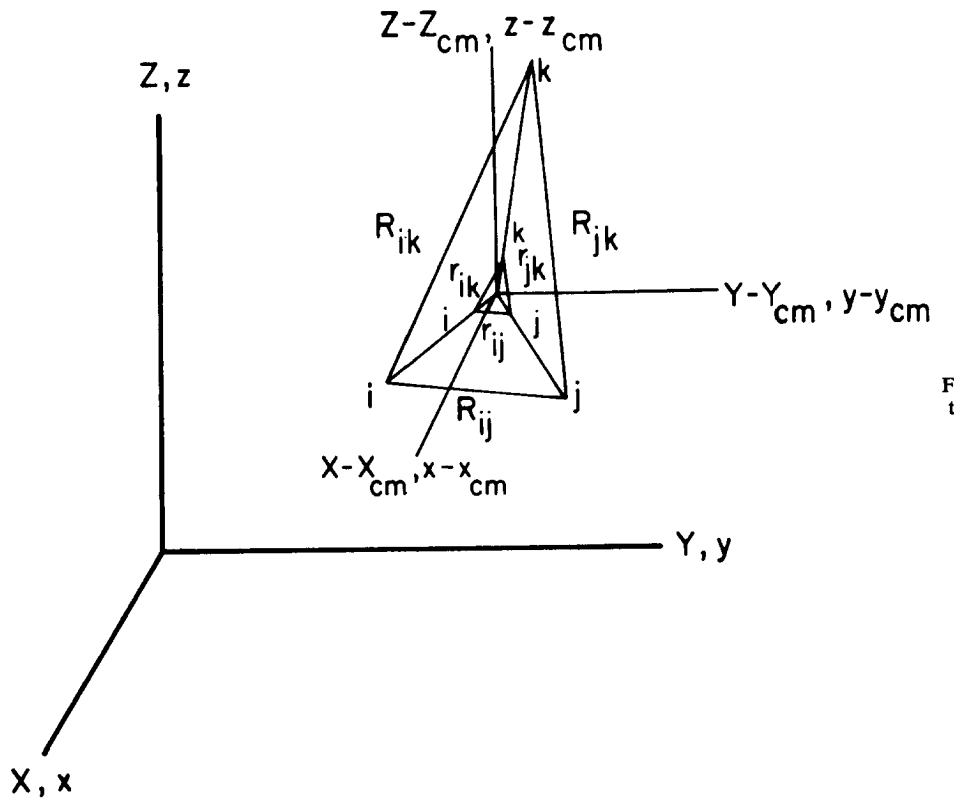


FIG. 2. Three-body (ijk) cluster transformation. $cm = cm(ijk)$.

Therefore, when

$R_{ij} < c$, $(R_{ik}, R_{jk}) > c$, the transformation gives

$$r_{ij} < \epsilon, (r_{ik}, r_{jk}) > \epsilon;$$

$$R_{ij} < R_{ik} < c, R_{jk} > c, \qquad r_{ij} < r_{ik} < \epsilon, r_{jk} > \epsilon;$$

etc.

Thus, the cluster transformation described achieves the desired result illustrated in Fig. 1. We denote the n-body cluster $(i_1 \cdots i_n)$ transformation of $X_{i_1\alpha}$ by $A_{i_1 \cdots i_n \alpha}$. We require the total transformation (the transformation of the $3N$ coordinates in the $3N$-dimensional space) to be one-to-one and the transformation for each set of particle coordinates $X_{i\alpha}$ to be continuous as the n-body cluster becomes an $(n+1)$-, $(n+2)$-, $\cdots$, or N-body cluster, or becomes an $(n-1)$-, $(n-2)$-, $\cdots$, or one-body cluster.

The transformation $A_{i_1 \cdots i_n \alpha}$, as described above, is not continuous at the boundaries between the cluster regions. To see this problem in detail, let us consider three bodies i, j, k. Take

$$R_{ik} = b - \delta, \quad R_{jk} = B > b, \quad R_{ij} = A \leq (R_{ik}, R_{jk}),$$

where δ is an arbitrarily small positive number and A and B are given distances. These particles thus form a three-body cluster (ijk). Under the transformation described, this three-body cluster shrinks about its center of mass, the shrinkage proportion depending upon r_{ij}. Let us now take

$$R_{ik} = b + \delta, \quad R_{jk} = B, \quad R_{ij} = A.$$

These three particles now form a two-body cluster (ij) and a one-body cluster (k). Under the transformation described, the two-body cluster shrinks about its center of mass, the shrinkage proportion depending upon r_{ij}. The one-body cluster is unchanged.[13] Thus, the transformation described is discontinuous at the boundary between these cluster regions. A generalization of this problem exists for all n-body clusters. To correct this deficiency, we introduce functions $h_{i_a i_b}$, whose roles are to make the transformation satisfy the following conditions for an n-body cluster $(i_1 \cdots i_n)$: (1) when none of the n particles are near a boundary of the cluster, the transformation of $X_{i_1\alpha}$ is essentially $A_{i_1 \cdots i_n \alpha}$; (2) when any of the n particles are near the boundary of the cluster, the transformation of $X_{i_1\alpha}$ approaches the

[13] The transformation describing the shrinkage of a one-body cluster about its center of mass is the identity transformation.

transformation of $X_{i_1\alpha}$ on the other side of the boundary. We are thereby led to the transformation

$$\begin{aligned}
X_{i\alpha} = &\left(\prod_{\substack{a=1\\a\neq i}}^{N}[\theta(r_{ia}-b)\right)[A_{i\alpha}] \\
&+\sum_{\substack{j=1\\j\neq i}}^{N}\left(\prod_{\substack{a=1\\a\neq i,j}}^{N}(\theta r_{ia}-b)\theta(r_{ja}-b)\right) \\
&\times[\theta(b-r_{ij})][A_{ij\alpha}(1-h_{ij})+A_{i\alpha}h_{ij}] \\
&+\sum_{\substack{j,k=1\\k<j\\k,j\neq i}}^{N}\left(\prod_{\substack{a=1\\a\neq i,j,k}}^{N}\theta(r_{ia}-b)\theta(r_{ja}-b)\theta(r_{ka}-b)\right) \\
&\times[\theta(b-r_{ik})\theta(b-r_{jk})\theta(r_{ij}-b) \\
&+\theta(b-r_{jk})\theta(b-r_{ij})\theta(r_{ik}-b) \\
&+\theta(b-r_{ij})\theta(b-r_{ik})\theta(r_{jk}-b) \\
&+\theta(b-r_{ij})\theta(b-r_{ik})\theta(b-r_{jk})] \\
&\times[A_{ijk\alpha}(1-h_{ik}h_{jk}(1-h_{ij}) \\
&-h_{ij}h_{kj}(1-h_{ik})-h_{ik}h_{ij})+A_{ij\alpha}h_{ik}h_{jk}(1-h_{ij}) \\
&+A_{ik\alpha}h_{ij}h_{kj}(1-h_{ik})+A_{i\alpha}h_{ik}h_{ij}]+\cdots, \qquad (2.2)
\end{aligned}$$

where

$$\begin{aligned}
\theta(r_{ia}-b)&=0, \quad r_{ia}<b, \\
&=1, \quad r_{ia}>b. \qquad (2.3)
\end{aligned}$$

The h_{ij} are the functions required for the continuity of the transformation. They are arbitrary functions subject to the following conditions:

$$\begin{aligned}
h_{ij}=h(r_{ij})&=0, \quad r_{ij}=0, \\
&=1, \quad r_{ij}=b. \qquad (2.4)
\end{aligned}$$

By substituting Eqs. (2.4) into Eq. (2.2) the boundary conditions are satisfied. To illustrate this point, consider the example described previously where the three-body cluster $(ijk)\rightarrow$ [two-body cluster (ij)] $\times$ [one-body cluster (k)] as $(r_{ik}, r_{jk})\rightarrow b$. Then, from Eq. (2.4), $(h_{ik}, h_{jk})\rightarrow 1$ and the three-body (ijk) cluster transformation,

$$\begin{aligned}
X_{i\alpha} = &A_{ijk\alpha}[1-h_{ik}h_{jk}(1-h_{ij}) \\
&-h_{ij}h_{kj}(1-h_{ik})-h_{ik}h_{ij}] \\
&+A_{ij\alpha}h_{ik}h_{jk}(1-h_{ij})+A_{ik\alpha}h_{ij}h_{kj}(1-h_{ik}) \\
&+A_{i\alpha}h_{ik}h_{ij},
\end{aligned}$$

becomes

$$X_{i\alpha}\xrightarrow[(r_{ik},r_{jk})\rightarrow b]{} A_{ij\alpha}(1-h_{ij})+A_{i\alpha}h_{ij},$$

the two-body (ij) cluster transformation for $X_{i\alpha}$. Similarly, the three-body cluster transformations for $X_{j\alpha}$ and $X_{k\alpha}$, in the limit of $(r_{ik}, r_{jk})\rightarrow b$, become

$$X_{j\alpha}\xrightarrow[(r_{ik},r_{jk})\rightarrow b]{} A_{ji\alpha}(1-h_{ij})+A_{j\alpha}h_{ij},$$

$$X_{k\alpha}\xrightarrow[(r_{ik},r_{jk})\rightarrow b]{} A_{k\alpha},$$

the two-body cluster (ij) transformation for $X_{j\alpha}$ and the one-body cluster (k) transformation for $X_{k\alpha}$.

We shall also find that we must require both dh_{ij}/dr_{ij} and d^2h_{ij}/dr_{ij}^2 be zero at $r_{ij}=b$. In addition, the h_{ij} must increase slowly enough with respect to the r_{ij} such that $R_{kl}\geq c$ when $r_{kl}\geq\epsilon'$, for all l, k, $1\leq l<k\leq N$, where ϵ' is the transformed [by Eq. (2.2)] strong-interaction radius, an arbitrarily chosen small distance.[14] The h_{ij} terms provide a continuous transition between the various n-body cluster regions and are thereby used as a cutoff for the transformation $A_{i_1\cdots i_n\alpha}$. Therefore, the cutoff factor $\phi(r)$, used in the pairwise transformation, does not need to be used in the $A_{i_1\cdots i_n\alpha}$.

The total transformation, Eq. (2.2), must be one-to-one. This condition leads to the requirement that $b>c$. For example, in the two-body problem the transformation gives $R_{ij}=b$ when $r_{ij}=b$, but $R_{ij}=c$ when $r_{ij}=\epsilon'$. The one-to-one character of the transformation in the $3N$-dimensional space depends upon the exact specification of the functions h and A. If we take each transformation $A_{i_1\cdots i_n\alpha}$ such that $R_{i_a}-R_{cm(i_1\cdots i_n)}$ is a monotonically increasing function of $r_{i_a}-r_{cm(i_1\cdots i_n)}$ for all $1\leq a\leq n$, then the only cause of $X_{i\alpha}$ being the same for two or more points in the transformed $3N$-dimensional space can be the combinations of the n-body clusters caused by the h's. Because of the form of the transformation Eq. (2.2), we do not expect this phenomenon to occur with the same two or more $3N$-dimensional points in the transformed space for the other $3N-1$ $X_{i\alpha}$ in the original space. Thus, we expect the total transformation to be one-to-one. This requirement should, of course, be checked for each specification of the h's. We note that for the two-body cluster we must require that h be a sufficiently slowly increasing function of r such that $dR/dr>0$. Consider the points $r_{i_ai_b}=b$. If we assume

$$\theta(x-y)=\tfrac{1}{2} \quad \text{for} \quad x=y,$$

then we have contributions from n- and $(n+1)$-body clusters at the boundary between these regions. Since we have taken the transformation to be continuous, the contributions from both cluster terms are the same. Therefore, for convenience we shall include $r_{ia}=b$ in the higher-order cluster, i.e., in the transformation (Eq. 2.2) [or Eq. (1.17)] we set

$$\begin{aligned}
\theta(r_{ia}-b)=0, \quad &\theta(b-r_{ia})=1, \\
&\text{at the point} \quad r_{ia}=b. \qquad (2.5)
\end{aligned}$$

We note that, if we so desired, we could eliminate the θ functions in Eq. (2.2) and keep only the highest-order cluster, i.e., the N cluster for an N-body system,

[14] Because of the h's, ϵ', although related to, is different from ϵ. However, as $\epsilon\rightarrow 0$, $\epsilon'\rightarrow 0$.

by changing the conditions of Eq. (2.3) to

$$h_{ij} = h(r_{ij}) = 0, \quad r_{ij} = 0, \\ = 1, \quad r_{ij} \geq b. \tag{2.6}$$

We now come to the determination of the $A_{i_1 \cdots i_n \alpha}$. In the transformation Eq. (2.2), the θ's determine the regions and the h's smooth the transformation between the various regions. The A's then are, in essence, the transformations which shrink the n-body clusters about their centers of mass. For a one-body cluster there is no interaction and no shrinkage; thus, we want the identity transformation

$$A_{i\alpha} = x_{i\alpha}. \tag{2.7}$$

For particle i in a two-body cluster with particle j, we want

$$X_{i\alpha} - X_{cm(ij)\alpha} = \alpha_{ij}(x_{i\alpha} - x_{cm(ij)\alpha}), \tag{2.8}$$

where α_{ij} is the shrinkage function. Replacing $X_{cm\alpha}$ and $x_{cm\alpha}$ by their values in the two-body (ij) system and requiring them to be equal, we have

$$X_{i\alpha} = \alpha_{ij} x_{i\alpha} + (1 - \alpha_{ij}) \tfrac{1}{2}(x_{i\alpha} + x_{j\alpha}) \\ = A_{ij\alpha}, \tag{2.9}$$

where the second line of Eq. (2.9) follows from the definition of $A_{ij\alpha}$ as the transformation of $X_{i\alpha}$ for the two-body (ij) cluster. In general,

$$A_{i_1 \cdots i_n \alpha} = \alpha_{i_1 \cdots i_n} x_{i_1 \alpha} \\ + (1 - \alpha_{i_1 \cdots i_n}) n^{-1} (x_{i_1 \alpha} + \cdots + x_{i_n \alpha}). \tag{2.10}$$

We note that the shrinkage functions α are dependent only on the interparticle distances $r_{i_a i_b}$, and not on the Cartesian variables α. This lack of dependence is due to the fact that only central potentials are being considered. To complete the specification of the transformation (neglecting the arbitrariness of the h's),we must specify the $\alpha_{i_1 \cdots i_n}$. From our previous discussion, $\alpha_{i_1 \cdots i_n}$ was to depend only on the smallest interparticle separation $r_{i_a i_b}$ (or $R_{i_a i_b}$). We must, therefore, first determine the smallest $r_{i_a i_b}$ (or $R_{i_a i_b}$) and then use a function dependent on $r_{i_a i_b}$ to effect the shrinkage. We are thereby led to the form

$$\alpha_{i_1 \cdots i_n} = (\tau_n)^{-1} \sum_{\substack{c,d=1 \\ c<d \\ i_a i_b \neq i_c i_d}}^{n} \prod_{\substack{a,b=1 \\ a<b}}^{n} \theta(r_{i_a i_b} - r_{i_c i_d}) e(r_{i_c i_d}), \tag{2.11}$$

where τ_n is the normalization function

$$\tau_n = \sum_{\substack{c,d=1 \\ c<d \\ i_a i_b \neq i_c i_d}}^{n} \prod_{\substack{a,b=1 \\ a<b}}^{n} \theta(r_{i_a i_b} - r_{i_c i_d}) \tag{2.12}$$

and $i_a i_b \neq i_c i_d$ means that $r_{i_a i_b}$ and $r_{i_c i_d}$ are not the same interparticle spacing. The product over the θ's is zero unless $r_{i_c i_d}$ is less than or equal to each one of the other interparticle distances $r_{i_a i_b}$ in the n-body cluster $(i_1 \cdots i_n)$. The function $e(r_{i_c i_d})$ is the actual shrinkage factor taken to be the same form for all $i_c i_d$. If three or more interparticle separations are equal, while being less than all other interparticle separations, then α defined with $\tau_n = 1$ would be discontinuous. Inclusion of τ_n defined by Eq. (2.12) corrects this problem. Let us now consider the actual shrinkage factor $e(r_{i_c i_d})$. From our previous discussion, we can take $(1 - h \cdots h)$ as the cutoff function. Thus, $e(r_{i_c i_d})$ can be taken to be just the shrinkage factor for $c \to \epsilon$ in each cluster transformation. One possible expression for e is the pairwise transformation form

$$e(r_{i_c i_d}) = 1 + c/(r_{i_c i_d} + \epsilon''), \tag{2.13}$$

where ϵ'' is a small number related to ϵ.

We have thus derived our set of transformations; each element in the set specified by our choice of the arbitrary functions h_{ij}, the cutoff distance b, and the contracted strong interaction radius ϵ. For hard cores[3] we let ϵ and thus ϵ', $\epsilon'' \to 0$. Let us now take this set of transformations, Eq. (2.2), assume two-body additivity, and compare it to the set of pairwise transformations.

3. PAIRWISE LIMIT

Neglecting three- and more-body clusters, $X_{i\alpha}$ is given by the first two terms of Eq. (2.2). These terms can be written as

$$X_{i\alpha} = \sum_{\substack{j=1 \\ j \neq i}}^{N} \prod_{\substack{a=1 \\ a \neq i,j}}^{N} \theta(r_{i\alpha} - b) \\ \times \{\theta(r_{ij} - b) A_{i\alpha} + \theta(r_{j\alpha} - b)\theta(b - r_{ij}) \\ \times [A_{ij\alpha}(1 - h_{ij}) + A_{i\alpha} h_{ij}]\}. \tag{3.1}$$

Since we are assuming the existence of only two-body clusters, we can set $\theta(r_{j\alpha} - b)$, $\theta(r_{i\alpha} - b) = 1$. For a particular (i,j) cluster, Eqs. (2.7) and (2.9) give

$$X_{i\alpha} = \theta(r_{ij} - b) x_{i\alpha} + \theta(b - r_{ij}) \\ \times \{[\alpha_{ij} x_{i\alpha} + \tfrac{1}{2}(1 - \alpha_{ij})(x_{i\alpha} + x_{j\alpha})] \\ \times (1 - h_{ij}) + x_{i\alpha} h_{ij}\}. \tag{3.2}$$

To put Eq. (3.2) into the form of Eq. (1.5), we change to the relative coordinates $X_{ij\alpha}$ and $x_{ij\alpha}$:

$$X_{ij\alpha} = \theta(r_{ij} - b) x_{ij\alpha} \\ + \theta(b - r_{ij})[\alpha_{ij} x_{ij\alpha}(1 - h_{ij}) + x_{ij\alpha} h_{ij}]. \tag{3.3}$$

Using Eq. (2.6) for the boundary conditions on h_{ij}, we rewrite Eq. (3.3) as

$$X_{ij\alpha} = x_{ij\alpha}[h_{ij} + \alpha_{ij}(1 - h_{ij})]. \tag{3.4}$$

For $b < r_0$, Eq. (3.4) is of the same form as Eq. (1.5). Comparing Eqs. (1.16) and (2.6), we can set

$$h_{ij} = 1 - \phi(r_{ij}). \tag{3.5}$$

From Eqs. (2.11) and (2.13), for two bodies,

$$\alpha_{ij} = e(r_{ij}) = 1 + c/(r_{ij} + \epsilon''). \tag{3.6}$$

Using Eqs. (3.5) and (3.6), we obtain

$$X_{ij\alpha} = x_{ij\alpha}\{1 + [c/(r_{ij} + \epsilon'')]\phi(r_{ij})\}, \tag{3.7}$$

which, in the limit of $\epsilon'' \to 0$, becomes the hard-core pairwise transformation given by Eqs. (1.5) and (1.15).

We have, therefore, developed a set of transformations which can be used for n-body clusters, n being arbitrary, $1 \leq n \leq N$, in considering an N-body system and which, in the limit of two-body clusters, reduces to the set of pairwise transformations. In the next section we show that, given a transformation of the general form of Eq. (2.2), the transformed Hamiltonian, for short range potentials, can be written as the sum of two-body H'_{ij}, three-body $H'_{ijk}, \cdots$, and N-body $H'_{1\cdots N}$ Hamiltonians as indicated in Eq. (1.18).

4. THE TRANSFORMED HAMILTONIAN

We start by considering the transformed Hamiltonian resulting from a transformation of the form of Eq. (2.2). Generalizing Eq. (1.16), we take the relationship between the transformed momenta $p_{j\beta}$ and the original momenta $P_{j\alpha}$ to be

$$P_{i\alpha} = \frac{1}{2}\sum_{j=1}^{N}\sum_{\beta=1}^{3}\left(\frac{\partial x_{j\beta}}{\partial X_{i\alpha}}p_{j\beta} + p_{j\beta}\frac{\partial x_{j\beta}}{\partial X_{i\alpha}}\right), \tag{4.1}$$

which, quantizing the momenta, can also be expressed as

$$\begin{aligned} P_{i\alpha} &= \sum_{j=1}^{N}\sum_{\beta=1}^{3}\frac{\partial x_{j\beta}}{\partial X_{i\alpha}}\left(p_{j\beta} - \tfrac{1}{2}i\hbar\frac{\partial \ln B}{\partial x_{j\beta}}\right) \\ &= \sum_{j=1}^{N}\sum_{\beta=1}^{3}\left(p_{j\beta} + \tfrac{1}{2}i\hbar\frac{\partial \ln B}{\partial x_{j\beta}}\right)\frac{\partial x_{j\beta}}{\partial X_{i\alpha}}, \end{aligned} \tag{4.2}$$

where B is the $3N$th-order Jacobian $|\partial x_{i\alpha}/\partial X_{l\tau}|$. By means of the point-transformation method (the generator of the point transforms being independent of time), the original Hamiltonian

$$H(\mathbf{R}, \mathbf{P}) = \sum_{i=1}^{N}\sum_{\alpha=1}^{3}P_{i\alpha}^2 + V(\mathbf{R}, \mathbf{P}) \tag{4.3}$$

is transformed into itself, expressed in terms of the transformed coordinates

$$H'(\mathbf{r}, \mathbf{p}) = H(\mathbf{R}(\mathbf{r})\mathbf{P}(\mathbf{r}, \mathbf{p})) \tag{4.4}$$

$$= \sum_{i,j=1}^{N}\sum_{\alpha,\beta=1}^{3}(p_{i\alpha}g_{i\alpha j\beta}p_{j\beta}) + V(\mathbf{r}, \mathbf{p}) + W(\mathbf{r}), \tag{4.5}$$

where

$$g_{i\alpha j\beta} = \sum_{l=1}^{N}\sum_{\tau=1}^{3}\frac{\partial x_{i\alpha}}{\partial X_{l\tau}}\frac{\partial x_{j\beta}}{\partial X_{l\tau}}, \tag{4.6}$$

$$W(\mathbf{r}) = \sum_{i,j=1}^{N}\sum_{\alpha,\beta=1}^{3}\left[\tfrac{1}{4}g_{i\alpha j\beta}\frac{\partial \ln B}{\partial x_{i\alpha}}\frac{\partial \ln B}{\partial x_{j\beta}} - \frac{1}{2}\frac{\partial}{\partial x_{j\alpha}}\left(g_{i\alpha j\beta}\frac{\partial \ln B}{\partial x_{j\beta}}\right)\right] \tag{4.7}$$

and $V(\mathbf{r}, \mathbf{p})$ is $V(\mathbf{R}, \mathbf{P})$ written in terms of $\mathbf{r}$, $\mathbf{p}$. In addition, the original normalized wavefunction $\psi(\mathbf{R})$ is related to the transformed normalized wavefunction $\psi'(\mathbf{r})$ by

$$\psi(\mathbf{R}(\mathbf{r})) = [B(\mathbf{r})]^{\frac{1}{2}}\psi'(\mathbf{r}). \tag{4.8}$$

To find the Hamiltonian transformed by Eq. (2.2), we must find the derivatives $\partial x_{k\lambda}/\partial X_{l\tau}$, the effective metrics $g_{i\alpha j\beta}$, the Jacobian B, and the new coordinate dependent potential $W(\mathbf{r})$. We shall now derive the regional form for these functions.[15]

Consider the $\partial x_{k\lambda}/\partial X_{l\tau}$. These terms are found by taking the derivative with respect to $X_{l\tau}$ of Eq. (2.2) for all $i = 1, \cdots, N$, $\alpha = 1, 2, 3$. Let us first discuss the derivatives of the θ functions explicitly displayed in Eq. (2.2). The θ functions are used to specify regions of space. The derivatives of the θ functions are nonzero only at the boundary between these regions, i.e., at some $r_{ij} = b$. In one region we have $\theta(r_{ij} - b)$. In the adjoining region we have $\theta(b - r_{ij})$. Now,

$$\frac{\partial\theta(r_{ij} - b)}{\partial X_{k\lambda}} = -\frac{\partial\theta(b - r_{ij})}{\partial X_{k\lambda}}.$$

Therefore, since the transformation is continuous at the boundary between these two regions, the contribution from the derivatives of the θ functions cancel each other. By considering the transformation Eq. (2.2), we see that this cancellation holds for all the explicitly displayed θ functions. We note that, by using the same argument, we also obtain cancellation of the derivatives of the θ functions occurring in the derivatives of the shrinkage functions $\alpha(r)$ and the normalization functions τ_n. The identical form of the $e(r_{ij})$ for all ij was required for the continuity of the transformation and, therefore, to effect this cancellation. Thus, for any of the terms in the derivatives of the transformation, as in the transformation itself, the θ functions are only used to define regions. Therefore, in our calculation of the derivatives $\partial x_{i\alpha}/\partial X_{l\tau}$, we are just left with the derivatives of the transformation functions themselves, continuously connected between the various regions. Applying the chain rule,

[15] A more thorough mathematical derivation of the form of these functions is given in Ref. 6.

the resulting set of simultaneous linear equations in the $3N$ variables $\partial x_{k\lambda}/\partial X_{l\tau}$ ($k = 1, \cdots, N, \lambda = 1, 2, 3$, $l\tau$ fixed) is then

$$\delta_{il}\delta_{\alpha\tau} = \prod_{\substack{a=1 \\ a\neq i}}^{N} \theta(r_{ia} - b)\left(\sum_{k=1}^{N}\sum_{\lambda=1}^{3}\frac{\partial \bar{A}_{i\alpha}}{\partial x_{k\lambda}}\frac{\partial x_{k\lambda}}{\partial X_{l\tau}}\right) + \sum_{\substack{i,j=1 \\ j<i}}^{N}\left(\prod_{\substack{a=1 \\ a\neq i,j}}^{N}\theta(r_{ia} - b)\theta(r_{ja} - b)\right) \times [\theta(b - r_{ij})]\left(\sum_{k=1}^{N}\sum_{\lambda=1}^{3}\frac{\partial \bar{A}_{ij\alpha}}{\partial x_{k\lambda}}\frac{\partial x_{k\lambda}}{\partial X_{l\tau}}\right) + \cdots + [\theta \text{ functions for the } N\text{-body } (N, N-1, \cdots, 3, 2, 1) \text{ cluster combinations}] \times \left(\sum_{k=1}^{N}\sum_{\lambda=1}^{3}\frac{\partial \bar{A}_{iN\cdots 321\alpha}}{\partial x_{k\lambda}}\frac{\partial x_{k\lambda}}{\partial X_{l\tau}}\right), \quad (4.9)$$

where for simplicity in notation we have used the symbolism of Eq. (1.17) instead of Eq. (2.2). Using Cramer's rule,[16] the set of equations for the derivatives $\partial x_{k\lambda}/\partial X_{l\tau}$ is of the form

$$\frac{\partial x_{k\lambda}}{\partial X_{l\tau}} = \frac{E_{k\lambda,l\tau}}{D}, \quad (4.10)$$

where D is the $3N$th-order determinant of the coefficients of the $\partial x_{k\lambda}/\partial X_{l\tau}$ in Eq. (4.9) and $E_{k\lambda,l\tau}$ is D with the $k\lambda$th column replaced by the coefficients on the left-hand side of Eq. (4.9), the $\delta_{il}\delta_{\alpha\tau}$. We write D in the form

$$D = |D_{i\alpha,k\lambda}| = \left|C_{i\alpha}\left(\frac{\partial x_{k\lambda}}{\partial X_{l\tau}}\right)\right|, \quad (4.11)$$

where dictionary ordering is used (for example, the ordering of the rows $i\alpha$ is 11, 12, 13, 21, $\cdots$, $N3$) and $C_{i\alpha}(\partial x_{k\lambda}/\partial X_{l\tau})$ is the coefficient of $\partial x_{k\lambda}/\partial X_{l\tau}$ in the $i\alpha$ equation of the set of equations given by Eq. (4.9). Now each term $C_{i\alpha}(\partial x_{k\lambda}/\partial X_{l\tau})$ in this determinant will include the sum over the contributions from all the n-body clusters containing the particle i. For example, in the three-body system 1, 2, 3, $C_{1\alpha}(\partial x_{k\lambda}/\partial X_{l\tau})$ will include contributions from the clusters (1), (12), (13), and (123). Thus, in the expansion of the determinant D, there will be products of the $C_{i\alpha}$ terms in which one term is from one cluster of particles and another term is from a different cluster of particles, where one or more of the particles in the first cluster is also included in the second cluster. For example, in the three-body system 1, 2, 3, there will be products of terms from the clusters (123) and (1), (123) and (12), (12) and (23), etc. Products containing two or more such combinations shall be called cross-term products. By their definition cross-term products will be proportional to terms of the form $\theta(r_{ij} - b)\theta(b - r_{ij})$. Thus, from Eqs. (2.3) and (2.5) the cross term products in D will be zero. Therefore, the only nonzero terms resulting from this determinant are the product of cluster terms where no particle is in two different clusters. From the form of Eq. (2.2) the determinant D can thus be put [by rearranging Eq. (4.11) where necessary] into the form

$$D = \sum_{\{n(N)\}}\theta(\{n(N)\})\prod_{i} D_{(n)_i} \quad (4.12)$$

where $\{n(N)\}$ represents a specific combination of clusters of the N particles, $\sum_{\{n(N)\}}$ is the sum over all possible combinations of clusters, $\theta(\{n(N)\})$ represents the product of the θ functions defining $\{n(N)\}$, $\prod_i$ is the product over the various i clusters $(n)_i$ which make up each cluster combination $\{n(N)\}$, and $D_{(n)_i}$ is the denominator we would obtain if we were only considering the isolated $(n)_i$ cluster. For example, in the three-body system, we have

$$D = \theta((1)(2)(3))D_{(1)}D_{(2)}D_{(3)} + \theta((12)(3))D_{(12)}D_{(3)} + \theta((13)(2))D_{(13)}D_{(2)} + \theta((23)(1))D_{(23)}D_{(1)} + \theta((123))D_{(123)}.$$

We note that, by applying the same techniques used in deriving Eq. (4.12), we can show that in the determinant D the cross terms from two different explicit θ-function combinations in the same n-body cluster [see Eq. (2.2) and diagrams following Eq. (2.1)] cancel. This type of cancellation will also occur in the remainder of our calculations. Although this elimination makes no difference in transformations of the form Eq. (1.17), it can be useful in their generalizations. Let us now calculate the numerators $E_{k\lambda,l\tau}$. From their definition, the $E_{k\lambda,l\tau}$, like the denominator D, can be written in the form

$$E_{k\lambda,l\tau} = \sum_{\{n(N)\}}\theta(\{n(N)\})\prod_{i} E_{k\lambda,l\tau(n)_i}. \quad (4.13)$$

In addition, $E_{k\lambda,l\tau(n)_i}$ is $D_{(n)_i}$ if the cluster $(n)_i$ does not contain either the particles k or l. Also, since each cluster has its own transformation dependent only on the coordinates of the particles contained in that cluster, $E_{k\lambda,l\tau(n)_i}$ equals zero if the cluster $(n)_i$ contains only one of the particles k or l. Therefore, substituting Eqs. (4.12) and (4.13) into Eq. (4.10), multiplying both sides of the equation by the denominator, and equating the various regions defined by the $\theta(\{n(N)\}$'s, we obtain

$$\frac{\partial x_{k\lambda}}{\partial X_{l\tau}} = \sum_{(.k.l.)}\theta((.k.l.))\frac{E_{k\lambda,l\tau(.k.l.)}}{D_{(.k.l.)}}, \quad (4.14)$$

[16] See, for example, G. B. Thomas, *Calculus and Analytic Geometry* (Addison-Wesley Publishing Co., Reading, Mass., 1953), Chap. 12.

where we have summed over the θ functions defining clusters not containing the particles k and l and notationally $\sum_{(.k.l.)}$ is the sum over all clusters including the particles k and l, $\theta((.k.l.))$ represents the product (and sums) of the θ functions defining the $(.k.l.)$ cluster, and $E_{k\lambda,lr(.k.l.)}$ is the numerator we would obtain if we were considering only the isolated $(.k.l.)$ cluster. As can be expected from the definition of $A_{i\alpha}$, Eq. (2.7), $E_{k\lambda,kr(k)}/D_{(k)} = \delta_{\lambda r}$, and the general form of $\partial x_{k\lambda}/\partial X_{lr}$ is

$$\frac{\partial x_{k\lambda}}{\partial X_{lr}} = \theta((k))\delta_{kl}\delta_{\lambda r} + \sum_{\substack{(.k.l.)\\(.k.l.)\neq(k)}} \theta((.k.l.))\left(\frac{\partial x_{k\lambda}}{\partial X_{lr}}\right)_{(.k.l.)}, \tag{4.15}$$

where $(\partial x_{k\lambda}/\partial X_{lr})_{(.k.l.)}$ is just $\partial x_{k\lambda}/\partial X_{lr}$ for the isolated $(.k.l.)$ cluster. From their definitions, both $E_{k\lambda,lr(.k.l.)}$ and $D_{(.k.l.)}$ are $3r$th-order determinants, where r is the number of particles in the $(.k.l.)$ cluster. In the Appendix we show that $D_{(.k.l.)}$ can be written as the sum of products of three rth-order determinants, where the sum omits many of the terms in the minor expansion of the original $3r$th-order determinant. A similar result can be obtained for $E_{k\lambda,lr(.k.l.)}$. We mention this point to indicate that this method is practicable for calculations assuming the existence of only few-body clusters. For example, in the N-body problem, assuming three-body additivity, we need only to evaluate third (not ninth)-order determinants.

Following our procedure for finding the transformed Hamiltonian H', we now consider the effective metrics $g_{i\alpha j\beta}$. Substituting Eq. (4.15) into Eq. (4.6) and eliminating cross-term products, we obtain

$$g_{i\alpha j\beta} = \theta((i))\delta_{ij}\delta_{\alpha\beta} + \sum_{\substack{(.i.j.)\\(.i.j.)\neq(i)}} \theta((.i.j.))g_{i\alpha j\beta(.i.j.)}, \tag{4.16}$$

where $g_{i\alpha j\beta(.i.j.)}$ is just $g_{i\alpha j\beta}$ for the isolated $(.i.j.)$ cluster.

We next consider the Jacobian $B = |\partial x_{k\lambda}/\partial X_{lr}|$. As in the preceding calculations, we obtain

$$B = \sum_{\{n(N)\}} \theta(\{n(N)\}) \prod_i B_{(n)_i}, \tag{4.17}$$

where $B_{(n)_i}$ is the Jacobian of the transformation for the isolated $(n)_i$ cluster. The determination of $B_{(n)_i}$ thus involves the calculation of a $3r$th-order determinant, where r is the number of particles in the $(n)_i$ cluster. We note that, by considering the form of the differences between $E_{k1,lr(.k.l.)}$, $E_{k2,lr(.k.l.)}$, and $E_{k3,lr(.k.l.)}$ and using the method illustrated in Appendix A, we can show that $B_{(n)_i}$, like the D's and E's, can be written as the sum of products of three rth-order determinants, where the sum omits many of the terms in the minor expansion of the original $3r$th-order determinant. We observe that, since the wavefunctions $\psi(R)$ and $\psi'(r)$ and their first derivatives must be continuous, the Jacobian B and its first derivatives must be continuous throughout the entire $3N$-dimensional space (including the region boundaries). Thus, we must require the first and second derivatives of the continuity functions h to be zero at the region boundaries and the cluster transformations A, their first and second derivatives to be continuous throughout the entire $3N$-dimensional space.

Continuing our procedure, we now consider the new coordinate-dependent potential $W(\mathbf{r})$. Equation (4.7) can be rewritten as

$$W(r) = \sum_{i,j=1}^{N} \sum_{\alpha,\beta=1}^{3} B^{\frac{1}{2}} \frac{\partial}{\partial x_{i\alpha}}\left[g_{i\alpha j\beta} \frac{\partial}{\partial x_{j\beta}}\left(\frac{1}{B^{\frac{1}{2}}}\right)\right]. \tag{4.18}$$

From the definition of the θ functions, Eqs. (2.3) and (2.5), $\theta(\{n(N)\})^{\frac{1}{2}}$ is just $\theta(\{n(N)\})$. From the requirements on the continuity functions h and the cluster transformations A, the effective cluster metrics $g_{i\alpha j\beta(.i.j.)}$ and the first derivatives of the cluster Jacobians $B_{(n)_i}$ are continuous at the region boundaries. Therefore, the derivatives of the θ functions occurring in Eq. (4.18) cancel each other. Applying the methods used in our previous derivations, we then obtain

$$W(\mathbf{r}) = \sum_{n=1}^{N} \sum_{(i_1\cdots i_n)} \left\{ \theta((i_1\cdots i_n)) \sum_{a,b=1}^{n} \sum_{\alpha,\beta=1}^{3} \left[B^{\frac{1}{2}}_{(i_1\cdots i_n)} \times \frac{\partial}{\partial x_{i_a\alpha}}\left(g_{i_a\alpha i_b\beta(i_1\cdots i_n)} \frac{\partial}{\partial x_{i_b\beta}} \frac{1}{B^{\frac{1}{2}}_{(i_1\cdots i_n)}}\right)\right]\right\} \tag{4.19}$$

$$\equiv \sum_{n=1}^{N} \sum_{(i_1\cdots i_n)} \theta((i_1\cdots i_n)) W_{(i_1\cdots i_n)}, \tag{4.20}$$

where $\sum_{(i_1\cdots i_n)}$ is the sum over all possible n-body clusters. For a one-body cluster, from Eqs. (4.15) and (4.16),

$$W_{(i)} = 0. \tag{4.21}$$

Thus, $W(\mathbf{r})$ is the sum of contributions from each region in which the transformation is not the identity transformation. We now combine these results to obtain the transformed Hamiltonian H'. Substituting Eqs. (4.16) and (4.20) into Eq. (4.5) and using our continuity requirements on the continuity functions h and the cluster transformations A, we obtain

$$H' = \left\{ \sum_{n=1}^{N} \sum_{(i_1\cdots i_n)} \theta((i_1\cdots i_n)) \times \left[\left(\sum_{a,b=1}^{n} \sum_{\alpha,\beta=1}^{3} p_{i_a\alpha} g_{i_a\alpha i_b\beta(i_1\cdots i_n)} p_{i_b\beta} \right) + W_{(i_1\cdots i_n)} \right] \right\} + V(\mathbf{r}, \mathbf{p}). \tag{4.22}$$

If the transformed potential has a range less than b, then $V(\mathbf{r}, \mathbf{p})$ can be written in the same regional form as Eq. (4.20), and Eq. (4.22) becomes

$$H' = \sum_{n=1}^{N} \sum_{(i_1 \cdots i_n)} \theta((i_1 \cdots i_n)) \times \left[\left(\sum_{a,b=1}^{n} \sum_{\alpha,\beta=1}^{3} p_{i_a\alpha} g_{i_a\alpha i_b\beta}(i_1 \cdots i_n) p_{i_b\beta} \right) + W_{(i_1 \cdots i_n)} + V_{(i_1 \cdots i_n)} \right], \quad (4.23)$$

$$= \sum_{n=1}^{N} \sum_{(i_1 \cdots i_n)} \theta((i_1 \cdots i_n)) H'_{(i_1 \cdots i_n)}, \quad (4.24)$$

where $V_{(i_1 \cdots i_n)}$ represents the transformed potentials acting between the particles $i_1 \cdots i_n$ in the isolated $(i_1 \cdots i_n)$ cluster and $H'_{(i_1 \cdots i_n)}$ is the transformed n-body Hamiltonian of the isolated $(i_1 \cdots i_n)$ cluster. We note that, if the original potential $V(\mathbf{R}, \mathbf{P})$ was only the strong short-range potential whose effects were weakened by the transformation (2.2), then $V(\mathbf{r}, \mathbf{p})$ would have a range $\epsilon' \ll c < b$ and the above condition would easily be satisfied. For a one-body cluster, from Eqs. (4.16) and (4.21),

$$H_{(i)} = \sum_{\alpha=1}^{3} p_{i\alpha}^2, \quad (4.25)$$

the unperturbed form. Equation (4.24) with Eq. (4.25) is just the condensed notation for the equation we sought to derive, Eq. (1.18). The sum over all clusters is expected from the regional dependence of the transformation. We note that, in deriving Eq. (4.23), we have not specified the exact form of the cluster transformations A other than by specifying the requirements that they, their first, and their second derivatives be continuous throughout the entire $3N$-dimensional space. In addition, we have not specified the exact form of the continuity functions h other than by specifying the requirements that

$$h_{ij} = 0, \quad r_{ij} = 0,$$
$$= 1, \quad r_{ij} = 1,$$

and

$$\frac{dh_{ij}}{dr_{ij}} = \frac{d^2h_{ij}}{dr_{ij}^2} = 0, \quad r_{ij} = b.$$

5. CONCLUSIONS

We have derived a set of regional transformations, Eq. (2.2), which, by properly choosing the cluster transformation A and the continuity functions h, includes part of the interparticle interaction in the transformation, thereby weakening the interaction terms dealt with by means of perturbation theory. Using Eqs. (2.10)–(2.13) with a suitable choice of ϵ, we can make the effects of a transformed strong short-range potential small. For example, in the case of hard-core potentials, the transformation in Eqs. (2.10)–(2.13) with ϵ (and thus ϵ', ϵ'') $\to 0$ removes the singular components. Each element in the set of transformed N-body Hamiltonians is equivalent to the original N-body Hamiltonian, is Fourier analyzable, and is amenable to ordinary perturbation and variational techniques. Both the transformed Hamiltonian H' and the Jacobian of the transformation B have been shown to be of the same regional character as that of the original regional transformation. Thus, they can be easily approximated by assuming n-body additivity ($1 \leq n \leq N$), i.e., by assuming the probability for $(n+1)$- and more-body clusters to be negligible where n is arbitrarily chosen as a compromise between accuracy and laborious mathematical computation. Assuming two-body additivity, the set of transformations developed for strong short range potentials and, thus, the set of transformed Hamiltonians associated with it, has been shown to be that derived by Eger and Gross.[3,5] We note that the choice of h (and thus b) is relatively arbitrary. If we solved the problem exactly (i.e., kept all the terms), the ground-state energy and the original wavefunction should be independent of this choice. However, when making approximations by specifying the maximum size of the clusters considered, we expect our results to be dependent upon the h's. Our results will, therefore, be accurate to the extent that they do not involve these parameters. As the size of the clusters considered increases, we expect the dependence of our results upon the choice of h to decrease.

Let us review the application of this method to a specific problem including strong short-range repulsive potentials. We first determine the shrinkage $c \to \epsilon$ required to make the effects of the transformed, strong, repulsive potential sufficiently small. The size of the clusters kept is determined by the amount of calculation we are willing to perform and by the properties of the physical system we wish to consider. A specific form of h satisfying the required conditions is chosen by trial and error. The range b and any other parameters in h are then determined by requiring the energy to be a minimum and/or by requiring the energy spectrum to have a specific form. The value of this method is that a perturbation expansion can be considered to any order by keeping large enough clusters.

ACKNOWLEDGMENTS

The author wishes to express his deep appreciation to Professor E. P. Gross for his suggestion of this problem and for his guidance, criticism, help, and

encouragement over the period in which this work was performed. In addition, the author would like to thank Dr. B. Guenther, Dr. D. W. Howgate, and Dr. J. D. Stettler for their suggestions in the preparation of this manuscript.

APPENDIX

In this Appendix we show that the calculation of the $3r$th-order determinant $D_{(n)_i}$ [where r is the number of particles in the $(n)_i$ cluster] can be reduced to the calculation of the sum of products of three rth-order determinants where the sum omits many of the terms in the minor expansion of the original $3r$th-order determinant. Consider $D_{(1\cdots r)}$. Rearranging the determinant in Eq. (4.11), we can write

$$D_{(1\cdots r)} = |D_{\alpha i,k\lambda(1\cdots r)}|, \tag{A1}$$

where dictionary ordering is used and the elements $D_{\alpha i,k\lambda(1\cdots r)}$ are the terms

$$C_{i\alpha(1\cdots r)}\left(\frac{\partial x_{k\lambda}}{\partial X_{l\tau}}\right) = \frac{\partial \bar{A}_{i1\cdots(i-1)(i+1)\cdots r\alpha}}{\partial x_{k\lambda}}. \tag{A2}$$

Comparing Eqs. (1.17) and (2.2), we observe that $\bar{A}$ is the sum of cluster transformations A multiplied by combinations of the continuity functions h. Taking the h's to be functions only of the radical components r and using Eq. (2.10) for $A_{i\cdots r\alpha}$, $C_{i\alpha(1\cdots r)}$ can be written in the form

$$C_{i\alpha(1\cdots r)}\left(\frac{\partial x_{k\lambda}}{\partial X_{l\tau}}\right) = \frac{x_{k\lambda}}{r_k}\frac{\partial F_{i\alpha k}}{\partial r_k} + G_{ik}\delta_{\alpha\lambda}, \tag{A3}$$

where $F_{i\alpha k}$ and G_{ik} are defined by comparing Eq. (A2) with Eq. (A3) and their dependence on $(1\cdots r)$ is implied. As an example, for the two-body (ij) cluster,

$$F_{i\alpha k} = \left(\frac{\partial \alpha_{ij}(1-h_{ij})}{\partial r_k}\right)x_{i\alpha} + \left(\frac{\partial(1-\alpha_{ij})(1-h_{ij})}{\partial r_k}\right) \times \tfrac{1}{2}(x_{i\alpha}+x_{j\alpha}) + \left(\frac{\partial h_{ij}}{\partial r_k}x_{i\alpha}\right),$$

$$G_{ik} = [\delta_{ik}\alpha_{ij}(1-h_{ij}) + (1-\alpha_{ij})(1-h_{ij}) \times \tfrac{1}{2}(\delta_{ik}+\delta_{jk})] + \delta_{ik}h_{ij}.$$

We now apply the theorem[17]

$$|a_{ij}+b_{ij}| = |a_{ij}| + |b_{ij}| + \sum \Delta a_{3r-1-\alpha}b_{\alpha+1}, \tag{A4}$$

where $3r-2 \geq \alpha \geq 0$; $\Delta a_{3r-1-\alpha}b_{\alpha+1}$ denotes the determinant formed as follows: The first $3r-1-\alpha$ columns are taken from $|a_{ij}|$ and the remaining $\alpha+1$ columns from $|b_{ij}|$, with the proviso that no two columns so taken have the same column number. $\sum$ denotes the sum over all the possible combinations. Identify the two terms on the right-hand side of Eq. (A3) with a_{ij} and b_{ij}, respectively. The first $3r$th-order determinant we obtain, $|(x_{k\lambda}/r_k)(\partial F_{i\alpha k}/\partial r_k)|$, can be written as

$$\left|\frac{x_{k\lambda}}{r_k}\frac{\partial F_{i\alpha k}}{\partial r_k}\right| = \left|\left(\frac{\partial F_{i\alpha k}}{\partial r_k}\right)\left(\frac{\partial F_{i\alpha k}}{\partial r_k}\right)\left(\frac{\partial F_{i\alpha k}}{\partial r_k}\right)\right|\left(\prod_{p=1}^{r}\prod_{\tau=1}^{3}\frac{x_{p\tau}}{r_p}\right), \tag{A5}$$

where $(\partial F_{i\alpha k}/\partial r_k)$ is a $3r \times r$ array. Since the determinant has identical columns,

$$\left|\frac{x_{k\lambda}}{r_k}\frac{\partial F_{i\alpha k}}{\partial r_k}\right| = 0. \tag{A6}$$

The second $3r$th-order determinant we obtain, $|G_{ik}\delta_{\alpha\lambda}|$, is diagonal in the rth-order determinant $|G_{ik}|$. Therefore, it reduces to

$$|G_{ik}\delta_{\alpha\lambda}| = [|G_{ik}|]^3. \tag{A7}$$

Consider the remaining terms K. To avoid repetition of the identical columns in $|(x_{k\lambda}/r_k)(\partial F_{i\alpha k}/\partial r_k)|$ shown in Eq. (A5), each nontrivially zero, $3r$th-order determinant K_m in K must be reducible to the form

$$K_m = \left|\begin{matrix}(G_{ik}) & 0 \\ 0 & (G_{ik}) \\ 0 & 0\end{matrix}\quad\begin{pmatrix}\\ J_{i\alpha k\lambda,m}\\ \\ \end{pmatrix}\right| \tag{A8}$$

where $(J_{i\alpha k\lambda,m})$ a $3r \times r$ array composed of r columns is chosen from both $|(x_{k\lambda}/r_k)(\partial F_{i\alpha k}/\partial r_k)|$ and $|G_{ik}|$. Because of the diagonal form of K_m,

$$K_m = [|G_{ik}|]^2\,|J'_{i\alpha k\lambda,m}|, \tag{A9}$$

where $|J'_{i\alpha k\lambda,m}|$ is the rth-order determinant of the lower $r \times r$ array of $(J_{i\alpha k\lambda,m})$ in Eq. (A8). Combining these equations, we obtain

$$K_m = (|G_{ik}|)^2\left(|G_{ik}| + \sum_m |J'_{i\alpha k\lambda,m}|\right). \tag{A10}$$

Therefore, the $3r$th-order determinant $D_{(1\cdots r)}$ can be reduced to the sum of products of three rth-order determinants, where the sum omits many of the terms in the minor expansion of the original $3r$th-order determinant, namely those not involving $(|G_{ik}|)^2$.

[17] Thomas Muir, *A Treatise on the Theory of Determinants* (Dover Publications, New York, 1960), Theorem 104, p. 89.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Formulation of the Many-Body Problem for Composite Particles. II. Equivalence of Physical and Ideal State Spaces

M. D. GIRARDEAU
Institute of Theoretical Science and Department of Physics, University of Oregon, Eugene, Oregon
(Received 17 March 1969)

It is shown that all inner products are preserved by the correspondence between the physical and ideal state spaces.

1. INTRODUCTION

Consider a nonrelativistic system of ln bosons or fermions, with wavefunctions $\psi(x_1 \cdots x_{ln})$, where each x_j stands for both position and spin variables and $\int dx_j$ denotes integration over position and summation over spin. Given any complete orthonormal set of symmetric (Bose) or antisymmetric (Fermi) l-particle wavefunctions $\varphi_\alpha(x_1 \cdots x_l)$, one can expand

$$\psi(x_1 \cdots x_{ln}) = \sum_{\alpha_1 \cdots \alpha_n} c(\alpha_1 \cdots \alpha_n)\varphi_{\alpha_1}(x_1 \cdots x_l) \cdots \varphi_{\alpha_n}(x_{ln-l+1} \cdots x_{ln}), \tag{1}$$

where

$$c(\alpha_1 \cdots \alpha_n) = \int \varphi^*_{\alpha_1}(x_1 \cdots x_l) \cdots \varphi^*_{\alpha_n}(x_{ln-l+1} \cdots x_{ln}) \times \psi(x_1 \cdots x_{ln})\, dx_1 \cdots dx_{ln}. \tag{2}$$

Such an expansion may be useful in cases where the particles have a tendency to group into composite particles composed of l particles each. This is the case, for example, for l-electron atoms or the electron pairs in a superconductor.

The symmetry or antisymmetry of ψ implies that the coefficients (2) cannot be arbitrary, but satisfy the following conditions:

$$Pc(\alpha_1 \cdots \alpha_n) = (\pm 1)^{lp(P)} c(\alpha_1 \cdots \alpha_n), \tag{3}$$

where P is any permutation of $\alpha_1 \cdots \alpha_n$ and $p(P)$ its parity;

$$I_{pq}c(\alpha_1 \cdots \alpha_n) \equiv \sum_{\alpha\beta} (\alpha_p\alpha_q|\, I\, |\alpha\beta) c(\alpha_1 \cdots \alpha_{p-1}\alpha\alpha_{p+1} \cdots \alpha_{q-1}\beta\alpha_{q+1} \cdots \alpha_n) = \pm c(\alpha_1 \cdots \alpha_n), \tag{4}$$

for $1 \le p < q \le n$, where the exchange matrix is defined as[1]

$$(\alpha\beta|\, I\, |\gamma\delta) \equiv \int \varphi^*_\alpha(x_1 \cdots x_l)\varphi^*_\beta(x'_1 \cdots x'_l)\varphi_\gamma(x'_1 x_2 \cdots x_l) \times \varphi_\delta(x_1 x'_2 \cdots x'_l)\, dx_1 \cdots dx_l\, dx'_1 \cdots dx'_l. \tag{5}$$

The upper or lower signs are to be taken in (3) and (4), depending on whether the particles are bosons or fermions. Because of the symmetry or antisymmetry (3) of c, the $\frac{1}{2}n(n-1)$ subsidiary conditions (4) are not independent; one of them (e.g., with $p = 1, q = 2$) implies all the rest. If c satisfies one (hence all) of these asymmetric subsidiary conditions, it will also satisfy the symmetric one:

$$Ic(\alpha_1 \cdots \alpha_n) \equiv \sum_{p<q}^{n} I_{pq}c(\alpha_1 \cdots \alpha_n) = \pm\tfrac{1}{2}n(n-1)c(\alpha_1 \cdots \alpha_n). \tag{6}$$

Conversely, since[1] ± 1 are the maximal and minimal eigenvalues of I_{pq} and $\pm\frac{1}{2}n(n-1)$ are the maximal and minimal eigenvalues of I, it follows that, if c satisfies the symmetric subsidiary condition, then it will also satisfy all the asymmetric ones. Conditions (3) and (6) are both necessary and sufficient[1] to ensure that the wavefunction (1) is totally symmetric or antisymmetric.

The state (1) has the standard quantized-field representation

$$|\psi\rangle = [(ln)!]^{-\frac{1}{2}} \int dx_1 \cdots dx_{ln}\psi(x_1 \cdots x_{ln}) \times \psi^\dagger(x_1) \cdots \psi^\dagger(x_{ln})\,|0\rangle, \tag{7}$$

where $\psi(x)$ and $\psi^\dagger(x)$ are Bose or Fermi field annihilation and creation operators satisfying the usual commutation or anticommutation relations. Defining composite-particle creation operators

$$A^\dagger_\alpha \equiv (l!)^{-\frac{1}{2}} \int dx_1 \cdots dx_l \varphi_\alpha(x_1 \cdots x_l) \times \psi^\dagger(x_1) \cdots \psi^\dagger(x_l), \tag{8}$$

one can write (7), with (1), in the form

$$|\psi\rangle = \left(\frac{(l!)^n}{(ln)!}\right)^{\frac{1}{2}} \sum_{\alpha_1 \cdots \alpha_n} c(\alpha_1 \cdots \alpha_n) A^\dagger_{\alpha_1} \cdots A^\dagger_{\alpha_n}\,|0\rangle. \tag{9}$$

It follows from (7) that, if ψ and ψ' are any two ln-particle wavefunctions with inner product (ψ, ψ') and corresponding state vectors $|\psi\rangle$ and $|\psi'\rangle$, then

$$(\psi, \psi') = \langle\psi\,|\,\psi'\rangle, \tag{10}$$

where $\langle\psi\,|\,\psi'\rangle$ is computed by the usual rules of Wick's theorem. On the other hand, no such simple

[1] M. Girardeau, J. Math. Phys. 4, 1096 (1963), herein denoted by I. Equation (n) of this paper will be denoted by (I.n). A slightly more general notation, motivated by atoms, for which one wishes to refer explicitly to the nuclei as well as the electrons, was used in I. The reader can easily see how to generalize the present discussion correspondingly; we shall not do so explicitly, in order to avoid irrelevant details.

rules hold in terms of the operators A_α and $A_\alpha^\dagger$, occurring if one uses the representation (9) in $\langle\psi \mid \psi'\rangle$. The trouble is that the A_α and $A_\alpha^\dagger$ are neither Bose nor Fermi operators, but satisfy

$$[A_\alpha, A_\beta]_\pm = [A_\alpha^\dagger, A_\beta^\dagger]_\pm = 0, ,$$

$$[A_\alpha, A_\beta^\dagger]_\pm = \delta_{\alpha\beta} + \int(\quad)\underbrace{\psi^\dagger\cdots\psi^\dagger}_{l-1 \text{ factors}}\underbrace{\psi\cdots\psi}_{l-1 \text{ factors}} + \cdots + \int(\quad)\psi^\dagger\psi, \tag{11}$$

where commutators are to be taken for bosons and commutators or anticommutators for fermions, depending on whether l is even or odd. The extra operator terms on the right-hand side of the expression for $[A_\alpha, A_\beta^\dagger]_\pm$ are the mathematical expression of the fact that a composite particle made of bosons or fermions is not a boson or fermion except in the limit, when it is so tightly bound that it may be regarded as always in its ground state. This is well known,[2] but also frequently ignored.

This difficulty was solved in I by the introduction, following Dyson's treatment of spin waves,[3] of an ideal state space generated from an ideal vacuum state $|0)$ by the application of creation operators $a_\alpha^\dagger$ satisfying normal Bose or Fermi relations

$$[a_\alpha, a_\beta]_\pm = [a_\alpha^\dagger, a_\beta^\dagger]_\pm = 0,$$
$$[a_\alpha, a_\beta^\dagger]_\pm = \delta_{\alpha\beta}. \tag{12}$$

We define

$$|\psi) = (n!)^{-\frac{1}{2}} \sum_{\alpha_1\cdots\alpha_n} c(\alpha_1\cdots\alpha_n) a_{\alpha_1}^\dagger\cdots a_{\alpha_n}^\dagger |0) \tag{13}$$

with the *same* coefficients as occur in (9). It was shown in I that the correspondence $|\psi\rangle \leftrightarrow |\psi)$ is 1-to-1, provided that conditions (3) and (6) are satisfied. Since the same conditions make the correspondence $\psi \leftrightarrow c$ 1-to-1, we see that the correspondence between the physical state space and the ideal state space [space of all states (13) with coefficients satisfying (3) and (6)] is 1-to-1. The method of transforming physical operators, e.g., a many-particle Hamiltonian, into the ideal state space was given in I. This transformation was carried out in such a way as to ensure that, if $O|\lambda\rangle = \lambda|\lambda\rangle$ and O_{ideal} is the corresponding operator in the ideal state space, expressed in terms of the a_α and $a_\alpha^\dagger$, then $O_{\text{ideal}}|\lambda) = \lambda|\lambda)$ with the same λ. However, there are physically important quantities which cannot be expressed in terms of eigenvalues, but require more general matrix elements for their description. In order to ensure the complete physical equivalence of the physical and ideal state spaces, it is necessary to show that all inner products are preserved by the correspondence $|\psi\rangle \leftrightarrow |\psi)$. The purpose of the present paper is to fill that gap in the logic of I. We shall show that, in fact,

$$\langle\psi \mid \psi'\rangle = (\psi \mid \psi') \tag{14}$$

for all $|\psi\rangle$ and $|\psi'\rangle$. When supplemented by the formulae for observables O_{ideal} given in I, this ensures the complete physical equivalence of the physical and ideal state spaces.

2. PROOF

We start by evaluating the inner product $\langle\psi \mid \psi'\rangle$ between two physical states expressed in the form (9). Substituting (8) and making use of Wick's theorem for vacuum expectation values

$$\langle 0|\,\psi\cdots\psi\psi^\dagger\cdots\psi^\dagger\,|0\rangle,$$

one finds

$$\begin{aligned}\langle\psi \mid \psi'\rangle &= [(ln)!]^{-1} \sum_{\alpha_1\cdots\alpha_n}\sum_{\beta_1\cdots\beta_n} c^*(\alpha_1\cdots\alpha_n)c'(\beta_1\cdots\beta_n) \\ &\times \sum_P (\pm 1)^{p(P)} \int dx_1\cdots dx_{ln} \\ &\times \varphi_{\alpha_n}^*(x_{ln-l+1}\cdots x_{ln})\cdots\varphi_{\alpha_1}^*(x_1\cdots x_l) \\ &\times \varphi_{\beta_1}(Px_1\cdots Px_l)\cdots\varphi_{\beta_n}(Px_{ln-l+1}\cdots Px_{ln}),\end{aligned} \tag{15}$$

where $(Px_1\cdots Px_{ln})$ is any one of the $(ln)!$ permutations P of $(x_1\cdots x_{ln})$ and $p(P)$ is its parity. Define

$$\begin{aligned}J_P(\alpha_1\cdots\alpha_n, \beta_1\cdots\beta_n) &\equiv \int \varphi_{\alpha_n}^*(x_{ln-l+1}\cdots x_{ln})\cdots\varphi_{\alpha_1}^*(x_1\cdots x_l) \\ &\times \varphi_{\beta_1}(Px_1\cdots Px_l)\cdots\varphi_{\beta_n}(Px_{ln-l+1}\cdots Px_{ln}) \\ &\times dx_1\cdots dx_{ln}.\end{aligned} \tag{16}$$

Then,

$$\begin{aligned}\langle\psi \mid \psi'\rangle = &[(ln)!]^{-1} \sum_{\alpha_1\cdots\alpha_n}\sum_{\beta_1\cdots\beta_n} c^*(\alpha_1\cdots\alpha_n)c'(\beta_1\cdots\beta_n) \\ &\times \sum_P (\pm 1)^{p(P)} J_P(\alpha_1\cdots\alpha_n, \beta_1\cdots\beta_n).\end{aligned} \tag{17}$$

We next show that

$$\sum_{\beta_1\cdots\beta_n} c'(\beta_1\cdots\beta_n) J_P(\alpha_1\cdots\alpha_n, \beta_1\cdots\beta_n) = (\pm 1)^{p(P)} c'(\alpha_1\cdots\alpha_n). \tag{18}$$

[2] P. Ehrenfest and J. R. Oppenheimer, Phys. Rev. **37**, 333 (1931).
[3] F. J. Dyson, Phys. Rev. **102**, 1217 (1956).

One has, by (1) and (2),

$$c'(\beta_1 \cdots \beta_n) = \int \varphi^*_{\beta_1}(y_1 \cdots y_l) \cdots \varphi^*_{\beta_n}(y_{ln-l+1} \cdots y_{ln}) \times \psi'(y_1 \cdots y_{ln})\, dy_1 \cdots dy_{ln}, \quad (19)$$

where

$$\psi'(y_1 \cdots y_n) = \sum_{\alpha_1 \cdots \alpha_n} c'(\alpha_1 \cdots \alpha_n) \times \varphi_{\alpha_1}(y_1 \cdots y_l) \cdots \varphi_{\alpha_n}(y_{ln-l+1} \cdots y_{ln}). \quad (20)$$

Hence,

$$\begin{aligned}
&\sum_{\beta_1 \cdots \beta_n} c'(\beta_1 \cdots \beta_n) J_P(\alpha_1 \cdots \alpha_n, \beta_1 \cdots \beta_n) \\
&= \sum_{\beta_1 \cdots \beta_n} \int dx_1 \cdots dx_{ln}\, dy_1 \cdots dy_{ln} \\
&\quad \times \varphi^*_{\alpha_n}(x_{ln-l+1} \cdots x_{ln}) \cdots \varphi^*_{\alpha_1}(x_1 \cdots x_l) \\
&\quad \times \varphi_{\beta_1}(Px_1 \cdots Px_l) \cdots \varphi_{\beta_n}(Px_{ln-l+1} \cdots Px_{ln}) \\
&\quad \times \varphi^*_{\beta_1}(y_1 \cdots y_l) \cdots \varphi^*_{\beta_n}(y_{ln-l+1} \cdots y_{ln}) \\
&\quad \times \psi'(y_1 \cdots y_{ln}). \qquad (21)
\end{aligned}$$

Use of the completeness relation (I.1) for the φ_α gives

$$\begin{aligned}
&\sum_{\beta_1 \cdots \beta_n} \varphi_{\beta_1}(Px_1 \cdots Px_l) \cdots \varphi_{\beta_n}(Px_{ln-l+1} \cdots Px_{ln}) \\
&\quad \times \varphi^*_{\beta_1}(y_1 \cdots y_l) \cdots \varphi^*_{\beta_n}(y_{ln-l+1} \cdots y_{ln}) \\
&= (l!)^{-n} \sum_{P_l} \sum_{Q_l} \cdots \sum_{R_l} (\pm 1)^{p(P_l)} (\pm 1)^{p(Q_l)} \cdots (\pm 1)^{p(R_l)} \\
&\quad \times \delta(Px_1 - P_l y_1) \cdots \delta(Px_l - P_l y_l) \\
&\quad \times \delta(Px_{l+1} - Q_l y_{l+1}) \cdots \delta(Px_{2l} - Q_l y_{2l}) \\
&\quad \times \cdots \delta(Px_{ln-l+1} - R_l y_{ln-l+1}) \cdots \delta(Px_{ln} - R_l y_{ln}), \qquad (22)
\end{aligned}$$

where $(P_l y_1 \cdots P_l y_l)$ is any permutation of $(y_1 \cdots y_l)$, $(Q_l y_{l+1} \cdots Q_l y_{2l})$ is any permutation of $(y_{l+1} \cdots y_{2l})$, $\cdots$, and $(R_l y_{ln-l+1} \cdots R_l y_{ln})$ is any permutation of $(y_{ln-l+1} \cdots y_{1n})$. Since c' satisfies the conditions (3) and (6), ψ' is symmetric or antisymmetric in $(y_1 \cdots y_{ln})$. Thus, on a renaming

$$y_1 \to P_l^{-1} y_1, \cdots, y_l \to P_l^{-1} y_l,$$
$$y_{l+1} \to Q_l^{-1} y_{l+1}, \cdots, y_{ln} \to R_l^{-1} y_{ln}$$

of the dummy integration variables, one finds

$$\begin{aligned}
&\sum_{\beta_1 \cdots \beta_n} c'(\beta_1 \cdots \beta_n) J_P(\alpha_1 \cdots \alpha_n, \beta_1 \cdots \beta_n) \\
&= (l!)^{-n} \sum_{P_l} \sum_{Q_l} \cdots \sum_{R_l} (\pm 1)^{p(P_l)} (\pm 1)^{p(Q_l)} \cdots (\pm 1)^{p(R_l)} \\
&\quad \times \int dx_1 \cdots dx_{ln}\, dy_1 \cdots dy_{ln} \\
&\quad \times \varphi^*_{\alpha_n}(x_{ln-l+1} \cdots x_{ln}) \cdots \varphi^*_{\alpha_1}(x_1 \cdots x_l) \\
&\quad \times \delta(Px_1 - y_1) \cdots \delta(Px_{ln} - y_{ln}) \\
&\quad \times \psi'(P_l^{-1} y_1 \cdots P_l^{-1} y_l \cdots R_l^{-1} y_{ln-l+1} \cdots R_l^{-1} y_{ln}) \\
&= \int dx_1 \cdots dx_{ln}\, dy_1 \cdots dy_{ln} \\
&\quad \times \varphi^*_{\alpha_n}(x_{ln-l+1} \cdots x_{ln}) \cdots \varphi^*_{\alpha_1}(x_1 \cdots x_l) \\
&\quad \times \delta(Px_1 - y_1) \cdots \delta(Px_{ln} - y_{ln}) \psi'(y_1 \cdots y_{ln}) \\
&= \int \varphi^*_{\alpha_1}(x_1 \cdots x_l) \cdots \varphi^*_{\alpha_n}(x_{ln-l+1} \cdots x_{ln}) \\
&\quad \times \psi'(Px_1 \cdots Px_{ln})\, dx_1 \cdots dx_{ln} \\
&= (\pm 1)^{p(P)} c'(\alpha_1 \cdots \alpha_n). \qquad (23)
\end{aligned}$$

Substitution of (18) into (17) then gives

$$\begin{aligned}
\langle \psi \mid \psi' \rangle &= [(ln)!]^{-1} \sum_{\alpha_1 \cdots \alpha n} c^*(\alpha_1 \cdots \alpha_n) c'(\alpha_1 \cdots \alpha_n) \sum_P 1 \\
&= \sum_{\alpha_1 \cdots \alpha_n} c^*(\alpha_1 \cdots \alpha_n) c'(\alpha_1 \cdots \alpha_n). \qquad (24)
\end{aligned}$$

On the other hand, it is a trivial consequence of (13), Wick's theorem, and (3) that

$$\begin{aligned}
(\psi \mid \psi') &= (n!)^{-1} \sum_{\alpha_1 \cdots \alpha_n} \sum_{\beta_1 \cdots \beta_n} c^*(\alpha_1 \cdots \alpha_n) \\
&\quad \times c'(\beta_1 \cdots \beta_n) (0 \mid a_{\alpha_n} \cdots a_{\alpha_1} a^\dagger_{\beta_1} \cdots a^\dagger_{\beta_n} \mid 0) \\
&= \sum_{\alpha_1 \cdots \alpha_n} c^*(\alpha_1 \cdots \alpha_n) c'(\alpha_1 \cdots \alpha_n). \qquad (25)
\end{aligned}$$

This establishes (14).

ACKNOWLEDGMENTS

I am very grateful to W. C. Schieve for impressing upon me the fact that the proof of the equivalence of the physical and ideal state spaces is not complete without a proof of (14).

This work was supported by the National Science Foundation under Grant GP-7689.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Pair Occupation, Fermi Condensation, and Phase Transitions in Many-Fermion Systems

M. D. GIRARDEAU
Institute of Theoretical Science and Department of Physics
University of Oregon, Eugene, Oregon

(Received 17 March 1969)

For a system of $2n$ fermions it is shown that the occupation number n_α of any fermion-pair state $\varphi_\alpha(x_1x_2)$ is $n_\alpha = n\int \varphi_\alpha^*(x_1x_2)\rho_2(x_1x_2, x_1'x_2')\varphi_\alpha(x_1'x_2')\,dx_1\,dx_2\,dx_1'\,dx_2'$, where ρ_2, assumed normalized to unity, is the two-particle density matrix. The known upper bound on the largest eigenvalue of ρ_2 implies that $n_\alpha \leq \frac{1}{2}$ in the thermodynamic limit $n \to \infty$, equality being approached, with suitable φ_α, only for certain limiting BCS states ψ_{BCS}. Bose condensation of fermion pairs, in the sense of macroscopic occupation of any n_α, is impossible. Fermi condensation into φ_α is defined to be present if $n_\alpha > 0$ in the thermodynamic limit. It occurs, for example, for suitable BCS states, but for a normal Fermi system all n_α are of order n^{-1} or smaller. It is argued that the physical criterion for ψ_{BCS} exhibiting Fermi condensation is that the pair state from which it is constructed must have a bound-state component of range $\lesssim k_F^{-1}$, where k_F is the Fermi momentum $(3\pi^2\rho)^{\frac{1}{3}}$. The maximally-occupied pair state is the eigenfunction of ρ_2 belonging to the largest eigenvalue, associated with off-diagonal long-range order of the type defined by Yang. A formula for n_0 for the original BCS state is derived. Some remarks are made concerning the interpretation of the Fermi condensation as a superconducting transition. The analysis is generalized to occupation of l-fermion states. It is conjectured that, when Fermi condensation first sets in at a given even l, this is associated with the formation of bound states of l fermions, and a formula for the maximal occupation of such states is exhibited. The implications, for the theory of liquid ^{4}He, of the fact that a helium atom contains electrons are examined. It is shown that Bose condensation into a single-^{4}He-atom state is impossible, but a Fermi condensation similar in some respects to that in a superconductor can and probably does occur. It is argued that the mechanism preventing Bose condensation in superconductors and liquid ^{4}He lies in the effect of collisions, acting via the exclusion principle, in causing virtual internal excitations.

1. INTRODUCTION

Consider a system of $2n$ fermions described by a normalized wavefunction $\psi(x_1 \cdots x_{2n})$, where each x_j stands for both position and spin variables, and $\int dx_j$ denotes integration over position and summation over spin variables. The same state has a quantized-field representation in terms of a normalized state vector $|\psi\rangle$ given by

$$|\psi\rangle = [(2n)!]^{-\frac{1}{2}}\int dx_1 \cdots dx_{2n}\psi(x_1 \cdots x_{2n}) \times \psi^\dagger(x_1)\cdots\psi^\dagger(x_{2n})|0\rangle, \quad (1)$$

where the $\psi^\dagger$ are fermion creation operators. Given any set of two-fermion wavefunctions $\varphi_\alpha(x_1x_2)$ which are antisymmetric, orthonormal, and complete, relative to the boundary conditions satisfied by ψ, one can expand ψ in the form[1]

$$\psi(x_1 \cdots x_{2n}) = \sum_{\alpha_1 \cdots \alpha_n} c(\alpha_1 \cdots \alpha_n) \times \varphi_{\alpha_1}(x_1x_2)\varphi_{\alpha_2}(x_3x_4)\cdots\varphi_{\alpha_n}(x_{2n-1}x_{2n}), \quad (2)$$

where

$$c(\alpha_1 \cdots \alpha_n) = \int \varphi_{\alpha_1}^*(x_1x_2)\cdots\varphi_{\alpha_n}^*(x_{2n-1}x_{2n}) \times \psi(x_1 \cdots x_{2n})\,dx_1 \cdots dx_{2n}. \quad (3)$$

[1] M. D. Girardeau, J. Math. Phys. 4, 1096 (1963). Equation (n) of this paper will be denoted by (I.n).

The antisymmetry of ψ implies that c is symmetric in $\alpha_1 \cdots \alpha_n$, corresponding to exchange of fermion pairs, and in addition satisfies the subsidiary condition[1]

$$\sum_{\alpha\beta}\sum_{p<q}^{n}(\alpha_p\alpha_q|\,I\,|\alpha\beta)c(\alpha_1 \cdots \alpha_{p-1}\alpha\alpha_{p+1}\cdots\alpha_{q-1}\beta\alpha_{q+1}\cdots\alpha_n) = -\tfrac{1}{2}n(n-1)c(\alpha_1 \cdots \alpha_n), \quad (4)$$

corresponding to pair-breaking fermion exchanges. Here the exchange matrix is defined by

$$(\alpha\beta|\,I\,|\gamma\delta) = \int \varphi_\alpha^*(x_1x_2)\varphi_\beta^*(x_1'x_2')\varphi_\gamma(x_1'x_2) \times \varphi_\delta(x_1x_2')\,dx_1\,dx_2\,dx_1'\,dx_2'. \quad (5)$$

The antisymmetry of the $\varphi_\alpha(xx')$ in x and x', symmetry of c in $\alpha_1 \cdots \alpha_n$, and the subsidiary condition (4) are also sufficient to ensure that a state of the form (2) is totally antisymmetric.[1]

Substitution of (2) into (1) yields

$$|\psi\rangle = [2^n/(2n)!]^{\frac{1}{2}}\sum_{\alpha_1 \cdots \alpha_n} c(\alpha_1 \cdots \alpha_n)A_{\alpha_1}^\dagger \cdots A_{\alpha_n}^\dagger|0\rangle, \quad (6)$$

where the $A_\alpha^\dagger$ are fermion-pair creation operators

$$A_\alpha^\dagger = 2^{-\frac{1}{2}}\int dx_1\,dx_2\varphi_\alpha(x_1x_2)\psi^\dagger(x_1)\psi^\dagger(x_2). \quad (7)$$

They satisfy the commutation relations

$$[A_\alpha, A_\beta] = [A_\alpha^\dagger, A_\beta^\dagger] = 0,$$
$$[A_\alpha, A_\beta^\dagger] = \delta_{\alpha\beta} - 2\int dx_1\, dx_2\, dy \varphi_\alpha^*(x_1x_2) \times \varphi_\beta(x_1\, y)\psi^\dagger(y)\psi(x_2). \quad (8)$$

Because of the extra operator term in $[A_\alpha, A_\beta^\dagger]$, the fermion pairs are not bosons, and $A_\alpha^\dagger A_\alpha$ cannot be validly interpreted as a boson occupation-number operator.

This difficulty can be overcome by introduction of the ideal state space. It has been shown[1,2] that, if one defines

$$|\psi) = (n!)^{-\frac{1}{2}} \sum_{\alpha_1 \cdots \alpha_n} c(\alpha_1 \cdots \alpha_n) a_{\alpha_1}^\dagger \cdots a_{\alpha_n}^\dagger |0), \quad (9)$$

where c is the same as in (6) and the a_α and $a_\alpha^\dagger$ satisfy ordinary Bose relations

$$\begin{aligned} a_\alpha |0) &= 0, \\ [a_\alpha, a_\beta] &= [a_\alpha^\dagger, a_\beta^\dagger] = 0, \\ [a_\alpha, a_\beta^\dagger] &= \delta_{\alpha\beta}, \end{aligned} \quad (10)$$

then all inner products are preserved, i.e.,

$$(\psi, \psi') = \langle \psi \mid \psi' \rangle = (\psi \mid \psi'), \quad (11)$$

provided that c satisfies the subsidiary condition (4). Here ψ and ψ' are ordinary Schrödinger wavefunctions, $|\psi\rangle$ and $|\psi'\rangle$ are expressed in the form (6) with coefficients c and c' given by (3), and $|\psi)$ and $|\psi')$ are the corresponding ideal state vectors (9). The method of transforming physical operators into the ideal state space has also been given.[1] Hence, all calculations can be performed in the ideal state space.

2. FERMION-PAIR OCCUPATION NUMBERS IN A PURE STATE

The ideal-state-space operator $N_\alpha = a_\alpha^\dagger a_\alpha$ is a boson occupation number by virtue of (10), so that its expectation value can validly be interpreted as the mean occupation number n_α of the fermion-pair state φ_α:

$$n_\alpha = (\psi| a_\alpha^\dagger a_\alpha |\psi). \quad (12)$$

Substitution from (9) and use of Wick's theorem together with the symmetry of $c(\alpha_1 \cdots \alpha_n)$ gives

$$n_\alpha = n \sum_{\alpha_1 \cdots \alpha_{n-1}} |c(\alpha_1 \cdots \alpha_{n-1}\alpha)|^2. \quad (13)$$

This is in agreement with a physical interpretation of $c(\alpha_1 \cdots \alpha_n)$ as the probability amplitude for observation of fermion pairs in states $\varphi_{\alpha_1} \cdots \varphi_{\alpha_n}$.

[2] M. D. Girardeau, J. Math. Phys. **11**, 682 (1970) (preceding article).

Substitution of the expression (3) for c yields

$$\begin{aligned} n_\alpha = n \sum_{\alpha_1 \cdots \alpha_{n-1}} \int & dx_1 \cdots dx_{2n}\, dy_1 \cdots dy_{2n} \\ & \times \varphi_{\alpha_1}(x_1x_2) \cdots \varphi_{\alpha_{n-1}}(x_{2n-3}x_{2n-2})\varphi_\alpha(x_{2n-1}x_{2n}) \\ & \times \varphi_{\alpha_1}^*(y_1y_2) \cdots \varphi_{\alpha_{n-1}}^*(y_{2n-3}y_{2n-2})\varphi_\alpha^*(y_{2n-1}y_{2n}) \\ & \times \psi^*(x_1 \cdots x_{2n})\psi(y_1 \cdots y_{2n}). \end{aligned} \quad (14)$$

The sums over $\alpha_1 \cdots \alpha_{n-1}$ and then the integrals over $y_1 \cdots y_{2n-2}$ can be performed with the aid of the closure relation (I.1).[1] Making use of the antisymmetry of ψ, one finds that

$$\begin{aligned} n_\alpha = n \int & dx_1 \cdots dx_{2n}\, dy_{2n-1}\, dy_{2n} \\ & \times \varphi_\alpha(x_{2n-1}x_{2n})\varphi_\alpha^*(y_{2n-1}y_{2n}) \\ & \times \psi^*(x_1 \cdots x_{2n})\psi(x_1 \cdots x_{2n-2}y_{2n-1}y_{2n}). \end{aligned} \quad (15)$$

Integration over $x_1 \cdots x_{2n-2}$ yields the two-particle density matrix of ψ, defined as

$$\rho_2(x_1x_2, x_1'x_2') = \int \psi(x_1 \cdots x_{2n})\psi^*(x_1'x_2'x_3 \cdots x_{2n})\, dx_3 \cdots dx_{2n}. \quad (16)$$

Thus, we obtain a formula for n_α in terms of ρ_2 and φ_α:

$$n_\alpha = n \int \varphi_\alpha^*(x_1x_2)\rho_2(x_1x_2, x_1'x_2') \times \varphi_\alpha(x_1'x_2')\, dx_1\, dx_2\, dx_1'\, dx_2'. \quad (17)$$

This bears a remarkable resemblance to the formula

$$n_k = n \int \varphi_k^*(x)\rho_1(x, x')\varphi_k(x')\, dx\, dx' \quad (18)$$

for the mean occupation number n_k of an orbital φ_k in an n-particle (boson or fermion) system with one-particle density matrix ρ_1. Summation of (17) over α, with the aid of the closure relation (I.1) for the φ_α and the normalization condition

$$\int \rho_2(x_1x_2, x_1x_2)\, dx_1\, dx_2 = 1, \quad (19)$$

yields

$$\sum_\alpha n_\alpha = n, \quad (20)$$

in accordance with one's intuitive expectation that the number of fermion pairs in a system of $2n$ fermions is n.

It is known[3–5] that the largest eigenvalue $\lambda_{\max}$ of ρ_2 for a system of $2n$ fermions has the following

[3] F. Sasaki, Quantum Chemistry Group, Upsala, Report No. 77 (1962) and Phys. Rev. **138**, B1338 (1965).
[4] C. N. Yang, Rev. Mod. Phys. **34**, 694 (1962).
[5] A. J. Coleman, Rev. Mod. Phys. **35**, 668 (1963), Sec. 7.

upper bound:

$$\lambda_{\max} < (2n-1)^{-1}. \tag{21}$$

The integral in (17) is the expectation value of ρ_2 in the state φ_α and cannot, therefore, exceed $\lambda_{\max}$; hence,

$$n_\alpha \leq n\lambda_{\max} < n/(2n-1). \tag{22}$$

3. STATISTICAL MECHANICS

For a system described by a statistical ensemble of orthonormal states $\psi_i(x_1 \cdots x_{2n})$ with statistical weights w_i satisfying

$$w_i \geq 0, \quad \sum_i w_i = 1, \tag{23}$$

(12) is replaced by

$$n_\alpha = \sum_i w_i(\psi_i|\, a_\alpha^\dagger a_\alpha\, |\psi_i). \tag{24}$$

Then, Eq. (17) is still valid with

$$\rho_2(x_1x_2, x_1'x_2') = \int \rho_{2n}(x_1 \cdots x_{2n}, x_1'x_2'x_3 \cdots x_{2n})\, dx_3 \cdots dx_{2n}, \tag{25}$$

where the $2n$-fermion density matrix

$$\rho_{2n}(x_1 \cdots x_{2n}, x_1' \cdots x_{2n}') = \sum_i w_i\psi_i(x_1 \cdots x_{2n})\psi_i^*(x_1' \cdots x_{2n}'). \tag{26}$$

Defining the density operator

$$\rho = \sum_i w_i\, |\psi_i\rangle\langle\psi_i|, \tag{27}$$

where the $|\psi_i\rangle$ are quantized-field-representation states (1), one can write[4]

$$\rho_2(x_1x_2, x_1'x_2') = [2n(2n-1)]^{-1} \times \mathrm{Tr}\,[\psi(x_1)\psi(x_2)\rho\psi^\dagger(x_2')\psi^\dagger(x_1')]. \tag{28}$$

4. BOSE VS FERMI CONDENSATION

Bose–Einstein condensation into the pair state φ_α is present if and only if

$$\text{lim therm } (n_\alpha/n) = f_\alpha > 0; \tag{29}$$

f_α is called the *condensed fraction*. Here "lim therm" denotes the thermodynamic limit, i.e., a limit where $n \to \infty$ and the system volume $\Omega \to \infty$ in such a way that the density n/Ω remains constant. However, in view of the upper bound (22), one has

$$f_\alpha = 0, \tag{30}$$

i.e., *Bose–Einstein condensation of fermion pairs cannot occur*. The fact that *complete* Bose condensation, i.e., $f_\alpha = 1$, cannot occur in Fermi systems was pointed out before.[6] However, the result (30) is much stronger. It is implicit in some remarks of Coleman.[6]

In view of (22), one has

$$\text{lim therm } n_\alpha \leq \tfrac{1}{2} \tag{31}$$

for any pair state φ_α. The attainment of this upper bound in the thermodynamic limit is an exceptional occurrence.[5] In fact, in a normal Fermi system n_α is only of order n^{-1} and, hence, vanishes in the thermodynamic limit, as we shall see in Sec. 7. We shall call the unusual situation where it does not vanish *Fermi condensation:*

$$\text{lim therm } n_\alpha > 0 \Leftrightarrow \text{Fermi condensation into } \varphi_\alpha. \tag{32}$$

It corresponds, according to (22), to a largest eigenvalue of order n^{-1} (with our normalization) for ρ_2, which Yang has shown[4] to be associated with off-diagonal long-range order (ODLRO) of ρ_2. For a normal Fermi system the largest eigenvalue of ρ_2 is only of order n^{-2}.

The distinction between Fermi condensation and true Bose condensation can be seen in another way by evaluating the expectation value of the operator $A_\alpha^\dagger A_\alpha$, which would be a Bose occupation number operator if the operators (7) obeyed Bose commutation relations. By (7),

$$\langle A_\alpha^\dagger A_\alpha\rangle = \tfrac{1}{2}\int \varphi_\alpha(x_1'x_2')\langle\psi^\dagger(x_1')\psi^\dagger(x_2')\psi(x_2)\psi(x_1)\rangle \times \varphi_\alpha^*(x_1x_2)\, dx_1\, dx_2\, dx_1'\, dx_2', \tag{33}$$

where the angular brackets denote an expectation value in the case of a pure state and an ensemble average in the case of a mixed state. Insertion of (28) and comparison with (17) give

$$\langle A_\alpha^\dagger A_\alpha\rangle = n(2n-1)\int \varphi_\alpha^*(x_1x_2)\rho_2(x_1x_2, x_1'x_2') \times \varphi_\alpha(x_1'x_2')\, dx_1\, dx_2\, dx_1'\, dx_2' = (2n-1)n_\alpha. \tag{34}$$

The upper bound (31) implies that

$$\text{lim therm } [n^{-1}\langle A_\alpha^\dagger A_\alpha\rangle] = 1. \tag{35}$$

One might jump to the conclusion, by comparison of (35) with (29), that $\langle A_\alpha^\dagger A_\alpha\rangle$ ought, therefore, to be interpreted as a close analog of a Bose occupation number and that attainment of the upper bound (35) represents extreme Bose condensation. That such an interpretation is incorrect can be seen from the normalization condition. Making use of the completeness relation for the φ_α and the normalization

[6] A. J. Coleman, Phys. Rev. Letters 13, 406 (1964).

condition (19) on ρ_2, one finds

$$\sum_\alpha \langle A_\alpha^\dagger A_\alpha\rangle = n(2n-1)\int \rho_2(x_1x_2, x_1x_2)\,dx_1\,dx_2$$
$$= n(2n-1) = \binom{2n}{2}. \tag{36}$$

Thus, even when a single $\langle A_\alpha^\dagger A_\alpha\rangle$ attains its upper bound n, all the other $\langle A_\alpha^\dagger A_\alpha\rangle$ give the overwhelming contribution to the sum rule (36). This is quite different from Bose condensation, for which the sum rule is exhausted by a single orbital occupation number n_α when n_α attains its upper bound n for a system of n bosons.

5. DENSITY MATRICES AND PAIR OCCUPATION IN BCS STATES

Since[3–5] the states with $n_\alpha = \frac{1}{2}$ for some φ_α are of BCS[7] type, we now investigate the n_α for such states. We define

$$\psi_{\rm BCS}(x_1\cdots x_{2n}) = \mathcal{C}A_{2n}[g(x_1x_2)g(x_3x_4)\cdots g(x_{2n-1}x_{2n})], \tag{37}$$

where $\mathcal{C}$ is a normalization constant, A_{2n} is the antisymmetrizer with respect to $x_1\cdots x_{2n}$, and g is a normalized and antisymmetric fermion-pair function. Such a state is called an "antisymmetrized geminal power" by Coleman[5,8]; our terminology $\psi_{\rm BCS}$ is motivated by the fact that, for any such state, there exists[9–11] a basis of orbitals in terms of which (37) takes on the pairing form characteristic of the variational trial states of the BCS theory.[7] The state (37) has a quantized-field representation of the form (1), which can be written

$$|\psi_{\rm BCS}\rangle = [2^n/(2n)!]^{\frac{1}{2}}\mathcal{C}(A^\dagger)^n|0\rangle, \tag{38}$$

with

$$A^\dagger = 2^{-\frac{1}{2}}\int dx_1\,dx_2 g(x_1x_2)\psi^\dagger(x_1)\psi^\dagger(x_2). \tag{39}$$

In order to transform (38) into the canonical BCS form, one considers $g(xx')$ as the kernel of an integral operator G on functions $f(x)$:

$$Gf(x) \equiv \int g(xx')f(x')\,dx'. \tag{40}$$

Defining G^* and $\tilde{G}$ as the integral operators with kernels $g^*(xx')$ and $g(x'x)$, and $G^\dagger = \tilde{G}^*$ as usual, one sees that

$$\tilde{G} = -G, \quad G^\dagger = -G^* \tag{41}$$

as a result of the antisymmetry of G. Since

$$K \equiv G^\dagger G = -G^*G \tag{42}$$

is Hermitian and positive semidefinite, its eigenvalues λ_k are real and nonnegative:

$$K\varphi_k = \lambda_k\varphi_k, \quad \lambda_k \geq 0, \tag{43}$$

where the $\varphi_k(x)$ are the eigenfunctions of K, orthonormal in the usual sense:

$$(\varphi_k, \varphi_{k'}) = \int \varphi_k^*(x)\varphi_{k'}(x)\,dx = \delta_{kk'}. \tag{44}$$

The number r of such eigenfunctions is, by definition, the rank of K. The set $\{\varphi_k\}$ is certainly not complete if r is finite, and it may not be complete even if $r = \infty$.

By a derivation paralleling those of Yang[9] and Bloch and Messiah,[10] it can be shown that the nonzero eigenvalues $\lambda_k > 0$ of K are at least twofold degenerate, and that the corresponding eigenfunctions $\varphi_{k\pm}$ can be chosen so that

$$G\varphi_{k\pm} = \mp\lambda_k^{\frac{1}{2}}\varphi_{k\mp}^*. \tag{45}$$

The desired canonical form of g is then

$$g(xx') = \sum_k \lambda_k^{\frac{1}{2}}[\varphi_{k+}^*(x)\varphi_{k-}^*(x') - \varphi_{k-}^*(x)\varphi_{k+}^*(x')]. \tag{46}$$

Substitution of (46) into (39) yields

$$A^\dagger = 2^{\frac{1}{2}}\sum_k \lambda_k^{\frac{1}{2}}a_{k+}^\dagger a_{k-}^\dagger \tag{47}$$

with fermion creation operators

$$a_{ks}^\dagger \equiv \int dx\,\varphi_{ks}^*(x)\psi^\dagger(x). \tag{48}$$

Substitution of (47) into (38) gives the standard paired form of a BCS state of $2n$ fermions.

The expressions for the one- and two-particle density matrices of such a state $|\psi_{\rm BCS}\rangle$ are well known[6]; in our notation they are given asymptotically for $n \to \infty$ by

$$\rho_1(x_1, x_1') = (2n)^{-1}\sum_k \left(\frac{p\lambda_k}{1+p\lambda_k}\right) \times [\varphi_{k+}^*(x_1)\varphi_{k+}(x_1') + \varphi_{k-}^*(x_1)\varphi_{k-}(x_1')],$$
$$\rho_2(x_1x_2, x_1'x_2') = \rho_1(x_1, x_1')\rho_1(x_2, x_2') - \rho_1(x_1, x_2')\rho_1(x_2, x_1') + \chi(x_1x_2)\chi^*(x_1'x_2'), \tag{49}$$

with

$$\chi(x_1x_2) = (2n)^{-1}\sum_k \left[\frac{(p\lambda_k)^{\frac{1}{2}}}{1+p\lambda_k}\right] \times [\varphi_{k+}^*(x_1)\varphi_{k-}^*(x_2) - \varphi_{k+}^*(x_2)\varphi_{k-}^*(x_1)]. \tag{50}$$

[7] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **108**, 1175 (1957). The fact that the $2n$-particle projection of a BCS state is of the form (37) was first shown by K. Nakamura [Progr. Theoret. Phys. (Kyoto) **21**, 713 (1959)].

[8] A. J. Coleman, J. Math. Phys. **6**, 1425 (1965).

[9] Ref. 4, Appendix A.

[10] C. Bloch and A. Messiah, Nucl. Phys. **39**, 95 (1962), Appendix 1.

[11] B. Zumino, J. Math. Phys. **3**, 1055 (1962).

The positive parameter p is determined by the condition

$$n^{-1}\sum_k \frac{p\lambda_k}{1+p\lambda_k} = 1. \tag{51}$$

A necessary condition for the asymptotic validity of the expressions (49) is that the orbitals φ_{ks} are spatially extended, as is the case for plane waves and, more generally, Bloch waves. Then, the φ_{ks} will be proportional to $\Omega^{-\frac{1}{2}}$, where Ω is the volume of the system, and each $\sum_k$ will introduce a factor of Ω. With the normalization corresponding to definition (16), ρ_1 will then be proportional to n^{-1} and ρ_2 to n^{-2} for large n.

Substitution of (49) into (17) gives the following expression for the mean occupation number n_α of any normalized and antisymmetric two-particle state φ_α:

$$n_\alpha = n\,|(\varphi_\alpha, \chi)|^2 + n'_\alpha, \tag{52}$$

where

$$(\varphi_\alpha, \chi) = \int \varphi_\alpha^*(x_1x_2)\chi(x_1x_2)\,dx_1\,dx_2 \tag{53}$$

and

$$n'_\alpha = 2n\int \varphi_\alpha^*(x_1x_2)\rho_1(x_1, x_1')\rho_1(x_2, x_2') \times \varphi_\alpha(x_1'x_2')\,dx_1\,dx_2\,dx_1'\,dx_2'. \tag{54}$$

For the special case that φ_α is chosen to be the pair function g occurring in the definition (37) of the BCS state, one has, by (46), (50), (51), (49), and the orthonormality of the φ_{ks},

$$(g, \chi) = n^{-1}p^{\frac{1}{2}}\sum_k \frac{\lambda_k}{1+p\lambda_k} = p^{-\frac{1}{2}} \tag{55}$$

and

$$n'_g = n^{-1}p^2\sum_k \frac{\lambda_k^3}{(1+p\lambda_k)^2}. \tag{56}$$

Thus, for $\varphi_\alpha = g$,

$$n_g = np^{-1} + n^{-1}p^2\sum_k \frac{\lambda_k^3}{(1+p\lambda_k)^2}. \tag{57}$$

6. FERMI CONDENSATION IN BCS STATES

For a system of fermions, $\rho_1(x, x')$ cannot exhibit ODLRO because this would imply an eigenvalue of ρ_1 exceeding the known upper bound.[4] Equivalently, ODLRO in ρ_1 would give $n_k = O(n)$ by (18) for suitable trial states φ_k, contradicting the limitation $n_k \leq 1$ of the exclusion principle. Thus, $\rho_1(x, x')$ falls to zero when $|\mathbf{r} - \mathbf{r}'|$ exceeds a certain volume-independent range; since it is of order n^{-1} when $|\mathbf{r} - \mathbf{r}'|$ is within this range, one concludes from (54) that n'_α is of order n^{-1} regardless of whether φ_α is bound or spatially extended. This is typical of the n_α of an ideal Fermi gas or, more generally, of a normal Fermi system. Thus, Fermi condensation in the sense (36) can only arise through the contribution of the first term in (52).

For orientation, consider first the case of a single Slater determinant. As has been pointed out by Coleman[5,6] and is, in fact, implicit in the interpretation of the normal state in the BCS theory,[7] a single Slater determinant is a special case of a BCS state (37), corresponding to a single $2n$-fold degenerate nonzero eigenvalue of K, with value $(2n)^{-1}$:

$$\lambda_k = (2n)^{-1}, \quad k \in \mathcal{S}_F, \quad \lambda_k = 0, \quad k \notin \mathcal{S}_F, \tag{58}$$

where $\mathcal{S}_F$, the Fermi sea, is a set of n k-values. Substitution of (58) into (46), with any choice of $2n$ orthonormal orbitals $\varphi_{k\pm}$, gives a K with the eigenvalues (58), and substitution of this g into (37) gives the Slater determinant constructed from these orbitals, as is easily seen from (47), (38), and the fact that $(a_{ks}^\dagger)^2 = 0$. In order to satisfy (51), one has to take $p \to \infty$. Then, χ vanishes by (50), so that Fermi condensation cannot occur in any pair state φ_α, as expected for a normal Fermi system.

More generally, one can construct g so as to have a single $2m$-fold degenerate nonzero eigenvalue, where m is any integer $\geq n$, by choosing

$$\lambda_k = (2m)^{-1}, \quad k \in \mathcal{S}, \quad \lambda_k = 0, \quad k \notin \mathcal{S}, \tag{59}$$

where $\mathcal{S}$ is any set of m k-values and $\{\varphi_{k\pm} \mid k \in \mathcal{S}\}$ is any set of $2m$ orthonormal orbitals. Such BCS states are not single Slater determinants if $m > n$; they are said to be of "extreme type" by Coleman,[5] since they have the property that, for given rank $r = 2m$ of K, these states maximize the largest eigenvalue of ρ_2 and, hence, maximize the occupation number n_α of a suitable pair state φ_α. Equation (51) is satisfied with

$$p = 2nm/(m-n). \tag{60}$$

Then, by (57),

$$n_g = (m-n)/2m + O(n^{-1}). \tag{61}$$

Since (57) is only asymptotically correct for large n, the same is true of (61). The upper bound (31) on n_g can be approached arbitrarily closely by letting both n and m get arbitrarily large with $m \gg n$:

$$\lim_{n\to\infty}\lim_{m\to\infty} n_g = \lim_{n\to\infty}[\tfrac{1}{2} + O(n^{-1})] = \tfrac{1}{2}. \tag{62}$$

This agrees with the known fact[4,5] that the upper bound (21) on the largest eigenvalue of ρ_2 is approached arbitrarily closely by BCS states of extreme type with arbitrarily large rank $r = 2m$. Yang has shown more generally that the largest eigenvalue of ρ_2 for BCS states of extreme type with rank r is[4] (with

our normalization of ρ_2)

$$\lambda_{\max} = (r - 2n + 2)/(2n - 1)r. \tag{63}$$

For large n, when the 2 in the numerator and the 1 in the denominator are negligible, it follows that the expression (61) for n_g is equal to its upper bound $n\lambda_{\max}$ [Eq. (22)], showing that g is an eigenfunction of ρ_2 belonging to the eigenvalue $\lambda_{\max}$. This is, in fact, exactly true for BCS states of extreme type even if n is not large, as has been shown by Coleman.[8]

Since BCS states of extreme type are a very special case, it is desirable to obtain criteria for Fermi condensation in general BCS states (37). We first ask the question: Given a BCS state, for what two-particle function φ_α is n_α largest? We denote the function φ_α maximizing n_α by φ_0. In view of (52), the answer is trivial in case n'_α is negligible, since the overlap (φ_0, χ) is maximized by choosing

$$\varphi_0(x_1x_2) = (\chi, \chi)^{-\frac{1}{2}}\chi(x_1x_2). \tag{64}$$

Since we have shown that n'_α is of order n^{-1}, it will indeed be negligible if Fermi condensation into the state φ_0 occurs, in which case n_0 will be of order unity. Dropping the negligible term n'_0, one has in this case, by (52),

$$n_0 = n(\chi, \chi). \tag{65}$$

The normalization of χ can be evaluated with the aid of (50):

$$(\chi, \chi) = \tfrac{1}{2}n^{-2}\sum_k \frac{p\lambda_k}{(1 + p\lambda_k)^2}; \tag{66}$$

hence,

$$n_0 = (2n)^{-1}\sum_k \frac{p\lambda_k}{(1 + p\lambda_k)^2}. \tag{67}$$

The criterion (32) for Fermi condensation will be satisfied if the summation is proportional to n. In the extreme case, one can easily verify with the aid of (59)–(61) that $n_0 = n_g$. In fact, for an extreme state, it follows from (46) and (50) that χ is proportional to g and, hence, $\varphi_0 = g$, in agreement with our previous statement.

More explicit results can be found by specializing to the case in which g is translationally invariant and a spin singlet:

$$\begin{aligned} g(xx') &= (2\Omega)^{-\frac{1}{2}}g(\mathbf{r} - \mathbf{r}')(\delta_{\sigma\uparrow}\delta_{\sigma'\downarrow} - \delta_{\sigma\downarrow}\delta_{\sigma'\uparrow}) \\ &= 2^{-\frac{1}{2}}\Omega^{-\frac{1}{2}}(\delta_{\sigma\uparrow}\delta_{\sigma'\downarrow} - \delta_{\sigma\downarrow}\delta_{\sigma'\uparrow})\sum_{\mathbf{k}} g_{\mathbf{k}}e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}, \end{aligned} \tag{68}$$

where $g_{\mathbf{k}}$ is the Fourier transform of $g(\mathbf{r})$, assuming periodic boundary conditions with periodicity volume Ω:

$$g_{\mathbf{k}} = \int_\Omega g(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}}\,d^3r \tag{69}$$

and $x = (\mathbf{r}, \sigma)$, $x' = (\mathbf{r}', \sigma')$. We assume that $g(\mathbf{r})$ is symmetric,

$$g(-\mathbf{r}) = g(\mathbf{r}), \quad g_{-\mathbf{k}} = g_{\mathbf{k}}, \tag{70}$$

so that $g(xx')$ is antisymmetric. This is the type of state assumed in the original BCS theory.[7] One has for the integral operator G with kernel $g(xx')$

$$\begin{aligned} G(e^{i\mathbf{k}\cdot\mathbf{r}}\delta_{\sigma\uparrow}) &= -(2\Omega)^{-\frac{1}{2}}g_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{r}}\delta_{\sigma\downarrow}, \\ G(e^{i\mathbf{k}\cdot\mathbf{r}}\delta_{\sigma\downarrow}) &= (2\Omega)^{-\frac{1}{2}}g_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{r}}\delta_{\sigma\uparrow}. \end{aligned} \tag{71}$$

It then follows from (42) that $e^{i\mathbf{k}\cdot\mathbf{r}}\delta_{\sigma\uparrow}$ and $e^{i\mathbf{k}\cdot\mathbf{r}}\delta_{\sigma\downarrow}$ are eigenfunctions of K with eigenvalue

$$\lambda_{\mathbf{k}} = (2\Omega)^{-1}\,|g_{\mathbf{k}}|^2. \tag{72}$$

The choice of eigenstates of K within the fourfold-degenerate subspace $(\mathbf{k}\uparrow)$, $(\mathbf{k}\downarrow)$, $(-\mathbf{k}\uparrow)$, $(-\mathbf{k}\downarrow)$ has to be made so that Eqs. (55) are satisfied. One finds that this is the case with

$$\begin{aligned} \varphi_{\mathbf{k}+}(x) &= \Omega^{-\frac{1}{2}}e^{i\mathbf{k}\cdot\mathbf{r}}\delta_{\sigma\uparrow}, \\ \varphi_{\mathbf{k}-}(x) &= \Omega^{-\frac{1}{2}}(g^*_{\mathbf{k}}/|g_{\mathbf{k}}|)e^{-i\mathbf{k}\cdot\mathbf{r}}\delta_{\sigma\downarrow}. \end{aligned} \tag{73}$$

The normalization condition on the $\lambda_{\mathbf{k}}$ is satisfied by virtue of the normalization condition for g:

$$\|g\|^2 = \int |g(\mathbf{r})|^2\,d^3r = \Omega^{-1}\sum_{\mathbf{k}} |g_{\mathbf{k}}|^2 = 2\sum_{\mathbf{k}} \lambda_{\mathbf{k}} = 1. \tag{74}$$

One then has by (67)

$$n_0 = (2n)^{-1}\sum_{\mathbf{k}} \frac{q\,|g_{\mathbf{k}}|^2}{(1 + q\,|g_{\mathbf{k}}|^2)^2}, \tag{75}$$

with $q = (2\Omega)^{-1}p$ determined, according to (51), by

$$n^{-1}\sum_{\mathbf{k}} \frac{q\,|g_{\mathbf{k}}|^2}{1 + q\,|g_{\mathbf{k}}|^2} = 1. \tag{76}$$

It then follows from (64), (65), and (50) that

$$\varphi_0(xx') = (2\Omega)^{-\frac{1}{2}}\varphi_0(\mathbf{r} - \mathbf{r}')(\delta_{\sigma\uparrow}\delta_{\sigma'\downarrow} - \delta_{\sigma\downarrow}\delta_{\sigma'\uparrow}), \tag{77}$$

with

$$\varphi_0(\mathbf{r}) = (n_0\rho)^{-\frac{1}{2}}\Omega^{-1}\sum_{\mathbf{k}} \frac{q^{\frac{1}{2}}g_{\mathbf{k}}}{1 + q\,|g_{\mathbf{k}}|^2}\,e^{i\mathbf{k}\cdot\mathbf{r}} \tag{78}$$

and $\rho = 2n/\Omega$, the number density of fermions.

Suppose that $g(\mathbf{r})$ is a bound state, i.e., that it falls to zero for $|\mathbf{r}| >$ some volume-independent range. Then its Fourier transform (69) will be volume independent. Making the usual replacement

$$\sum_{\mathbf{k}} \to (2\pi)^{-3}\Omega\int d^3k,$$

one finds in the thermodynamic limit $[n \to \infty, \Omega \to \infty, (2n/\Omega) \to \rho, 0 < \rho < \infty]$

$$\varphi_0(\mathbf{r}) = (2\pi)^{-3}\left(\frac{q}{n_0\rho}\right)^{\frac{1}{2}}\int \frac{g_{\mathbf{k}}}{1 + q\,|g_{\mathbf{k}}|^2}\,e^{i\mathbf{k}\cdot\mathbf{r}}\,d^3k, \tag{79}$$

with occupation number

$$n_0 = (2\pi)^{-3}\left(\frac{q}{\rho}\right)\int \frac{|g_{\mathbf{k}}|^2}{(1+q\,|g_{\mathbf{k}}|^2)^2}\,d^3k. \tag{80}$$

The parameter q is determined by

$$2(2\pi)^{-3}\left(\frac{q}{\rho}\right)\int \frac{|g_{\mathbf{k}}|^2}{1+q\,|g_{\mathbf{k}}|^2}\,d^3k = 1. \tag{81}$$

It is easy to see that the left side of (81) varies from 0 to ∞ as q varies from 0 to ∞. Hence, there is always a positive volume-independent solution q of (81). Substitution of this q into (79) and (80) gives a bound two-fermion state φ_0 with an occupation number n_0, which is positive and volume-independent and, hence, satisfies the criterion (32) for Fermi condensation. It follows from (80), (81), and the inequality

$$\frac{|g_{\mathbf{k}}|^2}{(1+q\,|g_{\mathbf{k}}|^2)^2} \leq \frac{|g_{\mathbf{k}}|^2}{1+q\,|g_{\mathbf{k}}|^2} \tag{82}$$

that

$$n_0 \leq \tfrac{1}{2}, \tag{83}$$

in agreement with (31).

In order to obtain more explicit results, consider the special case

$$g_{\mathbf{k}} = \text{const}, \quad \mathbf{k} \in \mathcal{S}; \quad g_{\mathbf{k}} = 0, \quad \mathbf{k} \notin \mathcal{S}, \tag{84}$$

where $\mathcal{S}$ is any region of $\mathbf{k}$ space. In view of (72), this is a special case of a state of extreme type, Eq. (59). The number of allowed $\mathbf{k}$ values in $\mathcal{S}$ is

$$m = (2\pi)^{-3}\Omega\omega, \tag{85}$$

where $\omega = \int_{\mathcal{S}} d^3k$ is the volume of $\mathcal{S}$. Thus, by (61),

$$n_0 = \tfrac{1}{2}[1 - \tfrac{1}{2}(2\pi)^3(\rho/\omega)]. \tag{86}$$

The same result can be found from (80) by determining the constant in (84) so that $g(\mathbf{r})$ is normalized; the corresponding expressions for q, g, and φ_0 are

$$q = \tfrac{1}{2}\rho[1 - \tfrac{1}{2}(2\pi)^3(\rho/\omega)]^{-1} \tag{87}$$

and

$$g(\mathbf{r}) = \varphi_0(\mathbf{r}) = (2\pi)^{-\frac{3}{2}}\omega^{-\frac{1}{2}}\int_{\mathcal{S}} e^{i\mathbf{k}\cdot\mathbf{r}}\,d^3k. \tag{88}$$

The limit $(\rho/\omega) \to 0$ corresponds to complete Fermi condensation ($n_0 = \frac{1}{2}$). As (ρ/ω) increases, n_0 decreases, vanishing at the limiting value $(\rho/\omega) = 2(2\pi)^3$, which corresponds to a single Slater determinant with $\mathcal{S}$ as the filled Fermi sea; (86) is not correct for larger values of (ρ/ω), for which $|\psi_{\mathrm{BCS}}\rangle$ vanishes identically as a result of the exclusion principle. In the special case that $\mathcal{S}$ is a sphere of radius k_0 centered on the origin, one has

$$n_0 = \tfrac{1}{2}(1 - 3\pi^2\rho k_0^{-3}) \tag{89}$$

and $g(\mathbf{r})$ is a spherical Bessel function of range k_0^{-1}. The limit $k_0 \to \infty$ of complete Fermi condensation corresponds to zero range of $g(\mathbf{r})$, i.e., $g(\mathbf{r})$ infinitely tightly bound. The very-high-momentum components required are an obvious consequence of this tight binding. The opposite limit $n_0 \to 0$ corresponds to $k_0 \to k_F = (3\pi^2\rho)^{\frac{1}{3}}$, the top of the normal Fermi sea.

The dependence of n_0 on the range of $g(\mathbf{r})$ in the more general case is similar to the special case (89). Suppose that $g(\mathbf{r})$ has range a, and introduce $f(\mathbf{r}^*)$ and $f_{\mathbf{k}^*}$ defined by

$$g(\mathbf{r}) = a^{-\frac{3}{2}}f(\mathbf{r}^*), \quad \mathbf{r}^* = \mathbf{r}a^{-1},$$

$$f_{\mathbf{k}^*} = \int f(\mathbf{r}^*)e^{-i\mathbf{k}^*\cdot\mathbf{r}^*}\,d^3r^* = a^{-\frac{3}{2}}g_{\mathbf{k}}, \quad \mathbf{k}^* = ka. \tag{90}$$

Then, expansion of the integrands of (80) and (81) for small q gives

$$q = \tfrac{1}{2}\rho\left[1 + \tfrac{1}{2}(2\pi)^{-3}\rho a^3\int |f_{\mathbf{k}^*}|^4\,d^3k^* + O((\rho a^3)^2)\right],$$

$$n_0 = \tfrac{1}{2}\left[1 - \tfrac{1}{2}(2\pi)^{-3}\rho a^3\int |f_{\mathbf{k}^*}|^4\,d^3k^* + O((\rho a^3)^2)\right], \tag{91}$$

which shows that the upper limit $n_0 = \frac{1}{2}$ is approached as $\rho a^3 \to 0$. At first sight, this appears to contradict the previous result that a state of extreme type with rank $r \to \infty$ is needed to attain $n_0 = \frac{1}{2}$. However, as the range a of $g(\mathbf{r})$ approaches zero, the range $k_0 = a^{-1}$ of $g_{\mathbf{k}}$ approaches infinity. This infinitely slow variation of $g_{\mathbf{k}}$, hence $\lambda_{\mathbf{k}}$, means that the situation of states of extreme type ($\lambda_{\mathbf{k}}$ constant wherever it is nonzero) is approached, in a certain sense, in the limit $\rho a^3 \to 0$. The opposite limit $n_0 = 0$ is reached when the rank r of K, here equal to the number of nonzero $g_{\mathbf{k}}$, drops to $2n$. One can estimate r as the number of allowed $\mathbf{k}$ values in a sphere of radius equal to the range of $g_{\mathbf{k}}$:

$$r \sim (2\pi)^{-3}\Omega\tfrac{4}{3}\pi a^{-3}. \tag{92}$$

Thus, n_0 drops to zero at

$$\rho a^3 \sim (3\pi^2)^{-1}. \tag{93}$$

Since, in the case in which g is a bound state, n_0 drops to zero as the range of g increases to a critical value given approximately by (93), one might conjecture that Fermi condensation into an unbound pair state cannot occur. This, however, is not the case. The essential property of g necessary for Fermi condensation is that it contain a component of

sufficiently short range; the presence of a nonvanishing tail, which makes g unbound, does not prevent Fermi condensation. In order to see this, decompose $g(\mathbf{r})$, Eq. (68), in the form

$$g(\mathbf{r}) = \Omega^{-1}g_0 + g'(\mathbf{r}), \quad g'(\mathbf{r}) = \Omega^{-1}\sum_{\mathbf{k}}{}' g_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{r}}, \quad (94)$$

where the prime implies omission of the $\mathbf{k} = 0$ component from g'. Suppose that $g'(\mathbf{r})$ is of finite range, i.e., of the same general form as $g(\mathbf{r})$ in the previous paragraph. Then,

$$g_{\mathbf{k}} = \int g'(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}}\,d^3r = O(1), \quad \mathbf{k} \neq 0, \quad (95)$$

where $O(1)$ denotes a function independent of Ω (and n). Suppose, furthermore, that g_0 is of order $\Omega^{\frac{1}{2}}$. Then, $g(\mathbf{r})$ will differ from a bound state by the presence of a constant tail of order $\Omega^{-\frac{1}{2}}$. By a suitable choice of phase of g, one can write

$$g_0 = (w\Omega)^{\frac{1}{2}}, \quad 0 < w < 1. \quad (96)$$

Then, (81) and (80) are replaced by [cf. (76) and (75)]

$$\frac{2\rho^{-1}wq}{1 + w\Omega q} + 2(2\pi)^{-3}\rho^{-1}q\int\frac{|g_{\mathbf{k}}|^2}{1 + q\,|g_{\mathbf{k}}|^2}\,d^3k = 1 \quad (97)$$

and

$$n_0 = \frac{\rho^{-1}wq}{(1 + w\Omega q)^2} + (2\pi)^{-3}\rho^{-1}q\int\frac{|g_{\mathbf{k}}|^2}{(1 + q\,|g_{\mathbf{k}}|^2)^2}\,d^3k. \quad (98)$$

If $q = O(1)$, the first term in (97) is negligible and the equation for the determination of q becomes identical with (81). The first term in (98) is also negligible, so that (98) reduces to (80). The normalization condition on the $g_{\mathbf{k}}$ is affected by the presence of the tail:

$$\|g\|^2 = w + \Omega^{-1}\sum_{\mathbf{k}}{}'|g_{\mathbf{k}}|^2 = 1. \quad (99)$$

Generalizing (90) to

$$g_{\mathbf{k}} = (1 - w)^{\frac{1}{2}}a^{\frac{3}{2}}f_{\mathbf{k}^*}, \quad \mathbf{k} \neq 0, \quad (100)$$

where $f_{\mathbf{k}^*}$ has range $\sim a$ and is of order unity for $k^* = ka \lesssim 1$, one has in the thermodynamic limit

$$\|g\|^2 = w + (2\pi)^{-3}(1 - w)\int|f_{\mathbf{k}^*}|^2\,d^3k^* = 1, \quad (101)$$

so that the normalization condition is satisfied if

$$(2\pi)^{-3}\int|f_{\mathbf{k}^*}|^2\,d^3k^* = 1, \quad (102)$$

as is the case for (90) and (91). The expansions of q and n_0 for $\rho a^3 \ll 1$ are then

$$q = \tfrac{1}{2}(1 - w)^{-1}\rho\left[1 + \tfrac{1}{2}(2\pi)^{-3}\rho a^3\int|f_{\mathbf{k}^*}|^4\,d^3k^* + \cdots\right],$$
$$n_0 = \tfrac{1}{2}\left[1 - \tfrac{1}{2}(2\pi)^{-3}\rho a^3\int|f_{\mathbf{k}^*}|^4\,d^3k^* + \cdots\right]. \quad (103)$$

Thus, the dependence of n_0 on ρa^3 is the same as in the case (90) and (91). It is easy to see that this equality is not restricted to small ρa^3, but is exact. The presence of a nonvanishing tail does not, therefore, qualitatively affect the Fermi condensation. The general criterion for Fermi condensation in a state of the type $|\psi_{\mathrm{BCS}}\rangle$ appears to be that the pair function g contain a bound-state component of range $a \lesssim k_F^{-1}$, where k_F is the Fermi momentum $(3\pi^2\rho)^{\frac{1}{3}}$. This remains true no matter how small the weight $(1 - w)$ of the bound-state component, so long as $1 - w > 0$. This remarkable result can be understood by noting from (79), (64), and (65) that the unbound tail $(w\Omega)^{\frac{1}{2}}$ present in g is not mirrored in φ_0 and χ and, hence, not in the term $\chi\chi^*$ in (49) responsible for the ODLRO in ρ_2. In fact, for $\rho a^3 \to 0$, φ_0 and χ reduce to[12]

$$\varphi_0(\mathbf{r}) \xrightarrow[\rho a^3 \to 0]{} (2\pi)^{-3}\int f_{\mathbf{k}^*}e^{i\mathbf{k}^*\cdot\mathbf{r}^*}\,d^3k^* + O(\Omega^{-\frac{3}{2}}),$$
$$\chi(\mathbf{r}) \xrightarrow[\rho a^3 \to 0]{} (2n)^{-\frac{1}{2}}\varphi_0(\mathbf{r}) \quad (104)$$

and, thus, are bound and independent of w so long as $1 - w > 0$. These results are valid in the thermodynamic limit for fixed $w < 1$; one may not set $w = 1$, since $|\psi_{\mathrm{BCS}}\rangle$ would then vanish identically because K would have rank $< 2n$.

We conclude this section by evaluating n_0 for the variational state of the original BCS theory.[7] It follows from the derivation of (49), BCS[7] Eq. (2.16), and (72), with $p = (2\Omega)^{-1}q$, that

$$q\,|g_{\mathbf{k}}|^2 = h_{\mathbf{k}}/(1 - h_{\mathbf{k}}). \quad (105)$$

The normalization condition (81) then becomes

$$2(2\pi)^{-3}\rho_s^{-1}\int h_{\mathbf{k}}\,d^3k = 1, \quad (106)$$

and Eq. (80) for n_0 becomes

$$n_0 = (2\pi)^{-3}\rho_s^{-1}\int h_{\mathbf{k}}(1 - h_{\mathbf{k}})\,d^3k, \quad (107)$$

where we have replaced ρ by ρ_s, the density of superconducting electrons, since the integrations are to be carried out only over the shell of energy width $2\hbar\omega$ about the Fermi surface. Making the BCS approximation

$$(2\pi)^{-3}\int\cdots d^3k \approx N(0)\int\cdots d\epsilon \quad (108)$$

[12] The $k = 0$ contribution to (79) is

$$\frac{(2\pi)^3}{\Omega}\frac{1}{(2\pi)^3}\left(\frac{q}{n_0\rho}\right)^{\frac{1}{2}}\frac{(w\Omega)^{\frac{1}{2}}}{1 + qw\Omega}$$

and is, thus, of order $\Omega^{-\frac{3}{2}}$ and, hence, negligible (the tail in g is of order $\Omega^{-\frac{1}{2}}$).

for integration over this shell, where $N(0)$ is the density of states at the Fermi surface and ϵ the single-particle energy relative to the Fermi surface, and substituting BCS Eq. (2.35) for $h(\epsilon)$, one finds for (106)

$$2\rho_s^{-1}N(0)\int_{-\hbar\omega}^{\hbar\omega}\frac{1}{2}\left[1-\frac{\epsilon}{(\epsilon^2+\epsilon_0^2)^{\frac{1}{2}}}\right]d\epsilon = 2\rho_s^{-1}N(0)\hbar\omega = 1. \quad (109)$$

The same method of evaluation gives for (107)

$$\begin{aligned} n_0 &= \rho_s^{-1}N(0)\int_{-\hbar\omega}^{\hbar\omega}\frac{\epsilon_0^2}{4(\epsilon^2+\epsilon_0^2)}d\epsilon \\ &= \tfrac{1}{2}\rho_s^{-1}N(0)\epsilon_0\tan^{-1}\frac{\hbar\omega}{\epsilon_0} \\ &= \frac{1}{4}\frac{\epsilon_0}{\hbar\omega}\tan^{-1}\frac{\hbar\omega}{\epsilon_0}. \end{aligned} \quad (110)$$

Finally, substitution of BCS Eq. (2.40) for ϵ_0 gives

$$n_0 = \frac{\tan^{-1}\{\sinh[1/N(0)V]\}}{4\sinh[1/N(0)V]}. \quad (111)$$

In the strong-coupling limit $N(0)V \gg 1$, one has

$$n_0 \xrightarrow[N(0)V\to\infty]{} \tfrac{1}{4}, \quad (112)$$

whereas n_0 is very small in the weak-coupling limit $N(0)V \ll 1$:

$$n_0 \approx (\pi/4)\exp[-1/N(0)V]. \quad (113)$$

The reason that n_0 does not approach the upper limit $\frac{1}{2}$ in the strong-coupling limit is that g and φ_0 do not approach zero range in that limit, since they are made out of Fourier components in a narrow shell about k_F. In fact, g and φ_0 both have range $\sim k_F^{-1}$ for all values of $N(0)V$.

7. PHASE TRANSITIONS

The foregoing analysis of BCS states can be extended to nonzero temperature, with the conclusion that the superconducting transition in the BCS model is an example of Fermi condensation: $n_0 = O(1)$ (independent of n and Ω) at all temperatures T below the superconducting transition temperature T_c, but $n_0 = O(n^{-1})$ for $T > T_c$. This is a consequence of the fact that[4] ρ_2 [Eq. (28)] has a largest eigenvalue $\lambda_{\max} = O(n^{-1})$ for $T < T_c$, because of ODLRO, but for $T > T_c$ there is no ODLRO and $\lambda_{\max} = O(n^{-2})$.

Since superconductivity is a consequence of Fermi condensation, it is natural to inquire whether other types of phase transitions in many-fermion systems might be interpretable as manifestations of Fermi condensation. In fact, it has been suggested by Coleman[6] that Fermi condensation into a pair function which is a spin triplet would manifest itself as ferromagnetism. However, it has been shown[13] that a state of the form $|\psi_{\rm BCS}\rangle$ with $(\mathbf{k}\sigma, -\mathbf{k}\sigma')$ pairing in a spin-triplet state is not ferromagnetic but superconducting; the weak paramagnetism arising from the spins is swamped by the orbital diamagnetism associated with a Meissner effect and superconductivity, as in the simpler BCS states with pairing in a spin singlet. Thus, Fermi condensation into a spin-triplet pair state gives superconductivity rather than ferromagnetism. Ferromagnetism is characterized not by ODLRO of either ρ_1 or ρ_2, but by *diagonal* long-range order which manifests itself in a nonvanishing spin-density correlation function in the limit of infinite separation:

$$\lim_{|\mathbf{r}-\mathbf{r}'|\to\infty}[\langle \boldsymbol{S}(\mathbf{r})\cdot\boldsymbol{S}(\mathbf{r}')\rangle - \langle \boldsymbol{S}(\mathbf{r})\rangle\cdot\langle \boldsymbol{S}(\mathbf{r}')\rangle] \neq 0, \quad (114)$$

where

$$\begin{aligned} S_z(\mathbf{r}) &= \tfrac{1}{2}[\psi^\dagger(\mathbf{r}\uparrow)\psi(\mathbf{r}\uparrow) - \psi^\dagger(\mathbf{r}\downarrow)\psi(\mathbf{r}\downarrow)], \\ S_+(\mathbf{r}) &= S_x(\mathbf{r}) + iS_y(\mathbf{r}) = \psi^\dagger(\mathbf{r}\uparrow)\psi(\mathbf{r}\downarrow), \\ S_-(\mathbf{r}) &= S_x(\mathbf{r}) - iS_y(\mathbf{r}) = \psi^\dagger(\mathbf{r}\downarrow)\psi(\mathbf{r}\uparrow). \end{aligned} \quad (115)$$

The same is true of the "excitonic insulator" phase, which has the property that[14] $\rho_2(x_1x_2, x_1x_2)$ has a nonvanishing periodic dependence on $|\mathbf{r}_1 - \mathbf{r}_2|$ as $|\mathbf{r}_1 - \mathbf{r}_2| \to \infty$, but no ODLRO and, hence, no Fermi condensation. It appears, then, that Fermi condensation and ODLRO of ρ_2 are uniquely associated with superconductivity or, more generally, superfluidity in a many-fermion system, as proposed by Yang.[4]

8. GENERALIZATION TO l-FERMION STATES

The wavefunctions $\psi(x_1\cdots x_{ln})$ of a system of ln fermions can be expanded, by the obvious generalization[1,2] of (2), in terms of a complete orthonormal set of antisymmetric l-fermion wavefunctions $\varphi_\alpha(x_1\cdots x_l)$. Equations (6) and (7) are replaced by[2]

$$|\psi\rangle = [(l!)^n/(ln)!]^{\frac{1}{2}}\sum_{\alpha_1\cdots\alpha_n} c(\alpha_1\cdots\alpha_n)A_{\alpha_1}^\dagger\cdots A_{\alpha_n}^\dagger|0\rangle, \quad (116)$$

with

$$A_\alpha^\dagger = (l!)^{-\frac{1}{2}}\int dx_1\cdots dx_l\varphi_\alpha(x_1\cdots x_l)\psi^\dagger(x_1)\cdots\psi^\dagger(x_l). \quad (117)$$

[13] R. Balian and N. R. Werthamer, Phys. Rev. **131**, 1553 (1963) and the earlier references cited therein.

[14] D. Jerome, T. M. Rice, and W. Kohn, Phys. Rev. **158**, 462 (1967) and the earlier references cited therein.

The ideal state corresponding to (116) is given by (9), where the a_α and $a_\alpha^\dagger$ are ordinary Bose or Fermi operators depending on whether l is even or odd. By the obvious generalization of Eqs. (12)–(17), one finds for the mean occupation number n_α of the state φ_α

$$\begin{aligned} n_\alpha &= (\psi|\, a_\alpha^\dagger a_\alpha\,|\psi) \\ &= n\int \varphi_\alpha^*(x_1\cdots x_l)\rho_l(x_1\cdots x_l, x_1'\cdots x_l') \\ &\quad\times \varphi_\alpha(x_1'\cdots x_l')\,dx_1\cdots dx_l\,dx_1'\cdots dx_l', \quad (118) \end{aligned}$$

where ρ_l is the l-fermion density matrix, defined and normalized in accordance with the obvious generalizations of (16) and (19). The normalization condition (20) on the n_α remains valid. The analog of (34) is

$$\langle A_\alpha^\dagger A_\alpha\rangle = n^{-1}\binom{ln}{l} n_\alpha. \quad (119)$$

Since the integral (118) is the expectation value of ρ_l in the state φ_α, one has

$$n_\alpha \le n\lambda_{\max}, \quad (120)$$

where $\lambda_{\max}$ is the largest eigenvalue of ρ_l. It is known[15,16] that

$$\begin{aligned} &\lambda_{\max} \le O(n^{-\frac{1}{2}l}), \qquad l \text{ even}, \\ &\lambda_{\max} \le O(n^{-\frac{1}{2}(l+1)}), \quad l \text{ odd}, \end{aligned} \quad (121)$$

where l is held constant as $n\to\infty$. It is not difficult to show that, for the ground state of the ideal Fermi gas or, more generally, for any Slater determinant constructed from spatially extended orbitals and for either bound or unbound φ_α,

$$n_\alpha \le O(n^{1-l}) \text{ (normal Fermi system)}. \quad (122)$$

This order of magnitude may be taken as typical of a normal Fermi system. Hence, attainment of the bounds (121) is an exceptional occurrence for $l > 1$ and may be defined as Fermi condensation.

Yang has shown[16] that the bounds (121) are attained by taking $|\psi\rangle$ to be a BCS state of extreme type [Eq. (59)] for even l and by a minor modification of such a state for odd l, and has conjectured that the maximum kinematically attainable values of the largest eigenvalues of the ρ_l are realized by such states. Accepting this conjecture, one has[17] asymptotically for $n\to\infty$

$$\begin{aligned} &\lambda_{\max} = [l!/(\tfrac{1}{2}l)!\,(2l)^{\frac{1}{2}l}]n^{-\frac{1}{2}l}, \quad l \text{ even}, \\ &\lambda_{\max} = \{(l-1)!/[\tfrac{1}{2}(l-1)]!\,(2l)^{\frac{1}{2}(l-1)}\}n^{-\frac{1}{2}(l+1)}, \\ &\qquad\qquad l \text{ odd} \ge 3, \end{aligned} \quad (123)$$

and, hence,

$$\begin{aligned} &n_\alpha \le [l!/(\tfrac{1}{2}l)!\,(2l)^{\frac{1}{2}l}]n^{-\frac{1}{2}(l-2)}, \quad l \text{ even}, \\ &n_\alpha \le \{(l-1)!/[\tfrac{1}{2}(l-1)]!\,(2l)^{\frac{1}{2}(l-1)}\}n^{-\frac{1}{2}(l-1)}, \\ &\qquad\qquad l \text{ odd} \ge 3. \end{aligned} \quad (124)$$

It should be noted that the n dependence in (123) and, hence, (124) is *not* a matter of conjecture. For $l = 4$, the expression (124) is $O(n^{-1})$, whereas the normal-Fermi-system value (122) is only $O(n^{-3})$. There is, thus, room for a less extreme type of Fermi condensation which would be manifested by some n_α attaining a value $O(n^{-2})$, with no other n_α larger. We conjecture that such a condensation would be associated with bound states of four fermions. On the other hand, the bound (124), and more generally the order of magnitude (121), is associated with bound states of two fermions.

For $l = 5$, the expression (124) is $O(n^{-2})$ and the normal-Fermi-system value is $O(n^{-4})$, again leaving room for a condensation which would be manifested by some $n_\alpha = O(n^{-3})$. We conjecture that such a condensation would again be associated with a large number of fermions occupying the same bound state of *four* fermions; a large occupation of a bound state of five fermions would be prevented by the exclusion principle.

More generally, we define l-fermion condensation for arbitrary *even* l by

$$\begin{aligned} &0 < \lim \text{therm}\; n^{l-2}n_\alpha < \infty \\ &\quad\Leftrightarrow l\text{-fermion condensation into } \varphi_\alpha(x_1\cdots x_l), \\ &\qquad\qquad l \text{ even}. \end{aligned} \quad (125)$$

The same limit can be between 0 and ∞ for odd $l \ge 3$, but then it would be associated with large occupation of some $\varphi_\alpha(x_1\cdots x_{l-1})$ and, hence, would be an indirect consequence of (137) being satisfied for the even value $l-1$. We conjecture that (137) can occur for a given even $l \ge 4$ only if the n_α for all smaller values of l exhibit their normal Fermi values (122).

[15] J. S. Bell, Phys. Letters 2, 116 (1962).
[16] C. N. Yang, J. Math. Phys. 4, 418 (1963).

[17] Make the substitutions $l\to\frac{1}{2}l$, $N\to ln$, $M\to\infty$ in Eq. (11) of Ref. 16, and divide by N^l (our l) because of the different normalization; make the related substitutions in Eq. (12) of Ref. 16.

In view of the results of Sec. 7, it seems plausible that the maximal condensation of the type (125) would be achieved by building $|\psi\rangle$ out of a tightly bound state of l fermions, generalizing (37):

$$\psi_g(x_1 \cdots x_{ln}) = \mathcal{C}A_{ln}[g(x_1 \cdots x_l) \times g(x_{l+1} \cdots x_{2l}) \cdots g(x_{ln-l+1} \cdots x_{ln})]. \quad (126)$$

More specifically, we conjecture that the least upper bound of the limit (125) *subject to the constraint that no Fermi condensation occurs for any smaller value of* l is attained, for even l, by states of the form (126) in the limit of infinitely tight binding of g, with $\varphi_\alpha = g$. The problem of the evaluation of ρ_l for such states, even asymptotically as $n \to \infty$, is at present unsolved. One can conjecture that (49) generalizes to

$$\rho_l(x_1 \cdots x_l, x_1' \cdots x_l') = \sum_P (-1)^{p(P)}[\rho_1(x_1, x_1')\rho_1(x_2, x_2') \cdots \rho_1(x_l, x_l')] + \chi(x_1 \cdots x_l)\chi^*(x_1' \cdots x_l') \quad (127)$$

with P an arbitrary permutation of $x_1 \cdots x_l$ and $p(P)$ its parity. One, furthermore, expects that

$$\chi(x_1 \cdots x_l) = c_l n^{-\frac{1}{2}(l-1)} g(x_1 \cdots x_l) \quad (128)$$

in the limit of infinitely tightly bound g, with c_l a numerical coefficient independent of n. The least upper bound of the left side of (125), subject to the constraint that no Fermi condensation occurs for smaller l, would then be

$$\lim \text{therm } n^{l-2} n_0 = c_l^2 \quad (129)$$

with $\varphi_0 = g$, infinitely tightly bound. We have a nonrigorous argument to the effect that

$$c_l^2 = l!/l^l. \quad (130)$$

The combinatorial argument suggesting (130) breaks down for odd l; we believe that states of the form (126) do not exhibit Fermi condensation for odd l.

The conditions under which Fermi condensation will actually occur in a given system are, of course, determined by minimization of the energy (more generally, free energy). The definition (125) is only useful in cases where it is energetically favorable for formation of bound[18] states of l fermions (l even) but not for formation of bound states of pairs or, more generally, of any even number $< l$.

[18] We are using the term "bound state" here in a generalized sense to include states which, though not bound, contain a bound component, as in Sec. 7; one is tempted to call such a state a "resonance."

9. IMPLICATIONS FOR LIQUID ^{4}He

Microscopic theories of liquid ^{4}He picture the system as composed of n bosons and attempt to relate the λ transition and superfluidity of the system to a Bose condensation in a system of interacting bosons. Such an approach is very plausible, since the internal excitation energy of an ^{4}He atom is so high ($\sim 10^4$ or 10^5 °K compared to thermal energies of a few °K). Nevertheless, from a more fundamental point of view, a system of n ^{4}He atoms is made up of $2n$ electrons and n nuclei; this suggests that the maximum occupancy of single-^{4}He-atom states ought to be limited by the exclusion principle, as in the foregoing analysis.

Let $\{\varphi_\alpha(x_1 x_2 \mathbf{R})\}$ be a complete orthonormal set of single-^{4}He-atom wavefunctions, where x_j stands for the position and spin of the jth electron, and $\mathbf{R}$ for the position of the nucleus. A wavefunction of n ^{4}He atoms can be expanded by the obvious generalization of (2):

$$\psi(x_1 \cdots x_{2n} \mathbf{R}_1 \cdots \mathbf{R}_n) = \sum_{\alpha_1 \cdots \alpha_n} c(\alpha_1 \cdots \alpha_n)\varphi_{\alpha_1}(x_1 x_2 \mathbf{R}_1) \cdots \varphi_{\alpha_n}(x_{2n-1} x_{2n} \mathbf{R}_n). \quad (131)$$

The antisymmetry of ψ in $x_1 \cdots x_{2n}$ implies and is implied by symmetry of c in $\alpha_1 \cdots \alpha_n$ together with a subsidiary condition[1] analogous to (4):

$$\sum_{\alpha\beta} \sum_{p<q}^{n} (\alpha_p \alpha_q | I_{\text{elec}} | \alpha\beta) \times c(\alpha_1 \cdots \alpha_{p-1}\alpha\alpha_{p+1} \cdots \alpha_{q-1}\beta\alpha_{q+1} \cdots \alpha_n) = -\tfrac{1}{2}n(n-1)c(\alpha_1 \cdots \alpha_n), \quad (132)$$

with

$$(\alpha\beta | I_{\text{elec}} | \gamma\delta) = \int \varphi_\alpha^*(x_1 x_2 \mathbf{R})\varphi_\beta^*(x_1' x_2' \mathbf{R}')\varphi_\gamma(x_1' x_2 \mathbf{R}) \times \varphi_\delta(x_1 x_2' \mathbf{R}')\, dx_1\, dx_2\, dx_1'\, dx_2'\, d^3R\, d^3R', \quad (133)$$

whereas symmetry of ψ in $\mathbf{R}_1 \cdots \mathbf{R}_n$ necessitates the additional subsidiary condition[1]

$$\sum_{\alpha\beta} \sum_{p<q}^{n} (\alpha_p \alpha_q | I_{\text{nuc}} | \alpha\beta) \times c(\alpha_1 \cdots \alpha_{p-1}\alpha\alpha_{p+1} \cdots \alpha_{q-1}\beta\alpha_{q+1} \cdots \alpha_n) = \tfrac{1}{2}n(n-1)c(\alpha_1 \cdots \alpha_n), \quad (134)$$

with

$$(\alpha\beta | I_{\text{nuc}} | \gamma\delta) = \int \varphi_\alpha^*(x_1 x_2 \mathbf{R})\varphi_\beta^*(x_1' x_2' \mathbf{R}')\varphi_\gamma(x_1 x_2 \mathbf{R}') \times \varphi_\delta(x_1' x_2' \mathbf{R})\, dx_1\, dx_2\, dx_1'\, dx_2'\, d^3R\, d^3R'. \quad (135)$$

The state (131) has the quantized-field representation

$$|\psi\rangle = [n!\,(2n)!]^{-\frac{1}{2}}\int dx_1 \cdots dx_{2n}\, d^3R_1 \cdots d^3R_n \times \psi(x_1 \cdots x_{2n}\mathbf{R}_1 \cdots \mathbf{R}_n)\psi^\dagger(x_1)\cdots\psi^\dagger(x_{2n}) \times \psi^\dagger(\mathbf{R}_1)\cdots\psi^\dagger(\mathbf{R}_n)\,|0\rangle, \quad (136)$$

where the $\psi^\dagger(x)$ are electron creation operators satisfying Fermi anticommutation relations and the $\psi^\dagger(\mathbf{R})$ are nucleus creation operators satisfying Bose commutation relations.[19] Equations (6) and (7) are then replaced by

$$|\psi\rangle = [2^n/n!\,(2n)!]^{\frac{1}{2}}\sum_{\alpha_1\cdots\alpha_n} c(\alpha_1\cdots\alpha_n)A^\dagger_{\alpha_1}\cdots A^\dagger_{\alpha_n}|0\rangle, \quad (137)$$

with

$$A^\dagger_\alpha = 2^{-\frac{1}{2}}\int dx_1\, dx_2\, d^3R\,\varphi_\alpha(x_1x_2\mathbf{R})\psi^\dagger(x_1)\psi^\dagger(x_2)\psi^\dagger(\mathbf{R}). \quad (138)$$

The fact that the A_α and $A^\dagger_\alpha$ are not boson operators can be dealt with by the introduction of the ideal state space as before, and Eqs. (9), (10), and (11) remain valid.

Equation (17) is derived essentially as before, yielding the following formula for the mean occupation number n_α of the single-^{4}He-atom wavefunction φ_α:

$$n_\alpha = n\int \varphi^*_\alpha(x_1x_2\mathbf{R})\rho_3(x_1x_2\mathbf{R}, x_1'x_2'\mathbf{R}') \times \varphi_\alpha(x_1'x_2'\mathbf{R}')\,dx_1\,dx_2\,d^3R\,dx_1'\,dx_2'\,d^3R', \quad (139)$$

with

$$\rho_3(x_1x_2\mathbf{R}, x_1'x_2'\mathbf{R}') = \int \psi(x_1\cdots x_{2n}\mathbf{R}\mathbf{R}_2\cdots\mathbf{R}_n) \times \psi^*(x_1'x_2'x_3\cdots x_{2n}\mathbf{R}'\mathbf{R}_2\cdots\mathbf{R}_n) \times dx_3\cdots dx_{2n}\,d^3R_2\cdots d^3R_n. \quad (140)$$

ρ_3 satisfies the normalization condition

$$\int \rho_3(x_1x_2\mathbf{R}, x_1x_2\mathbf{R})\,dx_1\,dx_2\,d^3R = 1, \quad (141)$$

and as a result the occupation numbers n_α satisfy the normalization condition (20), as one expects for n ^{4}He atoms.

[19] Nothing would be gained here by going down to the next level of the hierarchy, where the nonelementary nature of the α particle would have to be faced.

One expects that the largest eigenvalue of ρ_3, hence the maximum occupation of a state of two electrons and one α particle, would be achieved by a many-particle state in which the α particle component and the electron component each separately achieves maximal condensation. Maximal condensation of the α particle component is achieved by a totally-Bose-condensed state of the α particles; for periodic boundary conditions in volume Ω, the constant $\Omega^{-\frac{1}{2}n}$ is an α state in which all the α particles occupy the spatially constant orbital $\Omega^{-\frac{1}{2}}$. As shown in Sec. 7, maximal condensation of $2n$ fermions is achieved by a BCS state (37) in the limit of infinitely tight binding of g. Thus, we form the product state

$$\psi_{\text{prod}}(x_1\cdots x_{2n}\mathbf{R}_1\cdots\mathbf{R}_n) = \Omega^{-\frac{1}{2}n}\psi_{\text{BCS}}(x_1\cdots x_{2n}), \quad (142)$$

representing maximal condensation of the α particles and the electrons separately. Since ψ_{prod} is, in fact, independent of the α positions, its density matrix (140) is

$$\rho_3(x_1x_2\mathbf{R}, x_1'x_2'\mathbf{R}') = \Omega^{-1}\rho_2^{(\text{BCS})}(x_1x_2, x_1'x_2'), \quad (143)$$

where $\rho_2^{(\text{BCS})}$ is the two-particle density matrix (16) of ψ_{BCS}. The expectation value of (143) in the state

$$\varphi_0(x_1x_2\mathbf{R}) = \Omega^{-\frac{1}{2}}g(x_1x_2) \quad (144)$$

is

$$\int \varphi_0^*(x_1x_2\mathbf{R})\rho_3(x_1x_2\mathbf{R}, x_1'x_2'\mathbf{R}') \times \varphi_0(x_1'x_2'\mathbf{R}')\,dx_1\,dx_2\,d^3R\,dx_1'\,dx_2'\,d^3R' = \int g^*(x_1x_2)\rho_2^{(\text{BCS})}(x_1x_2, x_1'x_2')g(x_1'x_2')\,dx_1\,dx_2\,dx_1'\,dx_2' < (2n-1)^{-1}, \quad (145)$$

using the inequality (21). One then has asymptotically for $n \to \infty$

$$n_0 \leq \tfrac{1}{2} \quad (146)$$

for the occupation number (139) of φ_0, with equality attained in the limit of infinitely tight binding of g. We conjecture that the right side of (145) is, in fact, the largest kinematically attainable value of the largest eigenvalue of ρ_3 for a system of n bosons and $2n$ fermions and that the right side of (146) is the largest kinematically attainable value of the occupation number of any state of one boson and two fermions in such a system. If this conjecture is correct, it precludes Bose condensation into such states, which would require $n_0 = O(n)$. This is true in spite of the fact that the boson component can undergo Bose

condensation into a one-boson state (no electrons) and does exhibit such condensation in the state (142).

A wavefunction of the form (142) cannot by any stretch of the imagination be regarded as a possible wavefunction of liquid ^{4}He; instead, it represents a plasma composed of an interpenetrating ideal Bose gas and BCS Fermi system. It could formally be expanded in terms of a complete set of ^{4}He atom wavefunctions, but practically the entire weight would come from the unbound states which are not really ^{4}He in any physically meaningful sense. In investigating occupation of single-^{4}He-atom states, the relevant question to ask is: What is the maximal occupancy of a *bound* state of one α particle and two electrons in a many-particle wavefunction in which the electrons are likewise bound pairwise to α particles? Our previous analysis suggests that such maximal occupancy would be attained by a state of the form

$$\psi_g(x_1 \cdots x_{2n}\mathbf{R}_1 \cdots \mathbf{R}_n) = \mathcal{C}A_{2n}[g(x_1x_2\mathbf{R}_1) \cdots g(x_{2n-1}x_{2n}\mathbf{R}_n)] \quad (147)$$

in the limit of infinitely tight binding of g; the values of the n_α for a realistic[20] liquid ^{4}He wavefunction would then be less than this upper bound. An argument similar to that used in deriving (127)–(130) suggests that, for the state (147),

$$\rho_3(x_1x_2\mathbf{R}, x_1'x_2'\mathbf{R}') \sim \tfrac{1}{2}n^2 g(x_1x_2\mathbf{R})g^*(x_1'x_2'\mathbf{R}'), \quad (148)$$

where the symbol $\sim$ implies asymptotic equality in the limits $n \to \infty$, range of $g \to 0$, and separation between the groups $(\mathbf{r}_1\mathbf{r}_2\mathbf{R})$ and $(\mathbf{r}_1'\mathbf{r}_2'\mathbf{R}') \to \infty$. Then, one has for such states

$$n_\alpha \leq n_0 = \tfrac{1}{2}n^{-1}, \quad (149)$$

with equality attained for $\varphi_\alpha = \varphi_0 = g(x_1x_2\mathbf{R})$ in the limit of zero range of g. In spite of its small value, the bound (149) represents Fermi condensation, since in the absence of any condensation one would have[21] $n_\alpha \leq O(n^{-2})$.

On the other hand, states of the form (147) do not exhibit Bose condensation of the α particles, in contradistinction to the plasma states (142) which have complete Bose condensation of the α particles into the orbital $\Omega^{-\frac{1}{2}}$. The occupation number $n_\mathbf{k}$ of the orbital $\varphi_\mathbf{k} = \Omega^{-\frac{1}{2}}e^{i\mathbf{k}\cdot\mathbf{R}}$ is, by (18),

$$n_\mathbf{k} = n\Omega^{-1}\int e^{-i\mathbf{k}\cdot\mathbf{R}}\rho_1(\mathbf{R}, \mathbf{R}')e^{i\mathbf{k}\cdot\mathbf{R}'}\, d^3R\, d^3R'$$
$$= n\int \rho_1(\mathbf{R}_{12})e^{-i\mathbf{k}\cdot\mathbf{R}_{12}}\, d^3R_{12}, \quad (150)$$

where ρ_1 is defined by

$$\rho_1(\mathbf{R}, \mathbf{R}') = \int \psi(x_1 \cdots x_{2n}\mathbf{R}\mathbf{R}_2 \cdots \mathbf{R}_n)\psi^*(x_1 \cdots x_{2n}\mathbf{R}'\mathbf{R}_2 \cdots \mathbf{R}_n) \times dx_1 \cdots dx_{2n}\, d^3R_2 \cdots d^3R_n \quad (151)$$

and we have made use of the fact that $\rho_1(\mathbf{R}, \mathbf{R}')$ depends only on $\mathbf{R} - \mathbf{R}'$ if ψ is an eigenstate of total linear momentum, which we assume. The same argument which gives (148) and (149) suggests that, for the state (147),

$$n_\mathbf{k} \sim n\int g(x_1x_2\mathbf{R}_{12})g^*(x_1x_20)e^{-i\mathbf{k}\cdot\mathbf{R}_{12}}\, dx_1\, dx_2\, d^3R_{12} \quad (152)$$

in the limit $n \to \infty$ and range of $g \to 0$. Now

$$g(x_1x_2\mathbf{R}) = O(\Omega^{-\frac{1}{2}}a^{-3})$$

when $\mathbf{r}_1$, $\mathbf{r}_2$, and $\mathbf{R}$ are all within a distance $\sim a$ of each other, where a is the range of g, and g falls to zero outside this range. One, therefore, concludes that

$$n_\mathbf{k} = O(\rho a^3), \quad k \lesssim a^{-1}, \quad (153)$$

whereas $n_\mathbf{k}$ falls to zero for $k \gg a^{-1}$, where $\rho = n/\Omega$. Since (153) holds for $\mathbf{k} = 0$ as well as $\mathbf{k} \neq 0$, it shows that the states (147) do not have Bose condensation of the nuclei, such condensation being prevented by the binding of the electrons to the nuclei. This is hardly surprising; even on the Bohr model, the nucleus has a momentum which fluctuates over the range a_0^{-1}, even when the total momentum of the atom is zero.

The results (148), (149), (152), and (153) have all been verified for a special choice of g, for which it is found that they are valid in the limit of zero density for fixed finite range a of g; this is equivalent to zero range for fixed density. The first density corrections have also been evaluated for the same wavefunction. The details will be reported elsewhere.

[20] Infinitely tight binding of g gives an infinite energy for ψ_g; thus, a realistic wavefunction would have even the largest n_α less than the upper bound given by (147). A more realistic wavefunction than (145) would be given by replacing g with the ground state of the He^4 atom, but this would still not be realistic with regard to translational motion of the He^4 atoms.

[21] E.g., for a product state $\psi_{\text{elec}}(x_1 \cdots x_{2n})\psi_{\text{nuc}}(\mathbf{R}_1 \cdots \mathbf{R}_n)$, one has $\rho_3(x_1x_2\mathbf{R}, x_1'x_2'\mathbf{R}') = \rho_2^{(\text{elec})}(x_1x_2, x_1'x_2')\rho_1^{(\text{nuc})}(\mathbf{R}, \mathbf{R}')$. Let $\varphi_\alpha(x_1x_2\mathbf{R}) = \varphi_\alpha^{(\text{elec})}(x_1x_2)\varphi_\alpha^{(\text{nuc})}(\mathbf{R})$. Then, by (139), (17), and (18), the occupation of φ_α is $n_\alpha = n^{-1}n_\alpha^{(\text{elec})}n_\alpha^{(\text{nuc})}$. But, $n_\alpha^{(\text{elec})} \leq O(n^{-1})$ for a normal Fermi system [Eq. (122)], and $n_\alpha^{(\text{nuc})} \leq O(1)$ for a normal Bose system (no Bose condensation). Hence, $n_\alpha \leq O(n^{-2})$ in the absence of condensation. This can also be seen in a different way in the special case that the states φ_α are taken to be plane waves. Then, each α contains three momenta, one for each particle. In order that $\sum_\alpha n_\alpha = n$, the individual n_α must then be $O(n^{-2})$, since each summation over momentum gives a factor Ω.

We conclude, then, that the λ transition in liquid ^{4}He cannot be the result of a Bose–Einstein condensation of the atoms, such a condensation being incompatible with the upper bound on the single-^{4}He-atom occupation numbers imposed by the exclusion principle acting between the electrons in different ^{4}He atoms. If the λ transition and superfluidity of liquid ^{4}He are related to a condensation into a single-^{4}He-atom state, then this must be a Fermi condensation similar in some respects to that occurring in a superconductor. The occurrence of such a condensation for the simple wavefunctions (147) makes it plausible that it could also occur in the true ground state of liquid ^{4}He. It is also worth noting that a Fermi condensation into single-^{3}He-atom states cannot occur in liquid ^{3}He, being incompatible with the upper bound on the largest eigenvalue of ρ_3 for such a system.

10. DISCUSSION

The most important result of the foregoing analysis is that Bose condensation of composite particles containing even numbers of fermions cannot occur, being replaced by a Fermi condensation associated with ODLRO and large eigenvalues of the appropriate reduced density matrices of higher than first order, of the type defined by Yang.[4] The implications of this conclusion for the microscopic theory of liquid ^{4}He are so drastic that one is strongly tempted to deny the physical relevance of the ideal-state-space operators a_α and $a_\alpha^\dagger$, in terms of which we defined occupation numbers. However, the conclusion cannot be escaped that easily. Even if one defines occupation numbers in terms of mean values of the operators $A_\alpha^\dagger A_\alpha$, where $A_\alpha^\dagger$ creates the particle in the usual sense (7) or, more generally, (117), the relation

$$\langle A_\alpha^\dagger A_\alpha\rangle = \binom{ln}{l}\int \varphi_\alpha^*(x_1\cdots x_l)\rho_l(x_1\cdots x_l, x_1'\cdots x_l') \times \varphi_\alpha(x_1'\cdots x_l')\,dx_1\cdots dx_l\,dx_1'\cdots dx_l', \quad (154)$$

together with the known upper bounds on the largest eigenvalues of the ρ_l, shows that a single $\langle A_\alpha^\dagger A_\alpha\rangle$ can never contribute a nonzero fraction (in the limit $n\to\infty$) to the sum rule

$$\sum_\alpha \langle A_\alpha^\dagger A_\alpha\rangle = \binom{ln}{l}; \quad (155)$$

instead, a single $\langle A_\alpha^\dagger A_\alpha\rangle$ is at most of order n^{-1} relative to the entire sum. On the other hand, Bose condensation would require that a single $\langle A_\alpha^\dagger A_\alpha\rangle$ contribute a nonzero fraction of the sum in the limit $n\to\infty$. This was pointed out previously by Coleman,[6] from a somewhat different point of view.

It appears, then, that the very plausible expectation that the operator terms on the right side of (8) ought to be negligible for very tightly bound wavefunctions φ_α is incompatible with the exclusion principle in cases where the group of particles involved in φ_α, though bound together, is free to move as a group throughout the system. The value of the normalization constant (155) shows that internally excited states of the atoms in, e.g., liquid ^{4}He[22] are never negligible. It was, of course, known previously that virtual internal excitations of the atoms are not completely negligible in liquid helium, inasmuch as they are responsible for the van der Waals attractive tail in the effective interatomic interaction, which is responsible for the fact that the system is a liquid rather than a gas. However, it has always been supposed that the effects of such virtual excitations could be adequately approximated by an effective interatomic potential and the system then treated as composed of structureless particles. Although this may be true regarding the effects of long-range electromagnetic interactions, we believe that the results of this paper show that it is not true regarding the effects of collisions. We believe, in fact, that the mechanism preventing Bose condensation in superconductors and liquid ^{4}He lies in the effects of collisions, acting via the exclusion principle, in causing virtual internal excitations of the atoms. This mechanism is apparent from the observation that a two-atom product function $\varphi_0(x_1x_2\mathbf{R}_1)\varphi_0(x_3x_4\mathbf{R}_2)$ can be expanded, after exchange of a pair of electrons between the atoms, in the form

$$\varphi_0(x_3x_2\mathbf{R}_1)\varphi_0(x_1x_4\mathbf{R}_2) = \sum_{\alpha\beta}(\alpha\beta|\,I_{\text{elec}}\,|00)\,\varphi_\alpha(x_1x_2\mathbf{R}_1)\varphi_\beta(x_3x_4\mathbf{R}_2) \quad (156)$$

with expansion coefficients which are elements of the exchange matrix (133); it is clear that virtual excited states are important for this expansion. This phenomenon has no analog in boson models, for which $\varphi_0(\mathbf{R}_1)\varphi_0(\mathbf{R}_2)$ has no expansion in terms of

---

[22] For n ^{4}He atoms one has

$$\langle A_\alpha^\dagger A_\alpha\rangle = [2n(2n-1)n/2!]\int \varphi_\alpha^*(x_1x_2\mathbf{R})\rho_3(x_1x_2\mathbf{R}, x_1'x_2'\mathbf{R}') \times \varphi^\alpha(x_1'x_2'\mathbf{R}')\,dx_1\,dx_2\,d^3R\,dx_1'\,dx_2'\,d^3R' = n(2n-1)n_\alpha,$$

so that (155) is replaced by

$$\sum_\alpha \langle A_\alpha^\dagger A_\alpha\rangle = n^2(2n-1),$$

which follows directly from completeness of the φ_α and normalization of ρ_3.

$\varphi_\alpha(\mathbf{R}_1)\varphi_\beta(\mathbf{R}_2)$, except the trivial one with $\alpha = \beta = 0$. This distinction lies, we believe, at the heart of the difference between Bose and Fermi condensation. We expect it to be important when exchange is important, as in liquid ^{4}He.

In conclusion, we believe that the present results show that virtual internally excited atomic states are much more important than previously supposed, being at the heart of the striking difference between Fermi condensation, which occurs in superconductors and probably in liquid ^{4}He, and Bose condensation, which cannot occur in either.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation under Grant GP-7689.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Solution of the Lee Model in All Sectors by Dynamical Algebra

DANIEL I. FIVEL*
Center for Theoretical Physics, Department of Physics and Astronomy
University of Maryland, College Park, Maryland

(Received 28 April 1969)

A potential theory canonically equivalent to the Lee model in all sectors is deduced with algebraic techniques. For the $V\theta$ sector this potential is free of disconnected graph difficulties and so is soluble in closed form. For higher sectors the problem can be reduced to Fredholm equations.

1. INTRODUCTION

It is well known that the solubility of certain potential problems (such as the harmonic oscillator and Coulomb problems) derives from the existence of dynamical symmetries. Thus the problem of finding the canonical transformation which diagonalizes the Hamiltonian may be solved with finite Lie algebras.

In this paper we shall show that an analogous result holds for simple field-theoretic models—namely, the scalar field and the Lee model. Moreover, we shall show that the solubility of the Lee model in all sectors derives from the existence of a finite dynamical Lie algebra. The method, in fact, permits us to reduce the $V\theta$ sector, the $V\theta\theta$ sector, and all higher sectors to a potential scattering problem and in this sense provides a complete solution of the Lee model.

2. THE SCALAR FIELD

To illustrate the method we shall first examine the model in which there is one static fermion N and a moving boson θ with the Hamiltonian:

$$\begin{aligned} H &= H_0(\omega) + \lambda H_1(f), \\ H_0(\omega) &= mN^\dagger N + \int \omega(k)\theta_k^\dagger\theta_k\, dk, \\ H_1(f) &= N^\dagger N(f + f^\dagger), \end{aligned} \tag{2.1}$$

with

$$\omega(k) = (k^2 + m_0^2)^{\frac{1}{2}},$$

$$f \equiv \int f(k)\theta_k\, dk,$$

$f(k)$ a given real function. We note that if $g(k)$ is any real function and

$$g \equiv \int g(k)\theta_k\, dk, \tag{2.2}$$

then

$$[g, f^\dagger] = \int g(k)f(k)\, dk \equiv (f, g) = c\text{-number}; \tag{2.3}$$

and if

$$|f| \equiv (f,f)^{\frac{1}{2}} < \infty \quad \text{and} \quad |g| < \infty,$$

then

$$|(f, g)| < \infty.$$

We shall assume that all functions $f(k), g(k), h(k), \cdots$ etc., used in the sequel satisfy

$$|f| = |g| = |h| = \cdots = 1$$

and absorb any necessary factors to accomplish this into coefficients. Thus λ in (2.1) can be adjusted so that $|f| = 1$ as long as the given $f(k)$ satisfies $|f| < \infty$.

Since $N^\dagger N$ commutes with all operators and since N is a static fermion, we are interested only in states with one N particle and so set $N^\dagger N = 1$ throughout. We thus write (2.1) with

$$\begin{aligned} H_0(\omega) &= m + \int \omega(k)\theta_k^\dagger\theta_k\, dk, \\ H_1(f) &= f + f^\dagger. \end{aligned} \tag{2.1'}$$

We seek a canonical transformation U such that

$$H' = UHU^\dagger = H_0(\omega) + W, \tag{2.4}$$

where W has potential form, i.e., the one-particle states are eigenstates of H'. Thus W must not contain operators for the process

$$N \to N + \theta \quad \text{or} \quad N + \theta \to N. \tag{2.5}$$

Suppose that such a transformation exists and has the form

$$U = e^{\mu S(g)}, \tag{2.6}$$

where μ is a real constant and

$$S(g) = N^\dagger N(g - g^\dagger) = g - g^\dagger, \tag{2.7}$$

and g is given by (2.2). Here $g(k)$ is to be determined and the constant μ then fixed so that $|g| = 1$. Thus

$$\begin{aligned} W = {} & (e^{\mu S(g)}H_0(\omega)e^{-\mu S(g)} - H_0(\omega)) \\ & + \lambda e^{\mu S(g)}H_1(f)e^{-\mu S(g)}. \end{aligned} \tag{2.8}$$

* Supported in part by the National Science Foundation under Grant GP 6036.

Consider the first term in (2.8):

$$e^{\mu S(g)}H_0(\omega)e^{-\mu S(g)} - H_0(\omega) = \int_0^\mu \frac{d}{d\beta}[e^{\beta S(g)}H_0(\omega)e^{-\beta S(g)}]\,d\beta = \int_0^\mu e^{\beta S(g)}[S(g), H_0(\omega)]e^{-\beta S(g)}\,d\beta. \quad (2.9)$$

But

$$[S(g), H_0(\omega)] = (\omega g) + (\omega g)^\dagger = H_1(\omega g) \quad (2.10)$$

with

$$(\omega g) \equiv \int \omega(k)g(k)\theta_k\,dk,$$

so that

$$W = \int_0^\mu e^{\beta S(g)}H_1(\omega g)e^{-\beta S(g)}\,d\beta + \lambda e^{\mu S(g)}H_1(f)e^{-\mu S(g)}. \quad (2.11)$$

Now since, for any operators A and B,

$$e^A B e^{-A} = \sum_{n=0}^{\infty} \frac{1}{n!}[A^{(n)}, B],$$

where $[A^{(n)}, B]$ is the iterated commutator, we see that:

The first term on the right of (2.11) belongs to the Lie algebra generated by $S(g)$ and $H_1(\omega g)$, denoted by $D(g, \omega g)$; (2.12a)

The second term on the right of (2.11) belongs to $D(g,f)$. (2.12b)

By inspection, in virtue of (2.3), we have that $D(g, \omega g)$ consists of the elements

$$1, g, \omega g, g^\dagger, (\omega g)^\dagger \quad (2.13a)$$

and linear combinations thereof; and that $D(g,f)$ consists of the elements

$$1, g, f, g^\dagger, f^\dagger \quad (2.13b)$$

and linear combinations thereof. Hence if W contains any operator mediating (2.5), it must be of the form

$$(\alpha_1 g + \alpha_2 f + \alpha_3 \omega g) + \text{H.c.}, \quad (2.14)$$

where α_1, α_2, and α_3 are constants and we note that the relations such as (2.2) between operators and associated functions are linear. Thus, if W is to have potential form, we must have

$$\alpha_1 g(k) + \alpha_2 f(k) + \alpha_3 \omega(k)g(k) = 0. \quad (2.15)$$

Thus ωg is a linear combination of f and g so that

$$D(g, \omega g) = D(g,f). \quad (2.16a)$$

The function $g(k)$ can be written

$$g(k) = \alpha_0 f(k)/[\omega(k) + \delta], \quad (2.16b)$$

where α_0 and δ are constants and so, since $(g, g) = 1$,

$$\alpha_0 = \left[\int \left(\frac{f(k)}{\omega(k) + \delta}\right)^2 dk\right]^{-\frac{1}{2}}, \quad (2.16c)$$

i.e., $g(k)$ is now fixed by the parameter δ.

Because of (2.16a) the entire calculation of (2.11) may be carried out in the algebra $D(f, g)$.

Thus, from (2.16b),

$$H_1(\omega g) = -\delta H_1(g) + \alpha_0 H_1(f) = \alpha_0 f - \delta g + \text{H.c.},$$

so we obtain

$$e^{\beta S(g)}H_1(\omega g)e^{-\beta S(g)} = \alpha_0 f - \delta g + \alpha_0\beta(g,f) - \beta\delta + \text{H.c.},$$
$$e^{\mu S(g)}H_1(f)e^{-\mu S(g)} = f + \mu(g,f) + \text{H.c.}, \quad (2.17)$$

and hence (2.11) becomes

$$W = (\mu\alpha_0 + \lambda)f - \delta\mu g + (\tfrac{1}{2}\alpha_0\mu^2 + \mu\lambda)(g,f) - \tfrac{1}{2}\mu^2\delta + \text{H.c.} \quad (2.18)$$

Thus, to exclude (2.5) processes, we must have[1]

$$(\mu\alpha_0 + \lambda)f - \delta\mu g = 0. \quad (2.19)$$

Hence from (2.16b) one concludes that one must have

$$\delta = 0, \quad (2.20a)$$

$$\alpha_0\mu + \lambda = 0. \quad (2.20b)$$

Thus, from (2.16c),

$$\alpha_0 = \left[\int \left(\frac{f(k)}{\omega(k)}\right)^2 dk\right]^{-\frac{1}{2}}, \quad (2.21a)$$

$$\mu = -\lambda\left[\int \left(\frac{f(k)}{\omega(k)}\right)^2 dk\right]^{\frac{1}{2}}, \quad (2.21b)$$

$$\mu g(k) = -\lambda f(k)/\omega(k), \quad (2.21c)$$

$$\mu S(g) = \lambda \int \frac{f(k)}{\omega(k)}(\theta_k^\dagger - \theta_k)\,dk, \quad (2.21d)$$

so that (2.6) gives the well-known "clothing operator" for the scalar field.[2] The operator W then reduces to the c-number:

$$W = 2(\tfrac{1}{2}\alpha_0\mu^2 + \mu\lambda)(g,f) = -\lambda^2 \int \frac{[f(k)]^2}{\omega(k)}\,dk = -\lambda^2 \int \frac{[f(k)]^2}{\omega(k)}\,dk \cdot N^\dagger N, \quad (2.22)$$

[1] Note here that f and g are really multiplied by $N^\dagger N$, which we have set to unity.

[2] O. W. Greenberg and S. S. Schweber, Nuovo Cimento **8**, 378 (1958).

so the mass renormalization is the familiar result

$$\delta m = -\lambda^2 \int \frac{[f(k)]^2}{\omega(k)}\, dk. \tag{2.23}$$

3. THE LEE MODEL

The simplicity of the scalar-field problem derived from the fact that the fermion fields occurred only in the combination $N^\dagger N$, which could thus be treated as a c number. This, together with the separability of the interaction $H_1(f)$ in (2.1), led to a dynamical algebra (2.13) with a small number of elements. We turn now to the Lee model. Here we have

$$\begin{aligned} H &= H_0(\omega) + \lambda H_1(f), \\ H_0(\omega) &= m(V^\dagger V + N^\dagger N) + \int \omega(k)\theta_k^\dagger \theta_k\, dk, \\ H_1(f) &= N^\dagger V f^\dagger + V^\dagger N f, \end{aligned} \tag{3.1}$$

the other quantities defined as in (2.1). (In the second of these equations we assume equal bare masses of N and V to avoid inessential complications.) Now we have the combinations $N^\dagger V$ and $V^\dagger N$, which do not commute with one another and hence the Fermi fields play an essential role in the algebraic structure. We have a compensation, however, in that $N^\dagger V$ is only multiplied by $f^\dagger$ and $V^\dagger N$ by f. This leads to the sectoring of the model, i.e., one has $V \to N + \theta \to V$, but no $N \to V + \theta \to N$ processes. Also the number of heavy particles is conserved. Again, since N and V are static fermions, we may confine ourselves to states with one and only one fermion, i.e., we set

$$V^\dagger V + N^\dagger N = 1 \tag{3.2}$$

in all calculations and also

$$(V^\dagger N)^2 = (N^\dagger V)^2 = 0. \tag{3.3}$$

Thus the operators $V^\dagger N$, $N^\dagger V$, and $V^\dagger V - N^\dagger N$ behave like the $+$, $-$, 0 Pauli matrices.

We now seek a canonical transformation

$$U = e^{\mu S(g)} \tag{3.4}$$

such that

$$H' = UHU^\dagger = H_0(\omega) + W, \tag{3.5}$$

where W is a potential, i.e., it has no direct processes

$$V \to N + \theta \quad \text{or} \quad N + \theta \to V. \tag{3.6}$$

As in Sec. 2 we make an ansatz

$$S(g) = V^\dagger N g - N^\dagger V g^\dagger, \tag{3.7}$$

where g is defined by (2.2) and $g(k)$ is to be determined along with the real constant μ such that $(g, g) = 1$. As in Sec. 2 we write W in the form (2.11), and the statements (2.12) hold as before. Now, however, when we compute the algebras $D(g, \omega g)$ and $D(g, f)$, we do not get the simple result (2.12). Rather, using (3.3), we find that $D(g, \omega g)$ consists of all operators of the form

$$R_Q^\dagger R_Q', \quad Q = 1, 2, \cdots, \tag{3.8}$$

and their linear combinations, where R_Q and R_Q' stand for any of the operators among

$$\begin{aligned} &V(g)^{Q-n-1}(\omega g)^n, \quad n = 0, 1, \cdots, Q-1, \\ &N(g)^{Q-n}(\omega g)^n, \quad n = 0, 1, \cdots, Q. \end{aligned}$$

For $D(g, f)$ we have the same basis, but with f replacing ωg.

Now these algebras are infinite because Q is allowed to be arbitrarily large. However, because the same subscript Q appears in $R_Q^\dagger$ and R_Q', which is a manifestation of the sectoring of the model, we can make the following simplification: Let P_{Q_0} be the projection operator on the vector space of all states which contain an N particle and Q_0 θ-particles or else a V particle and $Q_0 - 1$ θ-particles. This is called the Q_0 sector. It consists of all states which are linear combinations of

$$V^\dagger f_1^\dagger f_2^\dagger \cdots f_{Q_0-1}^\dagger |0\rangle \quad \text{and} \quad N^\dagger f_1^\dagger f_2^\dagger \cdots f_{Q_0}^\dagger |0\rangle \tag{3.9}$$

with $f_i(k)$ *arbitrary* functions. Now clearly for any Q the operators $R_Q^\dagger R_Q'$ map this space into itself and no other state into it. Thus

$$P_{Q_0} R_Q^\dagger R_Q' P_{Q_0} = P_{Q_0} R_Q^\dagger R_Q' = R_Q^\dagger R_Q' P_{Q_0}. \tag{3.10}$$

Hence, if A and B are any pair of operators in $D(g, \omega g)$ or any pair of operators in $D(g, f)$, we see that

$$P_{Q_0}[A, B]P_{Q_0} = [P_{Q_0} A P_{Q_0}, P_{Q_0} B P_{Q_0}]. \tag{3.11}$$

Hence, if we denote by $D_{Q_0}(g, \omega g)$ and $D_{Q_0}(g, f)$ the algebras obtained by the mapping

$$A \to A_{Q_0} \equiv P_{Q_0} A P_{Q_0}, \tag{3.12}$$

we see that (3.12) is a homomorphism, i.e., if α and β are numbers,

$$\begin{aligned} (\alpha A + \beta B)_{Q_0} &= \alpha A_{Q_0} + \beta B_{Q_0}, \\ [A, B]_{Q_0} &= [A_{Q_0}, B_{Q_0}]. \end{aligned} \tag{3.13}$$

Now we see that

$$D_{Q_0}(g, \omega g) \quad \text{and} \quad D_{Q_0}(g, f) \tag{3.14}$$

are finite algebras because

$$P_{Q_0} R_Q^\dagger R_Q' P_{Q_0} = 0 \quad \text{for} \quad Q > Q_0. \tag{3.15}$$

Now since the total Hamiltonian H in (3.1) satisfies

$$[H, P_{Q_0}] = 0 \quad \text{for all} \quad Q_0,$$

we may write

$$H = \sum_{Q_0} P_{Q_0} H P_{Q_0} = \sum_{Q_0} H_{Q_0}. \tag{3.16}$$

Thus

$$H' = H_0(\omega) + \sum_{Q_0} W_{Q_0},$$

and so, because of the property (3.13), we have the equivalent of (2.11) for each W_{Q_0}:

$$W_{Q_0} = \int_0^{\mu} e^{\beta S_{Q_0}(g)} (H_1(\omega g))_{Q_0} e^{-\beta S_{Q_0}(g)}\, d\beta + \lambda e^{\mu S_{Q_0}(g)} [H_1(f)]_{Q_0} e^{-\mu S_{Q_0}(g)}. \tag{3.17}$$

We next choose the function $g(k)$ such that there is no $V \to N + \theta$ or $N + \theta \to V$. This means that

$$W_{Q_0} \text{ for } Q_0 = 1 \text{ must be free of operators of the form } V^\dagger N f_i \text{ and } N^\dagger f_i^\dagger V. \tag{3.18}$$

Once $g(k)$ has been found, we will have to evaluate (3.17) for all Q_0. For each Q_0 this may be carried out in the algebras $D_{Q_0}(g, \omega g)$ and $D_{Q_0}(g, f)$, which, as we have seen, are finite. As expected, we will find that W_{Q_0} for $Q_0 = 2$ contains operators mediating the processes

$$V + \theta \leftrightarrow V + \theta, \quad V + \theta \leftrightarrow N + \theta + \theta, \quad N + \theta + \theta \leftrightarrow N + \theta + \theta, \tag{3.19}$$

and that W_{Q_0} for $Q_0 = 3$ contains operators mediating

$$V + \theta + \theta \leftrightarrow N + \theta + \theta + \theta, \quad \text{etc.}$$

The important point here is that the processes indicated proceed *directly* and so the W_{Q_0} for $Q_0 > 1$ is simply a multichannel potential. For $Q_0 = 2$ (the so-called $V\theta$ sector) we will see that this potential is a separable potential free of disconnected graph problems and so the solution will be obtainable in closed form. For higher sectors the potentials are still separable but because of disconnected graphs lead to Fredholm equations. That the higher sectors cannot be soluble in closed form is clear from the problem of three bodies interacting via two-body separable potentials as described by Mitra.[3] Nonetheless, the reduction to Fredholm equations may be carried out systematically, as is done in Ref. 3, or with the more sophisticated Fadeev techniques.[4] We shall consider the problem solved when $g(k)$ has been determined and a *closed form* for the operators W_{Q_0} for *all* Q_0 has been obtained. This we shall now do.

We first observe that the only operators in $D_{Q_0}(g, f)$ and $D_{Q_0}(g, \omega g)$ for $Q_0 = 1$ which mediate

$$V \to N + \theta \quad \text{or} \quad N + \theta \to V$$

are linear combinations of $V^\dagger N f$, $V^\dagger N g$, $V^\dagger N \omega g$, and their Hermitian conjugates. Thus the most general such operator is

$$V^\dagger N(\alpha_1 f + \alpha_2 g + \alpha_3 \omega g) + \text{H.c.}$$

Thus, for such an operator to be absent, we must have

$$\alpha_1 f + \alpha_2 g + \alpha_3 \omega g = 0. \tag{3.20}$$

Hence, as in the scalar theory, ωg is a linear combination of f and g. Hence the statements (2.16) apply here as well. We shall thus be able to compute W_{Q_0} in the algebra $D_{Q_0}(f, g)$. In the scalar-field case we did this directly because the algebra was so trivial. Here it is a bit more complicated and so it will be useful to find a matric representation. To do so we must first obtain an orthonormal basis of our space of functions. Let $h(k)$ be defined by

$$f(k) = \cos\theta g(k) + \sin\theta h(k), \tag{3.21}$$

where

$$\cos\theta = (f, g) = \alpha_0 \int \frac{[f(k)]^2}{\omega(k) + \delta}\, dk. \tag{3.22}$$

Thus $\cos\theta$ is determined by δ. Now we have

$$(h, h) = 1 \quad \text{and} \quad (g, h) = 0; \tag{3.23}$$

and since h is a linear combination of f and g, we see that

$$D(g, f) = D(g, h). \tag{3.24}$$

Also

$$\omega g = -\delta g + \alpha_0 f = (\alpha_0 \cos\theta - \delta) g + \alpha_0 \sin\theta h, \tag{3.25}$$

so that

$$H_1(f) = \cos\theta H_1(g) + \sin\theta H_1(h),$$
$$H_1(\omega g) = (\alpha_0 \cos\theta - \delta) H_1(g) + \alpha_0 \sin\theta H_1(h). \tag{3.26}$$

The evaluation of the terms in (3.17) is thus reducible to the computation of

$$e^{\beta S_{Q_0}(g)} [H_1(g)]_{Q_0} e^{-\beta S_{Q_0}(g)} \tag{3.27a}$$

and

$$e^{\beta S_{Q_0}(g)} [H_1(h)]_{Q_0} e^{-\beta S_{Q_0}(g)}. \tag{3.27b}$$

We take as a basis of $D(g, h)$ the operators

$$T(Q; l, n) \equiv T_Q^\dagger(l) T_Q(n), \quad Q = 0, 1, 2, \cdots, \tag{3.28a}$$

[3] A. N. Mitra, Nucl. Phys. **32**, 529 (1962).
[4] L. D. Fadeev, Zh. Eksp. Teor. Fiz. **39**, 1459 (1960) [Sov. Phys.–JETP **12**, 1014 (1961)].

where

$$T_Q(l) \equiv [(Q-l)!\,l!]^{-\frac{1}{2}} N g^l h^{Q-l}, \quad l = 0, 1, \cdots, Q,$$
$$\equiv [(Q-|l|)!\,(|l|-1)!]^{-\frac{1}{2}} V g^{|l|-1} h^{Q-|l|}, \quad l = -1, -2, \cdots, -Q, \tag{3.28b}$$

so that, by (3.23), we have

$$\langle 0|\, T_{Q_1}(l_1) T^\dagger_{Q_2}(l_2)\,|0\rangle = \delta_{Q_1 Q_2}\delta_{l_1 l_2}. \tag{3.28c}$$

If $A \in D(g, h)$, there must exist numbers $\alpha(Q, l, n)$ such that

$$A = \sum_{Q=0}^{\infty} \sum_{l_1, l_2 = -Q}^{Q} \alpha(Q; l_1, l_2) T(Q; l_1, l_2). \tag{3.29}$$

Suppose that all $\alpha(Q, l_1, l_2)$ were known for $Q = 0, 1, \cdots, Q_0 - 1$. We may then compute the coefficients $\alpha(Q_0, l_1, l_2) l_1, l_2 = -Q_0, \cdots, Q_0$ as follows: Applying (3.12) and (3.15), we see that

$$A_{Q_0} = \sum_{Q=0}^{Q_0} \sum_{l_1, l_2 = -Q}^{Q} \alpha(Q; l_1, l_2) T_{Q_0}(Q; l_1, l_2)$$
$$= \bar{A}_{Q_0} + \sum_{l_1, l_2 = -Q_0}^{Q} \alpha(Q_0; l_1, l_2) T_{Q_0}(Q_0; l_1, l_2), \tag{3.30}$$

where

$$\bar{A}_{Q_0} \equiv \sum_{Q=0}^{Q_0 - 1} \sum_{l_1, l_2 = -Q}^{Q} \alpha(Q; l_1, l_2) T_{Q_0}(Q; l_1, l_2). \tag{3.31}$$

Then

$$\langle 0|\, T_{Q_0}(l_1)(A_{Q_0} - \bar{A}_{Q_0}) T^\dagger_{Q_0}(l_2)\,|0\rangle$$
$$= \sum_{l_1', l_2' = -Q_0}^{Q} \alpha(Q_0; l_{l_1}', l_2')$$
$$\times \langle 0|\, T_{Q_0}(l_1) T^\dagger_{Q_0}(l_1') T_{Q_0}(l_2') T^\dagger_{Q_0}(l_2)\,|0\rangle$$
$$= \alpha(Q_0; l_1, l_2). \tag{3.32}$$

Hence

$$\alpha(Q_0; l_1, l_2) = \langle 0|\, T_{Q_0}(l_1) A_{Q_0} T^\dagger_{Q_0}(l_2)\,|0\rangle - \sum_{Q=0}^{Q_0 - 1} \sum_{l_1', l_2' = -Q}^{Q} \alpha(Q; l_1', l_2') \begin{Bmatrix} Q_0 & l_1 & l_2 \\ Q & l_1' & l_2' \end{Bmatrix}, \tag{3.33}$$

where

$$\begin{Bmatrix} Q_0 & l_1 & l_2 \\ Q & l_1' & l_2' \end{Bmatrix} \equiv \langle 0|\, T_{Q_0}(l_1) T^\dagger_Q(l_1') T_Q(l_2') T^\dagger_{Q_0}(l_2)\,|0\rangle. \tag{3.34}$$

These numbers can be readily computed from (3.28).

Now we are interested in the expansion (3.29) for

$$A = e^{\beta S(g)} H_1(g) e^{-\beta S(g)}, \tag{3.35a}$$
$$A = e^{\beta S(g)} H_1(h) e^{-\beta S(g)}. \tag{3.35b}$$

In both cases one sees from the expansion in iterated commutators that the operator h occurs precisely once in every term of (3.35b) and not at all in (3.35a), so that the only nonvanishing coefficients in (3.29) are

$$\alpha(Q;\ Q\ Q) \quad \text{and} \quad \alpha(Q; \pm Q, -Q) \tag{3.36a}$$

for (3.35a),

$$\alpha(Q; \pm Q, Q-1), \quad \alpha(Q; \pm Q, -Q+1),$$
$$\alpha(Q; Q-1, \pm Q), \quad \alpha(Q; -Q+1, \pm Q) \tag{3.36b}$$

for (3.35b). In these cases the choice of l_1, l_2 on the left-hand side of (3.33) uniquely determines which pair l_1', l_2' can occur for each Q on the right-hand side. To see this, notice that, since at most one h operator can occur in the T's in (3.34), then either

$$T_{Q_0}(l_1) T^\dagger_Q(l_1') \quad \text{contains only } g\text{'s}$$

or

$$T_Q(l_2') T^\dagger_{Q_0}(l_2) \quad \text{contains only } g\text{'s}.$$

Suppose the former case. Then we must have either

$$l_1 = Q_0, \quad l_1' = Q \quad \text{or} \quad l_1 = -Q_0, \quad l_1' = -Q$$

or the matrix element vanishes. Moreover, one must have

$$0 = l_2 - l_2' + l_1' - l_1 \quad \text{and} \quad \operatorname{sign} l_2 = \operatorname{sign} l_2'$$

or the matrix element vanishes. Hence the choice of l_1 and l_2 determines for each Q which pair l_1', l_2' gives a nonvanishing matrix element. The result is

$$Q - |l'| = Q_0 - |l|,$$
$$\operatorname{sign} l' = \operatorname{sign} l. \tag{3.37}$$

We then find by direct calculation that for these values

$$\begin{Bmatrix} Q_0 & l_1 & l_2 \\ Q & l_1' & l_2' \end{Bmatrix} = \frac{\nu(l_1)}{\nu(l_1')} \frac{\nu(l_2)}{\nu(l_2')} \frac{1}{(Q - Q_0)!}, \tag{3.38}$$

where

$$\nu(l) = (l!)^{\frac{1}{2}}, \quad \text{for } l \geq 0,$$
$$= [(|l|-1)!]^{\frac{1}{2}}, \quad \text{for } l < 0.$$

Hence if we put

$$\alpha_j(Q) = \alpha(Q; l_1, l_2)/\nu(l_1)\nu(l_2) \tag{3.39}$$

and

$$\zeta_j(Q) = \langle 0|\, T_Q(l_1) A_Q T^\dagger_Q(l_2)\,|0\rangle/\nu(l_1)\nu(l_2), \tag{3.40}$$

where the subscript j in (3.39) and (3.40) means that for each Q the arguments l_1 and l_2 are chosen so that

$$Q - |l_1|, \quad Q - |l_2|, \quad \operatorname{sign} l_1, \quad \text{and} \quad \operatorname{sign} l_2$$

have prescribed values, we see that (3.33) becomes

$$\sum_{Q=0}^{Q_0} \frac{\alpha_j(Q)}{(Q_0 - Q)!} = \zeta_j(Q_0). \tag{3.41}$$

This recursion relation may be solved to give

$$\alpha_j(Q_0) = \sum_{Q=0}^{Q_0} \frac{(-1)^{Q_0 - Q}}{(Q_0 - Q)!} \zeta_j(Q). \tag{3.42}$$

Hence all $\alpha(Q; l_1, l_2)$ are determined from the $\zeta_j(Q)$ in (3.40). To evaluate these for the A's in (3.35), we note that from (3.7) and (3.28b) we have

$$S(g)T_Q^\dagger(l)\,|0\rangle = (|l|)^{-\frac{1}{2}}\,T_Q^\dagger(-l)\,|0\rangle, \tag{3.43}$$

and hence

$$e^{-\beta S(g)}T_Q^\dagger(l)\,|0\rangle = \cos(\beta\,|l|^{\frac{1}{2}})\cdot T_Q^\dagger(l)\,|0\rangle - \sin(\beta l/|l|^{\frac{1}{2}})\cdot T_Q(-l)\,|0\rangle. \tag{3.44}$$

The calculation of (3.40) is thus reduced to a calculation of

$$\langle 0|\,T_Q(l_1)H_1(g)T_Q^\dagger(l_2)\,|0\rangle$$

and

$$\langle 0|\,T_Q(l_1)H_1(h)T_Q^\dagger(l_2)\,|0\rangle \quad \text{for} \quad l_1 l_2 = \pm Q, \pm(Q-1),$$

which is trivial.

With (3.40) computed, we insert the result into (3.42); then, with (3.39), we obtain the $\alpha(Q; l_1, l_2)$ for the cases (3.36). This then gives the expansion (3.29) of the quantities (3.35). With (3.26) one then has all the ingredients of (2.11), and the quadrature in β is trivial since only sines and cosines in β enter via (3.44). The resulting expansion (3.29) can be rewritten with N, V, and θ operators with (3.28); the final result is

$$\begin{aligned} W = &\sum_{Q=1}^{\infty}\{N^\dagger N[\zeta_{NN}^{(1)}(Q)g^{\dagger Q}g^Q + \zeta_{NN}^{(2)}(Q)(g^{\dagger Q-1}h^\dagger g^Q + \text{H.c.})] \\ &+ V^\dagger V[\zeta_{VV}^{(1)}(Q)g^{\dagger Q-1}g^{Q-1} + \zeta_{VV}^{(2)}(Q)(g^{\dagger Q-2}h^\dagger g^{Q-1} + \text{H.c.})]\} \\ &+ \sum_{Q=1}^{\infty}\{N^\dagger V[\zeta_{NV}^{(1)}(Q)g^{\dagger Q}g^{Q-1} + \zeta_{NV}^{(2)}(Q)g^{\dagger Q-1}h^\dagger g^{Q-1} + \zeta_{NV}^{(3)}(Q)g^{\dagger Q}g^{Q-2}h] + \text{H.c.}\}, \end{aligned} \tag{3.45}$$

with[5]

$$\zeta_{NN}^{(1)}(Q) = \frac{\cos\theta(-1)^{Q+1}}{Q!}\sum_{Q'=1}^{Q}(-1)^{Q'}\binom{Q}{Q'}(Q')^{\frac{1}{2}}\cdot[\sin(2\mu(Q')^{\frac{1}{2}}) + \tfrac{1}{2}\gamma(Q')^{\frac{1}{2}}\cdot(1-\cos(2\mu(Q')^{\frac{1}{2}}))], \tag{3.46a}$$

$$\begin{aligned}\zeta_{NN}^{(2)}(Q) = &\frac{\sin\theta(-1)^{Q+1}}{[Q!\,(Q-1)!]^{\frac{1}{2}}}\sum_{Q'=1}^{Q}(-1)^{Q'}\left[\binom{Q}{Q'}\binom{Q-1}{Q'-1}\right]^{\frac{1}{2}} \\ &\times\{\sin(\mu(Q')^{\frac{1}{2}})\cos(\mu(Q')^{\frac{1}{2}}) + \alpha_0[(Q')^{\frac{1}{2}}(1-\cos(\mu(Q')^{\frac{1}{2}}) \\ &\times\cos(\mu(Q'-1)^{\frac{1}{2}}) - (Q')^{\frac{1}{2}}\cdot\sin(\mu(Q')^{\frac{1}{2}})\sin(\mu(Q'-1)^{\frac{1}{2}})]\},\end{aligned} \tag{3.46b}$$

$$\zeta_{VV}^{(1)}(Q) = \frac{\cos\theta(-1)^{Q}}{(Q-1)!}\sum_{Q'=1}^{Q}(-1)^{Q'}\binom{Q-1}{Q'-1}(Q')^{\frac{1}{2}}[\sin(2\mu(Q')^{\frac{1}{2}}) + \tfrac{1}{2}\gamma(Q')^{\frac{1}{2}}(1-\cos(2\mu(Q')^{\frac{1}{2}}))], \tag{3.46c}$$

$$\begin{aligned}\zeta_{VV}^{(2)}(Q) = &\frac{\sin\theta(-1)^{Q}}{[(Q-1)!\,(Q-2)!]^{\frac{1}{2}}}\sum_{Q'=2}^{Q}(-1)^{Q'}\left[\binom{Q-1}{Q'-1}\binom{Q-2}{Q'-2}\right]^{\frac{1}{2}} \\ &\times\{\sin(\mu(Q'-1)^{\frac{1}{2}})\cos(\mu(Q')^{\frac{1}{2}}) + \alpha_0[(Q')^{\frac{1}{2}}\sin(\mu(Q')^{\frac{1}{2}})\cdot\sin(\mu(Q'-1))^{\frac{1}{2}} \\ &-(Q'-1)^{\frac{1}{2}}(1-\cos(\mu(Q')^{\frac{1}{2}})\cos(\mu(Q'-1)^{\frac{1}{2}}))]\},\end{aligned} \tag{3.46d}$$

$$\begin{aligned}\zeta_{NV}^{(1)}(Q) = &\frac{\cos\theta(-1)^{Q}}{[Q!\,(Q-1)!]^{\frac{1}{2}}}\sum_{Q'=1}^{Q}(-1)^{Q'}\left[\binom{Q-1}{Q'-1}\binom{Q}{Q'}\right]^{\frac{1}{2}}(Q')^{\frac{1}{2}} \\ &\times[\cos(2\mu(Q')^{\frac{1}{2}}) + \tfrac{1}{2}\gamma(Q')^{\frac{1}{2}}\sin(2\mu(Q')^{\frac{1}{2}})],\end{aligned} \tag{3.46e}$$

$$\begin{aligned}\zeta_{NV}^{(2)}(Q) = &\frac{\sin\theta(-1)^{Q}}{(Q-1)!}\sum_{Q'=1}^{Q}(-1)^{Q'}\binom{Q-1}{Q'-1}\{\cos(\mu(Q')^{\frac{1}{2}})\cos(\mu(Q'-1)^{\frac{1}{2}}) \\ &+\alpha_0[(Q')^{\frac{1}{2}}\sin(\mu(Q')^{\frac{1}{2}})\cos(\mu(Q'-1)^{\frac{1}{2}}) - (Q'-1)^{\frac{1}{2}}\cos(\mu(Q')^{\frac{1}{2}})\sin(\mu(Q'-1)^{\frac{1}{2}})]\},\end{aligned} \tag{3.46f}$$

$$\begin{aligned}\zeta_{NV}^{(3)}(Q) = &\frac{\sin\theta(-1)^{Q+1}}{[Q!\,(Q-2)!]^{\frac{1}{2}}}\sum_{Q'=2}^{Q}(-1)^{Q'}\left[\binom{Q}{Q'}\binom{Q}{Q'-2}\right]^{\frac{1}{2}} \\ &\times\{\sin(\mu(Q')^{\frac{1}{2}})\sin(\mu(Q'-1)^{\frac{1}{2}}) + \alpha_0[(Q'-1)^{\frac{1}{2}}\sin(\mu(Q')^{\frac{1}{2}}) \\ &\times\cos(\mu(Q'-1)^{\frac{1}{2}}) - (Q')^{\frac{1}{2}}\cos(\mu(Q')^{\frac{1}{2}})\sin(\mu(Q'-1)^{\frac{1}{2}})]\},\end{aligned} \tag{3.46g}$$

where

$$\binom{n}{m} \equiv \frac{n!}{(n-m)!\,m!} \quad \text{and} \quad \gamma \equiv \frac{\alpha_0\cos\theta - \delta}{\cos\theta}.$$

[5] In the following we have assumed $\lambda = 1$ in the original Hamiltonian (3.1). This produces no loss in generality as the necessary scale factor can be absorbed in m and $\omega(k)$.

Notice that $\zeta^{(2)}_{VV}$ and $\zeta^{(3)}_{NV}$ are identically zero for $Q = 1$.

We must now choose the parameters so that there are no $V \to N + \theta$ or $N \to V + \theta$ processes. Hence we require

$$\zeta^{(1)}_{NV}(1) = \zeta^{(2)}_{NV}(1) = 0. \tag{3.47}$$

From this follows

$$\tan \mu = -1/\alpha_0 \quad \text{and} \quad \cos \theta = \alpha_0 \delta. \tag{3.48}$$

Since α_0 is determined by (2.16c) and $\cos \theta$ by (3.22), we must have

$$\delta = \int \frac{[f(k)]^2}{\omega(k) + \delta}\, dk. \tag{3.49}$$

With δ determined by this equation, then α_0, μ, $\cos \theta$, and $\gamma = (\alpha_0 \cos \theta - \delta)/\cos \theta$ are determined, and so then are all of the coefficients $\zeta(Q)$ in (3.46). Moreover, the functions $g(k)$ and $h(k)$ in the operators g and h in (3.45) are determined from $f(k)$ by (2.16b) and (3.21). W is thus completely specified and is the potential canonically equivalent to the Lee model. It is, of course, a multichannel potential. The remaining coefficients for $Q = 1$ are

$$\zeta^{(1)}_{VV}(1) = -\delta, \quad \zeta^{(1)}_{NN}(1) = \delta,$$
$$\zeta^{(2)}_{NN}(1) = (\sin \theta)[(1 + \alpha_0^2)^{\frac{1}{2}} - \alpha_0]. \tag{3.50}$$

The first of these produces the V-particle mass renormalization, and the second two give rise to an effective N–θ potential. Since the N-particle is static, we have a one-body potential for the θ in the $N\theta$ sector, which is in fact separable. Moreover, since there is no θ–θ interaction, we can write the wavefunctions for the eigenstates of

$$H_0' \equiv H_0 + W_1, \tag{3.51}$$

where W_1 are the $Q = 1$ terms (3.45) in *all higher sectors* as products of the eigenfunctions of H_0' in the $N\theta$ sector. Thus one may write the Lippman–Schwinger equation in any sector using these states as a basis; thus only the remaining terms in (3.45) are in the interaction term. Now the terms W_2 [the $Q = 2$ terms in (3.45)] have no disconnected graphs in the $V\theta$ sector, and all the remaining terms annihilate the states of the $V\theta$ sector. Thus the $V\theta$ sector reduces to a separable potential problem without disconnected graphs and the Lippman–Schwinger equation reduces to the inversion of a finite matrix. For the $V\theta\theta$ sector the W_2 terms produce disconnected graphs and contain genuine two-body potentials. Thus one can only reduce the problem to a Fredholm equation by means of Mitra's technique[3] or with Fadeev methods. Nonetheless, we have a problem which in all sectors may be solved by (albeit complicated) quadratures.

4. CONCLUSION

We have seen that the canonical equivalence of the Lee model and the scalar field to effective potential theories can be anticipated by investigation of the algebra of canonical transformations that must be considered in seeking to diagonalize them. This suggests that a similar approach ought to give some insight into the Chew model and other static models. We shall pursue this question further in a later paper.

ACKNOWLEDGMENTS

I wish to thank Professors O. W. Greenberg and M. Vaughn for helpful comments. I wish further to express my appreciation to the Aspen Institute, Physics Division, whose hospitality permitted much of this research.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Relation between the $O(4)$ and $O(3, 1)$ Partial-Wave Expansions

D. A. AKYEAMPONG,* J. F. BOYCE,† AND M. A. RASHID‡
International Centre for Theoretical Physics, Trieste, Italy

(Received 28 February 1968)

The $t = 0$, $O(3, 1)$, expansion of a square-integrable equal-mass-scattering amplitude is considered. By assuming the existence of an unsubtracted dispersion relation, it is explicitly shown to continue to an $O(4)$ expansion for $(m_1 - m_2)^2 \leq s \leq (m_1 + m_2)^2$. The continuation requires the introduction of $O(3, 1)$ signature; nonsense-channel terms arise; however, their complete sum is zero.

INTRODUCTION

Recent interest in families of Regge poles with spins differing by one or two units[1] has led to further study of the higher space–time symmetry $O(3, 1)$ of the two-particle scattering amplitude which holds when t or u is zero and the masses of the initial and final particles in the corresponding crossed channels are pairwise equal.[2] The two phenomena can be connected by the fact that a single $O(3, 1)$ representation contains within it sets of $O(2, 1)$ representations,[3] viz., a pole in an amplitude of the higher symmetry is equivalent to a set of poles in amplitudes of the lower symmetry with residues which are directly related to that of the parent.[4]

Two approaches have been made. The first, due to Toller,[5] invokes the $O(3, 1)$ symmetry to make a direct partial-wave expansion of the $t = 0$ amplitude in terms of $O(3, 1)$ reduced amplitudes. This suffers from the deficiency that direct group-theoretical results apply only to functions which are square integrable over the group manifold. This requirement imposes boundedness conditions on the amplitude as $s \to \infty$ which are not satisfied by general physical amplitudes. The second approach, as in Freedman and Wang,[6] proceeds by continuing the amplitude in s to the (unphysical) region where $O(4)$ symmetry applies, making an $O(4)$ expansion followed by a Sommerfeld–Watson transformation, and then continuing this back to the physical region. It replaces the difficulty of square integrability with that of the continuation in s.

A priori, both approaches should be equivalent. We wish to test this for amplitudes which are square integrable over the $O(3, 1)$ group manifold, and which satisfy an unsubtracted $t = 0$ dispersion relation. We find that, in general, the two expansions may be continued into each other, in analogy with the $O(3)$ to $O(2, 1)$ continuation.[7] The introduction of an $O(3, 1)$ signature is forced on us since the direct and exchange components of the amplitude must be continued separately, just as for the $O(2, 1)$ to $O(3)$ continuation. The analogy is not complete, however, for although a set of "nonsense-channel" $O(3, 1)$ representations arises, destructive interference occurs between them, yielding zero total contribution.

In Sec. 1 the expansions are stated; this leads, in Secs. 2–4, to a discussion of the $O(3, 1)$ representations involving their symmetries and analyticity structure in ch ζ and σ, respectively. The proof of an integral relation between representations of the first and second kinds in Sec. 5 enables the analytic continuation between the $O(3, 1)$ and $O(4)$ expansions to be made in Sec. 6. The cancellation of the nonsense-channel contributions is demonstrated in Appendix B.

1. THE PARTIAL-WAVE EXPANSION

We consider the scattering of particles having spins S_i and masses m_i, $i = 1, \cdots, 4$, the masses being pairwise equal, $m_1 = m_3$, $m_2 = m_4$. At $t = 0$ the helicity amplitude in the center-of-mass frame of the direct channel may be expressed as

$$\langle p_3\lambda_3, p_4\lambda_4 | T | p_1\lambda_1, p_2\lambda_2\rangle = \sum_{J,J'} (S_2\lambda_2, S_4 - \lambda_4 | Jm) T_{JmJ'}(s) (J'm | S_1\lambda_1, S_3 - \lambda_3), \tag{1.1}$$

where, subject to the condition of square integrability

$$\int_1^\infty dz (z^2 - 1)^{\frac{1}{2}} |T_{JmJ'}(s)|^2 < \infty \tag{1.2}$$

with

$$z = \text{ch } \zeta = (2m_1m_2)^{-1}(s - m_1^2 - m_2^2), \tag{1.3}$$

* Permanent address: Department of Mathematics, University of Ghana, Legon, Ghana.
† Permanent address: Department of Physics, Imperial College, London, England.
‡ Permanent address: Atomic Energy Commission, Lahore, Pakistan.
[1] D. Z. Freedman and J. M. Wang, Phys. Rev. **153**, 1596 (1967).
[2] J. A. Strathdee, Abdus Salam, R. Delbourgo, and J. F. Boyce, ICTP, Trieste, Preprint IC/67/9, 1967.
[3] A. Sciarrino and M. Toller, J. Math. Phys. **8**, 1252 (1967).
[4] D. A. Akyeampong, J. F. Boyce, and M. A. Rashid, ICTP Trieste, Preprint, IC/67/61, 1967.
[5] M. Toller, University of Rome, Nota Interna Nos. 76 and 84.
[6] D. Z. Freedman and J. M. Wang, Phys. Rev. **153**, 1596 (1967).
[7] J. F. Boyce, J. Math. Phys. **8**, 675 (1967).

the $O(3,1)$ expansion

$$T_{JmJ'}(s) = \sum_{j_0} \int_{-i\infty}^{i\infty} d\sigma (j_0^2 - \sigma^2) T_{JJ'}(j_0, \sigma)\, d^{j_0\sigma}_{JmJ'}(z), \quad |j_0| < J, J', \quad (1.4)$$

may be performed.[5] Notice that the $\Delta J_3 = 0$ law which applies at $t = 0$ is automatically taken into account by the Clebsch–Gordan coefficients.

It is our object to show that upon defining reduced amplitudes having $O(3,1)$ signature χ, the above expansion becomes

$$T_{JmJ'}(s) = \sum_{\chi, j_0} \int_{M+\frac{1}{2}-i\infty}^{M+\frac{1}{2}+i\infty} d\sigma \frac{(j_0^2 - \sigma^2)}{\sin \pi(\sigma - v)} T^{\chi}_{JJ'}(j_0, \sigma) \times \tfrac{1}{2}\chi[d^{j_0\sigma}_{JmJ'}(z) - \chi e^{-i\pi(v+j_0+m)} d^{j_0\sigma}_{JmJ'}(-z)], \quad (1.5)$$

where $M = \max(J, J')$, and $v = 0$ or $\frac{1}{2}$ as j_0 is integral or half integral. It may then be continued in z to $|z| < 1$, when it becomes expressible as

$$T_{JmJ'}(s) = -2i \sum_{\chi, j_0} \sum_{\sigma=M+1}^{\infty} (j_0^2 - \sigma^2) T^{\chi}_{JJ'}(j_0, \sigma) \times \tfrac{1}{2}\chi[1 + \chi(-1)^{(\sigma - v)}]\, d^{j_0\sigma}_{JmJ'}(z), \quad (1.6)$$

which may be identified as an $O(4)$ expansion.

In order to define analytically continuable reduced amplitudes, we must utilize the dispersion relation satisfied by $T_{JmJ'}(s)$, which follows from the assumption of an unsubtracted dispersion relation for the crossed amplitude[8]

$$\langle -p_2\lambda_2, p_4\lambda_4 | T | p_1\lambda_1, -p_3\lambda_3 \rangle = (1+z)^{\frac{1}{2}|\mu+\lambda|}(1-z)^{\frac{1}{2}|\mu-\lambda|} \left(\int_{z_R}^{\infty} dz' \frac{\rho^R(z')}{z' - z - i\epsilon} + \int_{-\infty}^{-z_L} dz' \frac{\rho^L(z')}{z' - z + i\epsilon} \right), \quad (1.7)$$

where $\mu = \lambda_2 - \lambda_4$, $\lambda = \lambda_1 - \lambda_3$, and $\rho(z) = \rho(z; \lambda_2\lambda_4; \lambda_1\lambda_3)$. By making use of the $t = 0$ crossing relation, we infer that

$$T_{JmJ'}(s) = \sum_{\mu,\lambda} \Phi^{\mu\lambda}_{JmJ'}(z) \left(\int_{z_R}^{\infty} dz' \frac{\rho^R(z')}{z' - z - i\epsilon} + \int_{-\infty}^{-z_L} dz' \frac{\rho^L(z')}{z' - z + i\epsilon} \right), \quad (1.8)$$

where[9]

$$\Phi^{\mu\lambda}_{JmJ'}(z) = d^J_{m\mu}(-\tfrac{1}{2}\pi)(1+z)^{\frac{1}{2}|\mu+\lambda|} \times (1-z)^{\frac{1}{2}|\mu-\lambda|}\, d^{J'}_{\lambda m}(\tfrac{1}{2}\pi). \quad (1.9)$$

[8] F. Calogero, J. M. Charap, and E. J. Squires, Ann. Phys. (N.Y.) **25**, 325 (1963).

[9] M. Andrews and J. Gunson, J. Math. Phys. **5**, 1391 (1964).

We shall choose the cuts of $\Phi^{\mu\lambda}_{JmJ'}(z)$ to be from $z = -\infty$ to $z = \pm 1$. Hence

$$\Phi^{\mu\lambda}_{JmJ'}(z - i\epsilon) = (-1)^{(\mu-\lambda)} \Phi^{\mu\lambda}_{JmJ'}(z + i\epsilon), \quad -1 < z < 1, \quad (1.10)$$

and

$$\Phi^{\mu\lambda}_{JmJ'}(z - i\epsilon) = (-1)^{2\mu} \Phi^{\mu\lambda}_{JmJ'}(z + i\epsilon), \quad z < -1. \quad (1.11)$$

Use will also be made of the reflection properties

$$\Phi^{\mu\lambda}_{JmJ'}(-z) = (-1)^{(J'+m)} \Phi^{\mu-\lambda}_{JmJ'}(z) \quad (1.12)$$

$$= (-1)^{(J+m)} \Phi^{-\mu\lambda}_{JmJ'}(z), \quad (1.13)$$

together with the index symmetries

$$\Phi^{\mu\lambda}_{JmJ'}(z) = (-1)^{(J-J')+(\mu-\lambda)} \Phi^{\mu\lambda}_{J-mJ'}(z) \quad (1.14)$$

and

$$\Phi^{\mu\lambda}_{J'mJ}(z) = \Phi^{\lambda\mu}_{JmJ'}(z). \quad (1.15)$$

2. SYMMETRIES OF $O(3,1)$ REPRESENTATIONS

Following Toller,[5] we adopt as the basic definition the integral

$$d^{j_0\sigma}_{JmJ'}(\zeta) = \tfrac{1}{2}\{[(2J+1)(2J'+1)]\}^{\frac{1}{2}} \int_0^{\pi} d\theta \sin\theta \times \left(\frac{\exp(-\frac{1}{2}\zeta)\cos\frac{1}{2}\theta}{\cos\frac{1}{2}\theta'} \right)^{(2\sigma-1)} d^{J^*}_{j_0 m}(\theta)\, d^{J'}_{j_0 m}(\theta'), \quad (2.1)$$

where $\tan\frac{1}{2}\theta' = \exp(\zeta)\tan\frac{1}{2}\theta$ and $d^J_{j_0 m}(\theta)$ is an $O(3)$ function as defined by Andrews and Gunson.[9] From its index symmetries, we at once obtain

$$d^{j_0\sigma}_{JmJ'}(\zeta) = d^{m\sigma}_{Jj_0J'}(\zeta) = d^{-j_0\sigma}_{J-mJ'}(\zeta), \quad (2.2)$$

while the symmetry of the $O(3)$ functions under $\theta \to \pi - \theta$ yields

$$d^{j_0\sigma}_{JmJ'}(-\zeta) = d^{j_0-\sigma}_{J'mJ}(\zeta) = (-1)^{(J-J')}\, d^{-j_0\sigma}_{JmJ'}(\zeta). \quad (2.3)$$

The weak equivalence of the $O(3,1)$ representations (j_0, σ) and $(-j_0, -\sigma)$ leads to

$$d^{-j_0-\sigma}_{JmJ'}(\zeta) = \frac{\Gamma(J+\sigma+1)\Gamma(J'-\sigma+1)}{\Gamma(J-\sigma+1)\Gamma(J'+\sigma+1)} d^{j_0\sigma}_{JmJ'}(\zeta). \quad (2.4)$$

By expressing the integral of Eq. (2.1) in terms of $x = \exp(-\zeta)\cos^2\frac{1}{2}\theta(\cos^2\frac{1}{2}\theta')^{-1}$ and using the binomial theorem, the integrand may be expanded as a finite sum, each term of which is integrable,[3] yielding

$$d^{j_0\sigma}_{JmJ'}(\zeta) = e^{j_0\sigma}_{JmJ'}(\zeta) + (-1)^{(J-J')} e^{-j_0-\sigma}_{J'mJ}(\zeta), \quad (2.5)$$

with

$$e^{j_0\sigma}_{JmJ'}(\zeta) = \Delta^J_{j_0m}\Delta^{J'}_{j_0m}(2\,\mathrm{sh}\,\zeta)^{-J-J'-1}$$
$$\times \sum_{\substack{\alpha,\alpha'\\ r,n}} (-1)^{J'-m+n}\Gamma(J+J'+m-j_0-\alpha-\alpha'+1)\Gamma(j_0-m+\alpha+\alpha'+1)$$
$$\times [\Gamma(J-j_0-\alpha+1)\Gamma(J+m-\alpha+1)\Gamma(\alpha+j_0-m+1)\Gamma(\alpha+1)\Gamma(J'-j_0-\alpha'+1)$$
$$\times \Gamma(J'+m-\alpha'+1)\Gamma(\alpha'+j_0-m+1)\Gamma(\alpha'+1)\Gamma(r-\alpha+1)\Gamma(\alpha+J'-j_0-n-r+1)$$
$$\times \Gamma(J+J'+m-j_0-\alpha'-r+1)\Gamma(2j_0-m-J'+\alpha'+r+n+1)(j_0+n-\sigma)]^{-1}$$
$$\times \exp[\zeta(-\sigma+J+J'-j_0+m-2r)], \tag{2.6}$$

in which

$$\Delta^J_{j_0m} = [(2J+1)\Gamma(J+j_0-1)\Gamma(J-j_0+1)\Gamma(J+m+1)\Gamma(J-m+1)]^{\frac{1}{2}}, \tag{2.7}$$

and the summation is over those values of α, α', r, and n for which the denominator is nonsingular.

We may identify

$$e^{j_0\sigma}_{JmJ'}(\zeta) = [(2J+1)(2J'+1)]^{\frac{1}{2}}(2\,\mathrm{sh}\,\zeta)^{-1}$$
$$\times \int_{\exp(-\zeta)}^{0} dx x^{\sigma-1} d^J_{j_0m}(\theta)\, d^{J'}_{j_0m}(\theta'),$$
$$\text{for}\quad \mathrm{Re}\,(\sigma - J') > 0, \tag{2.8}$$
$$= [(2J+1)(2J'+1)]^{\frac{1}{2}}(2\,\mathrm{sh}\,\zeta)^{-1}$$
$$\times \int_{\exp(-\zeta)}^{\infty} dx x^{\sigma-1} d^J_{j_0m}(\theta)\, d^{J'}_{j_0m}(\theta'),$$
$$\text{for}\quad \mathrm{Re}\,(\sigma + J) < 0, \tag{2.9}$$

with

$$\cos\theta\,\mathrm{sh}\,\zeta = \mathrm{ch}\,\zeta - x, \quad \cos\theta'\,\mathrm{sh}\,\zeta = x^{-1} - \mathrm{ch}\,\zeta,$$
$$\sin\theta\,\mathrm{sh}\,\zeta = \{[\exp(\zeta) - x][x - \exp(-\zeta)]\}^{\frac{1}{2}},$$
$$\sin\theta'\,\mathrm{sh}\,\zeta = \{[\exp\zeta - x^{-1})(x^{-1} - \exp(-\zeta)]\}^{\frac{1}{2}}.$$

The index symmetries

$$e^{j_0\sigma}_{JmJ'}(\zeta) = e^{m\sigma}_{Jj_0J'}(\zeta) = e^{-j_0\sigma}_{J-mJ'}(\zeta) \tag{2.10}$$

follow at once, while the symmetry of the $O(3)$ functions under $\theta \to \pi - \theta$ leads to

$$e^{j_0\sigma}_{JmJ'}(-\zeta) = e^{j_0-\sigma}_{J'mJ}(\zeta), \tag{2.11}$$

and the weak equivalence of the representations to

$$e^{j_0\sigma}_{J'mJ}(\zeta) = (-1)^{(J-J')}\frac{\Gamma(J+\sigma+1)\Gamma(J'-\sigma+1)}{\Gamma(J-\sigma+1)\Gamma(J'+\sigma+1)}$$
$$\times e^{j_0\sigma}_{JmJ'}(\zeta). \tag{2.12}$$

As the summation variable r which occurs in Eq. (2.6) is always an integer, it follows that

$$e^{j_0\sigma}_{JmJ'}(\zeta + i\pi) = \exp[-i\pi(\sigma + j_0 - m + 1)]e^{j_0\sigma}_{JmJ'}(\zeta), \tag{2.13}$$

and upon combining this with Eq. (2.5) we obtain

$$d^{j_0\sigma}_{JmJ'}(\zeta + i\pi) - \exp[i\pi(\sigma + j_0 + m - 1)]\, d^{j_0\sigma}_{JmJ'}(\zeta)$$
$$= 2i\sin\pi(\sigma + j_0)\exp(i\pi m)e^{j_0\sigma}_{JmJ'}(\zeta). \tag{2.14}$$

3. ANALYTICITY STRUCTURE IN ch ζ

For physical applications we are interested in the $O(3, 1)$ representations as functions of $y = \mathrm{ch}\,\zeta$, and in particular with the analyticity structure. As a function of $y = \exp\zeta$, $e^{j_0\sigma}_{JmJ'}(\zeta)$ has poles of order $(J + J' + 1)$ at $y = \pm 1$:

$$y = z + (z^2 - 1)^{\frac{1}{2}}, \tag{3.1}$$

where we choose the square root cut to be between ± 1 and the phase to be positive on the principal sheet. As a function of z, then ζ is defined by

$$\zeta = \log[z + (z^2 - 1)^{\frac{1}{2}}], \tag{3.2}$$

with branch points at $z = \pm 1$ and cuts from these to $-\infty$. The resulting phases are shown in Fig. 1, with

$$\mathrm{ch}\,\alpha = |z|, \quad \mathrm{sh}\,\alpha = |(z^2 - 1)^{\frac{1}{2}}|,$$
$$\mathrm{ch}\,\theta = |z|, \quad \sin\theta = |(1 - z^2)^{\frac{1}{2}}|.$$

As a result, $e^{j_0\sigma}_{JmJ'}(z)$ has the same analyticity structure. Between $-\infty$ and -1 the cut is due to the logarithmic

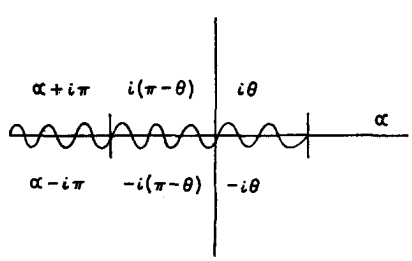


FIG. 1. Phase of $\zeta(z)$.

discontinuity, yielding

$$e^{j_0\sigma}_{JmJ'}(z - i\epsilon) = \exp[2\pi i(\sigma + j_0 - m + 1)] \times e^{j_0\sigma}_{JmJ'}(z + i\epsilon), \quad z < -1, \quad (3.3)$$

while for $-1 \leq z \leq 1$, the discontinuity is due to the square root of Eq. (3.2). Upon crossing this cut $\zeta \to -\zeta$, and hence

$$e^{j_0\sigma}_{JmJ'}(z - i\epsilon) = e^{j_0-\sigma}_{J'mJ}(z + i\epsilon), \quad -1 \leq z \leq 1, \quad (3.4)$$

from which it follows that

$$d^{j_0\sigma}_{JmJ'}(z - i\epsilon) = (-1)^{(J-J')} d^{-j_0\sigma}_{JmJ'}(z + i\epsilon), \quad -1 \leq z \leq 1; \quad (3.5)$$

it is a consequence of the choice of phases that $z \to -z$ implies $\zeta \to \zeta \mp i\pi$ as $\operatorname{Im} z \gtrless 0$. Hence

$$e^{j_0\sigma}_{JmJ'}(-z) = \exp[\pm i\pi(\sigma + j_0 - m + 1)] e^{j_0\sigma}_{JmJ'}(z), \quad \pm \text{ as } \operatorname{Im} z \gtrless 0, \quad (3.6)$$

and, upon using this in conjunction with (2.14),

$$2 \sin \pi(\sigma + j_0) e^{j_0\sigma}_{JmJ'}(z) = \pm i e^{\pm i(m-1)\pi}[e^{\mp i(\sigma+j_0+m-1)\pi} d^{j_0\sigma}_{JmJ'}(z) - d^{j_0\sigma}_{JmJ'}(-z)], \quad \pm \text{ as } \operatorname{Im} z \gtrless 0. \quad (3.7)$$

Asymptotically, as $|\sigma| \to \infty$, we have

$$e^{j_0\sigma}_{JmJ'}(z) \to K\sigma^{-1} \exp\{-\sigma \log[z + (z^2 - 1)^{\frac{1}{2}}]\}, \quad (3.8)$$

while, as $|z| \to \infty$, we have

$$e^{j_0\sigma}_{JmJ'}(z) \to \left[(2J + 1)(2J' + 1) \frac{\Gamma(J - m + 1)\Gamma(J + j_0 + 1)\Gamma(J' - m + 1)\Gamma(J' + j_0 + 1)}{\Gamma(J + m + 1)\Gamma(J - j_0 + 1)\Gamma(J' + m + 1)\Gamma(J' - j_0 + 1)}\right]^{\frac{1}{2}} \times (-1)^{(J'+m)} \frac{\Gamma(\sigma + J' + 1)\Gamma(-\sigma - j_0)}{\Gamma(j_0 - m + 1)\Gamma(\sigma - m + 1)\Gamma(J' - \sigma + 1)} (2z)^{-\sigma-|j_0-m|-1}[1 + O(z)^2], \quad (3.9)$$

and the behavior of $d^{j_0\sigma}_{JmJ'}(\zeta)$ follows at once from Eq. (2.5).

4. ANALYTICITY STRUCTURE IN σ

It may be seen from Eq. (2.6) that $e^{j_0\sigma}_{JmJ'}(z)$ is an analytic function of σ having a logarithmic branch point at $\sigma = 0$, and with possible poles at those integer values of $(\sigma - j_0)$ for which $-J \leq \sigma \leq J'$. If we choose the cut to lie along the positive imaginary axis, then, from (2.5), $d^{j_0\sigma}_{JmJ'}(z)$ is cut along the imaginary axis. From Eq. (2.1), $d^{j_0\sigma}_{JmJ'}(z)$ is finite for all σ and, consequently, from (2.4), if $(\sigma - j_0)$ is an integer,

$$d^{j_0\sigma}_{JmJ'}(z) = 0, \quad \text{if} \quad J + 1 \leq \sigma \leq J' \quad \text{or} \quad J' \leq \sigma \leq -J + 1. \quad (4.1)$$

If σ is restricted to positive values and we consider $d^{j_0\sigma}_{JmJ'}(z)$ as a matrix in (J, J'), then its form suggests its redefinition by

$$d^{j_0\sigma}_{JmJ'}(z) \to d'^{j_0\sigma}_{JmJ'}(z) = \left(\frac{\Gamma(J + \sigma + 1)\Gamma(J' - \sigma + 1)}{\Gamma(J - \sigma + 1)\Gamma(J' + \sigma + 1)}\right)^{\frac{1}{2}} d^{j_0\sigma}_{JmJ'}(z), \quad (4.2)$$

which is block diagonal. The separate pieces form the finite-dimensional unitary representation of $O(4)$ and the infinite-dimensional nonunitary representation of $O(3, 1)$ which correspond to the value of σ considered. This behavior is in complete analogy with that of $d^j_{mn}(z)$ when $j - m$ is an integer; although its diagonalization is threefold into two infinite-dimensional unitary representations $D^{(\pm)}(j)$ of $O(2, 1)$, $m, n > j$ and $m, n < -j$, respectively, and a finite-dimensional unitary representation $D(j)$ of $O(3)$, $|m|, |n| \leq j$.[9]

However, the redefinition represented by Eq. (4.2) would introduce extra cuts into the σ plane, and therefore we shall not make it. The results are, of course, independent of whether we make it or not. For values of σ at which $e^{j_0\sigma}_{JmJ'}(z)$ is finite, viz., $\sigma > J'$, $\sigma < -J$, then, from Eq. (3.7), we have

$$d^{j_0\sigma}_{JmJ'}(-z) = -e^{\mp i(\sigma+j_0+m)\pi} d^{j_0\sigma}_{JmJ'}(z), \quad \mp \text{ as } \operatorname{Im} z \gtrless 0. \quad (4.3)$$

The asymptotic behavior as $|\sigma| \to \infty$ has already been given in Eq. (3.8).

5. AN INTEGRAL RELATION

In order to continue from $O(3, 1)$ to $O(4)$, it is necessary to establish the integral relation which is analogous to that between $P_j(z)$ and $Q_j(z)$, as it is this relation which provides the link between the two expansions. Consider the function

$$F^{j_0\sigma}_{J_\mu,J'_\lambda}(z) = \sum_m \Phi^{\mu\lambda}_{JmJ'}(z) e^{j_0\sigma}_{JmJ'}(z)(z^2 - 1)^{\frac{1}{2}}, \quad (5.1)$$

where $\Phi^{\mu\lambda}_{JmJ'}(z)$ is defined by (1.9). $F^{j_0\sigma}_{J_\mu,J'_\lambda}(z)$ is analytic in the plane cut from $z = 1$ to $z = -\infty$, with branch points at $z = \pm 1$. The discontinuities across

the cuts are

$$\text{Disc}\,[F^{j_0\sigma}_{J\mu,J'\lambda}(z)] = \sum_m \Phi^{\mu\lambda}_{JmJ'}(z)\, d^{j_0\sigma}_{JmJ'}(z)(z^2-1)^{\frac{1}{2}}, \quad -1 \le z \le 1. \tag{5.2}$$

and

$$\text{Disc}\,[F^{j_0\sigma}_{J\mu,J'\lambda}(z)] = -2i \exp[i\pi(\sigma - j_0)] \sin \pi(\sigma - j_0) F^{j_0\sigma}_{J\mu,J'\lambda}(z), \quad z < -1, \tag{5.3}$$

where the functions on the right-hand side are obtained by approaching the cut from above.

It follows from Cauchy's integral theorem that if z is any point not on the real axis less than 1 and L is any contour which surrounds z and does not enclose any of the branch points, then

$$F^{j_0\sigma}_{J\mu,J'\lambda}(z) = (2\pi i)^{-1}\int_L dz' F^{j_0\sigma}_{J\mu,J'\lambda}(z')(z'-z)^{-1}. \tag{5.4}$$

It we now distort the contour L into an infinite circle plus the integral of the discontinuity along the cut, then from Eq. (3.9) it follows that the infinite circle contribution is zero for $\text{Re}\,\sigma + |j_0 - m| > \mu$ and thus, upon incorporating the results of Eqs. (4.2) and (4.3), Eq. (4.4) becomes

$$\sum_m \Big[(z^2-1)^{\frac{1}{2}}\Phi^{\lambda\mu}_{JmJ'}(z) e^{j_0\sigma}_{JmJ'}(z) - (2\pi i)^{-1}\int_{-1}^{1} dz'(z'^2-1)^{\frac{1}{2}}\Phi^{\mu\lambda}_{JmJ'}(z')\, d^{j_0\sigma}_{JmJ'}(z')(z'-z)^{-1} - \pi^{-1}\exp[i\pi(\sigma-j_0)]\sin\pi(\sigma-j_0) \times \int_{-\infty}^{-1} dz' \frac{(z^2-1)^{\frac{1}{2}}}{(z'-z)}\cdot \Phi^{\mu\lambda}_{JmJ'}(z') e^{j_0\sigma}_{JmJ'}(z')\Big] = 0, \quad \text{Re}\,\sigma > \mu. \tag{5.5}$$

This is essentially the relation for which we were looking; however, it bears somewhat closer examination, for it must be shown that the poles of $e^{j_0\sigma}_{JmJ'}(z)$ at $z = \pm 1$ do not cause the above integrals to diverge. This we perform in Appendix A.

Notice that Eq. (5.5) admits the two equivalent forms

$$\sum_m \Big((z^2-1)^{\frac{1}{2}}\Phi^{\mu\lambda}_{JmJ'}(z) e^{j_0\sigma}_{JmJ'}(z) - (2\pi i)^{-1}\int_{-1}^{1} dz'(z'^2-1)^{\frac{1}{2}}\Phi^{\mu\lambda}_{JmJ'}(z')\, d^{j_0\sigma}_{JmJ'}(z')(z'-z)^{-1} - \pi^{-1} e^{i\pi J'}\sin\pi(\sigma-j_0) \times \int_{1}^{\infty} dz'(z'^2-1)^{\frac{1}{2}}\Phi^{\mu-\lambda}_{JmJ'}(z') e^{j_0\sigma}_{JmJ'}(z')(z'+z)^{-1}\Big) = 0, \quad \text{Re}\,\sigma > \mu, \tag{5.6}$$

and

$$\sum_m \Big((z^2-1)^{\frac{1}{2}}\exp i\pi[(J'+m) \pm (\sigma + j_0 - m + 1)]\Phi^{\mu-\lambda}_{JmJ'}(z) e^{j_0\sigma}_{JmJ'}(z) - (2\pi i)^{-1}\int_{-1}^{1} dz'(z'^2-1)^{\frac{1}{2}}\Phi^{\mu\lambda}_{JmJ'}(z')\, d^{j_0\sigma}_{JmJ'}(z')(z'+z)^{-1} - \pi^{-1}e^{i\pi J'}\sin\pi(\sigma-j_0) \times \int_{1}^{\infty} dz'(z'^2-1)^{\frac{1}{2}}\Phi^{\mu-\lambda}_{JmJ'}(z') e^{j_0\sigma}_{JmJ'}(z')(z'-z)^{-1}\Big) = 0, \quad \text{Re}\,\sigma > \mu, \ \pm \text{ as } \text{Im}\, z \gtrless 0, \tag{5.7}$$

while use of Eq. (1.15) means that we are able to interchange J and J' on Φ.

6. THE $O(3,1)$ TO $O(4)$ CONTINUATION

Our object is to express the $O(3,1)$ expansion, Eq. (1.4), in such a form that it may be continued in z to $|z| < 1$ and expressed as an infinite sum over positive-integral values of $\sigma - J$, which will then be identified with an $O(4)$ expansion. In order to perform the continuation, we are forced to consider direct and exchange forces independently and thereby to introduce $O(3,1)$ signature.

The completeness relation for unitary $O(3,1)$ representations

$$\sum_m \int_0^\infty d\zeta\, \text{sh}^2\,\zeta\, d^{j_0\sigma}_{JmJ'}(\zeta)\, d^{j_0'\sigma'^*}_{JmJ'}(\zeta) = \frac{\pi(2J+1)(2J'+1)}{(j_0^2-\sigma^2)}\delta_{j_0j_0'}\delta(\sigma-\sigma') \tag{6.1}$$

enables the partial-wave amplitude to be easily obtained from (1.4):

$$T_{JJ'}(j_0,\sigma) = [\pi(2J+1)(2J'+1)]^{-1} \times \sum_m \int_0^\infty d\zeta\, \text{sh}^2\,\zeta\, d^{j_0-\sigma}_{JmJ'}(\zeta) T_{JmJ'}(s). \tag{6.2}$$

However, we wish the reduced amplitude to coincide with the $O(4)$ reduced amplitude for positive-integer values of $(\sigma - J)$. We shall use the invariance of Eq. (1.4) under $(j_0,\sigma)\to(-j_0,-\sigma)$, together with Eq. (4.5), to ensure that this is indeed the case.

We begin by dividing the amplitude into direct and exchange components:

$$T_{JmJ'}(s) = A_{JmJ'}(z) - B_{JmJ'}(z), \tag{6.3}$$

with

$$A_{JmJ'}(z) = \sum_{\mu,\lambda}\Phi^{\mu\lambda}_{JmJ'}(z)\int_{z_R}^\infty dz'\,\rho^R(z')(z'-z-i\epsilon)^{-1} \tag{6.4}$$

and

$$B_{JmJ'}(z) = \sum_{\mu,\lambda} \Phi^{\mu\lambda}_{JmJ'}(z) \int_{z_L}^{\infty} dz' \rho^L(-z')(z' + z - i\epsilon)^{-1}, \tag{6.5}$$

where the $(J\mu, J'\lambda)$ subscripts on the ρ's have been omitted. The expansions of the direct and exchange components will be considered separately; we shall show presently why this is essential.

$$B_{JmJ'}(z) = \sum_{j_0} \int_{-i\infty}^{i\infty} d\sigma (j_0^2 - \sigma^2) \mathfrak{B}_{JJ'}(j_0, \sigma)\, d^{j_0\sigma}_{JmJ'}(z), \tag{6.6}$$

where, by using the reflection invariance of the above under $(j_0, \sigma) \to (-j_0, -\sigma)$, together with (2.5),

$$\mathfrak{B}_{JJ'}(j_0, \sigma) = 2(-1)^{(J-J')}[\pi(2J + 1)(2J' + 1)]^{-1} \times \sum_m \int_1^{\infty} dz (z^2 - 1)^{\frac{1}{2}} e^{-j_0\sigma}_{J'mJ}(z) B_{JmJ'}(z) \tag{6.7}$$

or,[10] equivalently, as far as the principal-series integral is concerned,

$$\begin{aligned}\mathfrak{B}_{JJ'}(j_0, \sigma) &= 2(-1)^{(J-J')}[\pi(2J + 1)(2J' + 1) \sin \pi(\sigma - j_0)]^{-1} \\ &\times \sum_m \Big(\sin \pi(\sigma - j_0) \\ &\times \int_1^{\infty} dz (z^2 - 1)^{\frac{1}{2}} e^{-j_0\sigma}_{J'mJ}(z) B_{JmJ'}(z) \\ &- (2i)^{-1} e^{+i\pi(J+J'+m)} \\ &\times \int_{-1}^{1} dz (z^2 - 1)^{\frac{1}{2}}\, d^{-j_0\sigma}_{J'mJ}(z) B_{JmJ'}(-z) \Big).\end{aligned} \tag{6.8}$$

Due to the condition of square integrability Eq. (1.2), the above definition converges for all Re $\sigma \geq 0$. For Re $\sigma > J$, we may substitute the dispersion relation for $B_{JmJ'}(z)$, interchange the orders of integration, and then use Eq. (5.6) to identify

$$\begin{aligned}\mathfrak{B}_{JJ'}(j_0, \sigma) &= 2e^{-i\pi J'}[(2J + 1)(2J' + 1) \sin(\sigma + j_0)]^{-1} \\ &\times \sum_{\substack{m \\ \mu,\lambda}} \int_{z_L}^{\infty} dz (z^2 - 1)^{\frac{1}{2}} \Phi^{\mu,-\lambda}_{JmJ'}(z) e^{-j_0\sigma}_{J'mJ}(z) \rho^L(-z),\end{aligned} \qquad \text{Re } \sigma > J. \tag{6.9}$$

The form of the above is in close analogy with the Froissart–Gribov continuation of the normal $O(3)$ amplitude.

[10] For j_0 a nonzero integer, Eq. (6.7) has a pole at $\sigma = 0$ which yields an additional term to Eq. (6.6); it does not affect the arguments which follow and is dealt with in Appendix B.

Upon defining

$$B_{JJ'}(j_0, \sigma) = \sin \pi(\sigma - j_0) \mathfrak{B}_{JJ'}(j_0, \sigma), \tag{6.10}$$

Eq. (6.6) becomes

$$B_{JmJ'}(z) = \sum_{j_0} \int_{-i\infty}^{i\infty} d\sigma \frac{(j_0^2 - \sigma^2)}{\sin \pi(\sigma - j_0)} \times B_{JJ'}(j_0, \sigma)\, d^{j_0\sigma}_{JmJ'}(z), \tag{6.11}$$

where $B_{JJ'}(j_0, \sigma)$ is finite for all Re $\sigma \geq 0$. The asymptotic behavior of $e^{j_0\sigma}_{JmJ'}(z)$, Eq. (3.8), enables us to continue in z to $|z| < 1$ and then perform an inverse Sommerfeld–Watson transformation, yielding

$$B_{JmJ'}(z) = 2i \sum_{j_0} \sum_{\sigma=1-v}^{\infty} (j_0^2 - \sigma^2) B_{JJ'}(j_0, \sigma) \times e^{i\pi(\sigma - j_0)}\, d^{j_0\sigma}_{JmJ'}(z), \tag{6.12}$$

where $v = 0$ or $\frac{1}{2}$ as j_0 is integral or half-integral. But since $d^{j_0\sigma}_{JmJ'}(z)$ is zero for $J + 1 \leq \sigma < J'$, it follows that the sum over σ splits into two parts, since either $B_{JJ'}(j_0, \sigma)$ or $d^{j_0\sigma}_{JmJ'}(z)$ is zero for min $(J, J') + 1 \leq \sigma \leq$ max (J, J'). This is a consequence of the block diagonalization already referred to in Sec. 4. The sum max $(J, J') + 1 \leq \sigma < \infty$ is over unitary representations of $O(4)$, while the nonsense-channel terms $1 - v \leq \sigma \leq$ min (J, J') are components of the infinite-dimensional nonunitary representations.

For $\sigma - 1 \geq (J, J')$, Eq. (6.9) yields

$$\begin{aligned}B_{JJ'}(j_0, \sigma) &= \frac{i(-1)^{(\sigma - j_0)}}{\pi(2J + 1)(2J' + 1)} \\ &\times \frac{\Gamma(J + \sigma + 1)\Gamma(J' - \sigma + 1)}{\Gamma(J - \sigma + 1)\Gamma(J' + \sigma + 1)} \\ &\times \sum_m \int_{-1}^{1} dz (z^2 - 1)^{\frac{1}{2}}\, d^{-j_0\sigma}_{JmJ'}(z) B_{JmJ'}(z) \\ &= \frac{i(-1)^{(\sigma - j_0)}}{\pi(2J + 1)(2J' + 1)} \\ &\times (-1)^{(J-J')} \frac{\Gamma(J + \sigma + 1)\Gamma(J' - \sigma + 1)}{\Gamma(J - \sigma + 1)\Gamma(J' + \sigma + 1)} \\ &\times \sum_m \int_{-1}^{1} dz (z^2 - 1)^{\frac{1}{2}}\, d^{j_0\sigma^*}_{JmJ'}(z) B_{JmJ'}(z),\end{aligned} \tag{6.13}$$

which may be identified as an $O(4)$ partial-wave amplitude apart from the factor

$$(-1)^{(J-J')}\Gamma(J + \sigma + 1)\Gamma(J' - \sigma + 1) \times [\Gamma(J - \sigma + 1)\Gamma(J' + \sigma + 1)]^{-1}.$$

At the nonsense-channel values $\sigma < (J, J')$, we have

$$B_{JJ'}(j_0, \sigma) = i(-1)^{(J-J')}[\pi(2J+1)(2J'+1)]^{-1} \times \sum_m \left(\int_1^\infty dz(z^2-1)^{\frac{1}{2}}[(-1)^{(\sigma-j_0)} d^{-j_0\sigma}_{J'mJ}(z) + e^{im\pi} d^{-j_0\sigma}_{J'mJ}(-z)] + e^{-i\pi(J-J'-m)} \int_{-1}^1 dz(z^2-1)^{\frac{1}{2}} d^{-j_0\sigma}_{J'mJ}(z) B_{JmJ'}(-z) \right). \quad (6.14)$$

Consider now the expansion of the direct force amplitude

$$A_{JmJ'}(z) = \sum_{j_0} \int_{-i\infty}^{i\infty} d\sigma (j_0^2 - \sigma^2) \mathcal{A}_{JJ'}(j_0, \sigma)\, d^{j_0\sigma}_{JmJ'}(z), \quad (6.15)$$

with

$$\mathcal{A}_{JJ'}(j_0, \sigma) = 2(-1)^{(J-J')}[\pi(2J+1)(2J'+1)]^{-1} \times \sum_m \int_1^\infty dz(z^2-1)^{\frac{1}{2}} e^{-j_0\sigma}_{J'mJ}(z) A_{JmJ'}(z), \quad (6.16)$$

where we have again made use of the invariance of the principal series under $(j_0, \sigma) \to (-j_0, -\sigma)$, as in Eq. (6.7). We utilize it further to redefine

$$\mathcal{A}_{JJ'}(j_0, \sigma) = 2(-1)^{(J-J')}[\pi(2J+1)(2J'+1)\sin\pi(\sigma - j_0)]^{-1} \times \sum_m \left(\sin\pi(\sigma - j_0) \int_1^\infty dz(z^2-1)^{\frac{1}{2}} e^{-j_0\sigma}_{J'mJ}(z) A_{JmJ'}(z) + \tfrac{1}{2} i e^{i\pi(J+J'+m)} \int_{-1}^1 dz(z^2-1)^{\frac{1}{2}} d^{-j_0\sigma}_{J'mJ}(z) A_{JmJ'}(-z) \right). \quad (6.17)$$

As was the case with $\mathcal{B}_{JJ'}(j_0, \sigma)$, this definition converges for all $\operatorname{Re}\sigma \geq 0$. For $\operatorname{Re}\sigma > J$, we may substitute the dispersion relation for $A_{JmJ'}(z)$, interchange the orders of integration, and then use Eq. (5.6) to identify

$$\mathcal{A}_{JJ'}(j_0, c) = -2ie^{i\pi(\sigma+j_0+J-J'+1)} \times [(2J+1)(2J'+1)\sin\pi(\sigma - j_0)]^{-1} \times \sum_{\substack{m \\ \mu,\lambda}} \int_{z_R}^\infty dz(z^2-1)^{\frac{1}{2}} \Phi^{\mu\lambda}_{JmJ'}(z) e^{-j_0\sigma}_{J'mJ}(z) \rho^R(z), \quad \operatorname{Re}\sigma > J, \quad (6.18)$$

$$= e^{i\pi(\sigma+j_0-1)}[\sin\pi(\sigma - j_0)]^{-1} A_{JJ'}(j_0, \sigma). \quad (6.19)$$

The factor $e^{i\pi\sigma}$, which occurs in Eq. (6.18) and which did not occur in Eq. (6.9), prevents the continuation of Eq. (6.15) in z to $|z| < 1$. It is possible to absorb its effect by changing $d^{j_0\sigma}_{JmJ'}(z)$ to $d^{j_0\sigma}_{JmJ'}(-z)$ before continuing in z; we are thereby forced to consider direct forces independently from exchange forces and consequently to introduce signature.

Before introducing signature, however, we shall shift the contour of integration of Eq. (6.15) from $\operatorname{Re}\sigma = 0$ to $\operatorname{Re}\sigma = M + \frac{1}{2}$, where $M = \max(J, J')$. As the contributions from $\operatorname{Im}\sigma = \pm\infty$ are zero, this yields

$$A_{JmJ'}(z) = \sum_{j_0} \int_{M+\frac{1}{2}-i\infty}^{M+\frac{1}{2}+i\infty} d\sigma \frac{(j_0^2 - \sigma^2)}{\sin\pi(\sigma - j_0)} \times e^{i\pi(\sigma+j_0-1)} A_{JJ'}(j_0, \sigma)\, d^{j_0\sigma}_{JmJ'}(z) + 2i \sum_{\sigma=1-v}^{\min(J,J')} (j_0^2 - \sigma^2)(-1)^{2j_0} A_{JJ'}(j_0, \sigma)\, d^{j_0\sigma}_{JmJ'}(z), \quad (6.20)$$

where the nonsense-channel amplitudes are given by

$$A_{JJ'}(j_0, \sigma) = (-1)^{(\sigma+j_0+J-J')}[i\pi(2J+1)(2J'+1)]^{-1} \times \sum_m \left(\int_1^\infty dz(z^2-1)^{\frac{1}{2}}[(-1)^{(\sigma-j_0)} d^{-j_0\sigma}_{J'mJ}(z) + e^{im\pi} d^{-j_0\sigma}_{J'mJ}(-z)] A_{J'mJ}(z) + e^{-i\pi m} \int_{-1}^1 dz(z^2-1)^{\frac{1}{2}} d^{-j_0\sigma}_{J'mJ}(z) A_{JmJ'}(-z) \right)_{\sigma \leq M} \quad (6.21)$$

The signature factor $e^{i\pi\sigma}$ is now eliminated from Eq. (6.20) by using

$$e^{i\pi(\sigma+j_0)\pi} d^{j_0\sigma}_{JmJ'}(z) = e^{-i(m-1)\pi} d^{j_0\sigma}_{JmJ'}(-z) + 2i\sin\pi(\sigma + j_0) e^{j_0\sigma}_{JmJ'}(z), \quad (3.7)$$

together with

$$\sum_{j_0} \int_{M+\frac{1}{2}-i\infty}^{M+\frac{1}{2}+i\infty} d\sigma (j_0^2 - \sigma^2) A_{JJ'}(j_0, \sigma) e^{j_0\sigma}_{JmJ'}(z) = 0, \quad (6.22)$$

where the latter relation follows by closing the contour in the right-hand σ plane. Hence the "sense component" of Eq. (6.20) becomes

$$A_{JmJ'}(z) = \sum_{j_0} \int_{M+\frac{1}{2}-i\infty}^{M+\frac{1}{2}+i\infty} d\sigma \frac{(j_0^2 - \sigma^2)}{\sin\pi(\sigma - j_0)} A_{JJ'}(j_0, \sigma) \times e^{-im\pi} d^{j_0\sigma}_{JmJ'}(-z). \quad (6.23)$$

This expression may be continued in z to $|z| < 1$ and an inverse Sommerfeld–Watson transformation

then performed, yielding

$$A_{Jm J'}(z) = -2i \sum_{j_0} \sum_{\sigma=M+1}^{\infty} (j_0^2 - \sigma^2) A_{JJ'}(j_0, \sigma) \times e^{-i\pi(\sigma - j_0 + m)} d^{j_0\sigma}_{JmJ'}(-z) \quad (6.24)$$

with, from (6.18),

$$A_{JJ'}(j_0, \sigma) = [i\pi(2J+1)(2J'+1)]^{-1}(-1)^{(J-J')} \times \frac{\Gamma(J+\sigma+1)\Gamma(J'-\sigma+1)}{\Gamma(J-\sigma+1)\Gamma(J'+\sigma+1)} \times \sum_m \int_{-1}^{1} dz (z^2-1)^{\frac{1}{2}} d^{j_0\sigma}_{JmJ'}(z) A_{JmJ'}(z), \quad (6.25)$$

which, analogously to Eq. (6.13), may be identified with an $O(4)$ partial-wave amplitude, apart from a normalization factor.

If we neglect the nonsense-channel terms, then Eqs. (6.11) and (6.23) enable us to express the amplitude *in toto* as

$$T_{JmJ'}(z) = \sum_{j_0} \int_{M+\frac{1}{2}-i\infty}^{M+\frac{1}{2}+i\infty} d\sigma \frac{(j_0^2 - \sigma^2)}{\sin \pi(\sigma - j_0)} \times [A_{JJ'}(j_0, \sigma) e^{-i\pi m} d^{j_0\sigma}_{JmJ'}(-z) - B_{JJ'}(j_0, \sigma) d^{j_0\sigma}_{JmJ'}(z)], \quad M = \max(J, J'), \quad (6.26)$$

where the partial-wave amplitudes are defined by (6.9) and (6.18), respectively. Upon defining $O(3,1)$ signature eigenamplitudes by

$$(-1)^{(v-j_0)} T^{\chi}_{JJ'}(j_0, \sigma) = (-1)^{(v+j_0+1)} A_{JJ'}(j_0, \sigma) - \chi B_{JJ'}(j_0, \sigma), \quad (6.27)$$

where $v = 0$ or $\frac{1}{2}$ as j_0 is integral or half-integral, the expansion becomes

$$T_{JmJ'}(z) = \sum_{j_0} \sum_{\chi} \int_{M+\frac{1}{2}-i\infty}^{M+\frac{1}{2}+i\infty} d\sigma \frac{(j_0^2 - \sigma^2)}{\sin \pi(\sigma - v)} T^{\chi}_{JJ'}(j_0, \sigma) \times \tfrac{1}{2}\chi[d^{j_0\sigma}_{JmJ'}(z) - \chi e^{-i\pi(v+j_0+m)} d^{j_0\sigma}_{JmJ'}(-z)], \quad (6.28)$$

which, upon continuing to $|z| < 1$ and making an inverse Sommerfeld–Watson transformation, becomes

$$T_{JmJ'}(z) = -2i \sum_{j_0,\chi} \sum_{\sigma=J+1} (j_0^2 - \sigma^2) \times \tfrac{1}{2}[1 + \chi(-1)^{(\sigma - v)}] d^{j_0\sigma}_{JmJ'}(z) \quad (6.29)$$

and is then identifiable as an $O(4)$ expansion.

Incorporating the result of Appendix B, that the complete set of nonsense-channel terms sums to zero, we infer that the $O(3,1)$ expansion of the amplitude for $\operatorname{ch}\zeta > 1$, i.e., $s > (m_1 + m_2)^2$, has been continued into an $O(4)$ expansion for $|\operatorname{ch}\zeta| < 1$, i.e., $(m_1 - m_2)^2 \leq s \leq (m_1 + m_2)^2$. For amplitudes of the above class, *viz.*, square integrable over the $O(3,1)$ group-manifold and satisfying an unsubtracted dispersion relation, the two expansions are therefore equivalent.

7. CONCLUSION

For pairwise equal mass-scattering amplitudes which satisfy the condition of square-integrability expressed by Eq. (2.1) and which obey the unsubtracted dispersion relation of Eq. (4.7), the $O(3,1)$ expansion which holds at t or $u = 0$ and $s > (m_1 + m_2)^2$ may be continued into the $O(4)$ expansion which holds for $(m_1 - m_2)^2 \leq s \leq (m_1 + m_2)^2$. In order to make this continuation between the two expansions, it is necessary to consider direct and exchange forces independently, thereby introducing $O(3,1)$ signature. In performing the continuation, "nonsense channel," terms analogous to those of $O(2,1)$ arise; however, they destructively interfere, as the sum of the complete set is zero. This might be anticipated, as the representations to which they correspond are neither unitary nor square integrable over $O(3,1)$ or $O(4)$.

The $O(3,1)$ expansion in terms of signature eigenamplitudes $T^{\chi}_{JJ'}(j_0, \sigma)$ is

$$T_{JmJ'}(z) = \sum_{\chi, j_0} \int_{M+\frac{1}{2}-i\infty}^{M+\frac{1}{2}+i\infty} d\sigma \frac{(j_0^2 - \sigma^2)}{\sin \pi(\sigma - v)} T^{\chi}_{JJ'}(j_0, \sigma) \times \tfrac{1}{2}\chi[d^{j_0\sigma}_{JmJ'}(z) - \chi e^{-i\pi(v+j_0+m)} d^{j_0\sigma}_{JmJ'}(-z)],$$

where $z = \operatorname{ch}\zeta$, $M = \max(J, J')$, $|j_0| \leq J, J'$, and $v = 0$ or $\frac{1}{2}$ as j_0 is integral or half integral.

ACKNOWLEDGMENTS

The authors wish to thank Professor Abdus Salam for suggesting this problem and for his interest throughout its development. They also wish to acknowledge much assistance from Dr. J. A. Strathdee and many discussions with Dr. R. Delbourgo. One of the authors (J. F. B.) wishes to thank Dr. A. Sciarrino and Dr. K. Koller for illuminating conversations. They wish to thank Professor Abdus Salam, Professor P. Budini, and the IAEA for hospitality at the International Centre for Theoretical Physics, Trieste.

APPENDIX A

Owing to the poles of $e^{j_0\sigma}_{JmJ'}(z)$ of order $J + J' + 1$ at $z = \pm 1$ and of $d^{j_0\sigma}_{JmJ'}(z)$ at $z = +1$, it might at first sight appear that the integrals within Eq. (5.5) diverge; however, as a result of the zeros of $\Phi^{\mu\lambda}_{JmJ'}(z)$ at $z = \pm 1$, the complete expressions are finite. To

show this, we utilize the $O(3,1) \to O(2,1)$ decomposition made by Sciarrino and Toller[3]:

$$d^{j_0\sigma}_{JmJ'}(\zeta) = -i\sum_{n,n'}\int_{-\frac{1}{2}-i\infty}^{-\frac{1}{2}+i\infty} dj\,\frac{(2j+1)}{\tan\pi(j-v)}\,K_n^{j_0-\sigma}(j,+,J) \times K_{n'}^{j_0\sigma}(j,+,J')\,d^J_{mn}(-\tfrac{1}{2}\pi)\,d^j_{nn'}(\zeta)\,d^{J'}_{n'm}(\tfrac{1}{2}\pi) + \text{discrete series.} \quad \text{(A1)}$$

The above expansion decomposes a principal series representation of $O(3,1)$ onto the principal and discrete series of unitary representations of $O(2,1)$. In order to obtain the decomposition of a nonunitary representation, we must continue in C and simultaneously distort the contour of integration to avoid the poles of $K_n^{j_0\sigma}(j,+,J)$ at $j=-\sigma+N, j=\sigma-N-1$, $(N=0,1,\cdots)$.

Now, from its definition,[9] we may show that, as $|\operatorname{Im} j| \to \infty$, we have

$$|e^j_{mn}(z)| < \text{const}\,|j|^{-\frac{1}{2}}\exp(\pi\,|\operatorname{Im} j|), \quad \text{(A2)}$$

while

$$|K_m^{j_0\sigma}(j,+,J)| < \text{const}\,|j|^a\exp(-\pi\,|\operatorname{Im} j|), \quad \text{(A3)}$$

where a is a finite real number whose value depends on the maximum value attained by the variable on the defining integral of K. Hence the integral appearing in Eq. (A1) is uniformly convergent, and consequently we have

$$\sum_m \Phi^{\mu\lambda}_{JmJ'}(z)\,d^{j_0\sigma}_{JmJ'}(z) = -i\int_{-\frac{1}{2}-i\infty}^{-\frac{1}{2}+i\infty} dj\,\frac{(2j+1)}{\tan\pi(j-v)}\,K_\mu^{j_0-\sigma}(j,+,J) \times K_\lambda^{j_0\sigma}(j,+,J')\,d^j_{\mu\lambda}(z)(1+z)^{\frac{1}{2}|\mu+\lambda|}(1-z)^{\frac{1}{2}|\mu-\lambda|} + \text{discrete series.} \quad \text{(A4)}$$

It now follows directly from the definition of $d^j_{mn}(z)$, together with H.T.F.[11] (2.10.14), that the right-hand side is finite at $z=\pm 1$.

APPENDIX B

If we take explicit account of the pole of $\mathfrak{B}_{JJ'}(j_0,\sigma)$, then, for j_0 integral, Eq. (6.6) becomes

$$B_{JmJ'}(z) = \sum_{j_0}\int_{-i\infty}^{i\infty} d\sigma(j_0^2-\sigma^2)\mathfrak{B}_{JJ'}(j_0,\sigma)\,d^{j_0\sigma}_{JmJ'}(z) - i\sum_{j_0} j_0^2(-1)^{j_0}B_{JJ'}(j_0,0)\,d^{j_00}_{JmJ'}(z). \quad \text{(B1)}$$

[11] *Higher Transcendental Functions*, A. Erdélyi, Ed. (McGraw-Hill Book Co., New York, 1953), Vol. I.

Since

$$\lim_{\sigma\to 0}\{\sigma\mathfrak{B}_{JJ'}(j_0,\sigma)\} = \pi^{-1}(-1)^{j_0}B_{JJ'}(j_0,0), \quad \text{(B2)}$$

where $B_{JJ'}(j_0,\sigma)$ is defined by Eq. (6.14); it follows that the complete set of nonsense-channel terms are

$$-2i\sum_{j_0}\sum_{\sigma=1-v}^{N}(j_0^2-\sigma^2)B_{JJ'}(j_0,\sigma)(-1)^{(\sigma-j_0)}\,d^{j_0\sigma}_{JmJ'}(z) - i\sum_{j_0} j_0^2B_{JJ'}(j_0,0)(-1)^{j_0}\,d^{j_00}_{JmJ'}(z), \quad \text{(B3)}$$

where $N=\min(J,J')$ and the second term occurs only for integral j_0.

From the definition of $B_{JJ'}(j_0,\sigma)$ given in Eq. (6.14), if we can show by using the discrete symmetries that

$$\Psi = 2\sum_{j_0}\sum_{\sigma=1-v}^{N}(j_0^2-\sigma^2)(-1)^{(\sigma-j_0)}\,d^{j_0\sigma}_{JmJ'}(z)\,d^{-j_0\sigma}_{J'm'J}(z') + \sum_{j_0} j_0^2(-1)^{j_0}\,d^{j_00}_{JmJ'}(z)\,d^{-j_00}_{J'm'J}(z') \quad \text{(B4)}$$

is zero, then the sum of the nonsense-channel contributions will be zero. Now since

$$d^{j_0\sigma}_{JmJ'}(z)\,d^{-j_0\sigma}_{J'm'J}(z') = d^{-j_0-\sigma}_{JmJ'}(z)\,d^{j_0-\sigma}_{J'm'J}(z'), \quad \text{(B5)}$$

$$\therefore\quad \Psi = \sum_{j_0,\sigma}(j_0^2-\sigma^2)(-1)^{(\sigma-j_0)}\,d^{j_0\sigma}_{JmJ'}(z)\,d^{-j_0\sigma}_{J'm'J}(z'), \quad |j_0|,|\sigma| \le J,J'; \quad \text{(B6)}$$

while, for $\sigma-j_0$ an integer, $\sigma \le J,J'$,[12]

$$d^{j_0\sigma}_{JmJ'}(z) = \left(\frac{\Gamma(J-\sigma+1)\Gamma(J+j_0+1)\Gamma(J'+\sigma+1)\Gamma(J'-j_0+1)}{\Gamma(J+\sigma+1)\Gamma(J-j_0+1)\Gamma(J'-\sigma+1)\Gamma(J'+j_0+1)}\right)^{\frac{1}{2}} \times d^{\sigma j_0}_{JmJ'}(z). \quad \text{(B7)}$$

Hence

$$d^{j_0\sigma}_{JmJ'}(z)\,d^{-j_0\sigma}_{J'm'J}(z') = \frac{\Gamma(J+j_0+1)\Gamma(J'-j_0+1)}{\Gamma(J-j_0+1)\Gamma(J'+j_0+1)}\,d^{\sigma j_0}_{JmJ'}(z)\,d^{\sigma-j_0}_{J'm'J}(z') = d^{\sigma j_0}_{JmJ'}(z)\,d^{-\sigma j_0}_{J'm'J}(z'), \quad \text{(B8)}$$

and consequently

$$\Psi = \sum_{j_0,\sigma}(j_0^2-\sigma^2)(-1)^{(\sigma-j_0)}\,d^{\sigma j_0}_{JmJ'}(z)\,d^{-\sigma j_0}_{J'm'J}(z') = -\Psi = 0. \quad \text{(B9)}$$

Thus we infer that the complete sum of nonsense-channel terms of the exchange amplitude is zero. The same result will apply to the direct part of the amplitude and hence to its entirety.

[12] See Ref. 5, Eq. (A11).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

On the Green's Functions for the Bethe–Salpeter Equation*

JUSTIN C. HUANG AND BRIAN DEFACIO
Physics Department, University of Missouri, Columbia, Missouri

(Received 21 June 1968)

We have studied four possible Green's functions of the Bethe–Salpeter equation for equal mass scalar bosons by considering contours in an off-the-mass-shell complex energy (Ω) plane and an on-the-mass-shell magnitude of 3-momenta (k) or energy (ω_k) complex plane. We find that three of the Green's functions—causal, retarded, and advanced Green's function—cannot lead to scattering solutions. Only a new Green's function, which we have called the scattering Green's function, with appropriate restrictions on the interaction $V(x)$, can lead to scattering states.

1. INTRODUCTION

The Bethe–Salpeter (BS) equation was first introduced by Bethe and Salpeter[1] as a relativistic bound-state equation. Wick[2] studied the equation in detail and obtained a simpler (and different) equation by utilizing a Euclidean metric in the interaction propagator for the exchanged particle. This equation is customarily the "Wick-rotated Bethe–Salpeter equation." Cutkosky[3] solved the Wick-rotated BS equation for the ladder approximation with zero-mass quanta exchanged.

Kemmer and Salam[4] extended the BS equation to scattering energies, and recently Tiktopoulos,[5] Schwartz and Zemach,[6] Nakanishi,[7] Saenger,[8] and others have studied or used the BS equation with scattering energies. Since the above papers were concerned with interactions and since the present work is confined to Green's functions of the BS equation, the interested reader is referred to the references contained in those papers. The work by Keam,[9] Ciafaloni and Menotti,[10] and Schwartz and Zemach[6] contain most of the existing results on the causal Green's function of the BS equation. All three of these papers,[6,9,10] as well as those papers they referred to, used causal boundary conditions[11] in order to obtain their scattering boundary conditions.

In the present work on the BS equation, we will reexamine the equal mass causal Green's function, will study the advanced and retarded Green's function, and will introduce a new Green's function which we will call the scattering Green's function. Two independent determinations[6,9] of the causal Green's function have used the same inapplicable formula from a mathematical reference book. We consider two scalar bosons with mass m and 4-momentum operators $\hat{p}_1$ and $\hat{p}_2$. In configuration space, the BS equation can be written as

$$(\hat{p}_1^2 + m^2)(\hat{p}_2^2 + m^2)\Psi(x_1, x_2) = V(x_1, x_2)\Psi(x_1, x_2), \tag{1.1}$$

where $\Psi(x_1, x_2)$ is the BS amplitude, x_1 and x_2 are the space–time 4-vectors which locate particles 1 and 2, and $V(x_1, x_2)$ is the interaction potential operator between the two particles. We choose conventions and units such that $\hbar = c = 1$, and the inner product between the two 4-vectors p and x is $px = \mathbf{p}\cdot\mathbf{x} - P_0x_0$.

In Sec. 2, we present the kinematics and the pole structure of the Green's functions in the off-the-mass-shell energy plane. In Sec. 3, the causal Green's function is explicitly calculated by the specification of contours in an on-the-energy-shell magnitude of momentum plane. In Sec. 4, the same method is used to study the retarded and advanced Green's function in detail. In Sec. 5, a new Green's function which we call the scattering Green's function is considered. In Sec. 6, the summary and conclusions are presented, and some calculational details are relegated to appendices.

2. POSSIBLE POLES FOR THE GREEN'S FUNCTION

Let us introduce the following transformations on the 4-momenta operators and the space–time coordinates mentioned in Sec. 1:

$$\begin{aligned} \hat{p} &= \mu_2\hat{p}_1 - \mu_1\hat{p}_2 = (\hat{\mathbf{p}}, i\hat{p}_0), \\ \hat{P} &= \hat{p}_1 + \hat{p}_2, \\ x &= x_1/\mu_2 - x_2/\mu_1 = (\mathbf{x}, it), \\ X &= x_1 + x_2, \end{aligned} \tag{2.1}$$

* This work was supported in part by a grant from the University of Missouri Research Council.

[1] H. A. Bethe and E. E. Salpeter, Phys. Rev. **84**, 1232 (1954).
[2] G.-C. Wick, Phys. Rev. **96**, 1124 (1954).
[3] R. E. Cutkosky, Phys. Rev. **96**, 1136 (1954).
[4] N. Kemmer and A. Salam, Proc. Roy. Soc. (London) **A230**, 266 (1955).
[5] G. Tiktopoulos, Phys. Rev. **136**, B275 (1964).
[6] C. Schwartz and C. Zemach, Phys. Rev. **141**, 1454 (1966).
[7] N. Nakanishi, Progr. Theoret. Phys. (Kyoto) **38**, 226 (1967).
[8] R. M. Saenger, J. Math. Phys. **8**, 2366 (1967).
[9] R. F. Keam, J. Math. Phys. **7**, 2196 (1966).
[10] M. Ciafaloni and P. Menotti, Phys. Rev. **140**, B929 (1965).
[11] For an explanation of why the causal Green's function was initially used, the reader is referred to S. S. Schweber, *An Introduction to Quantum Field Theory* (Harper and Row, New York, 1962), 2nd ed., Chap. 17, p. 714.

where $\mu_1 + \mu_2 = 1$ is the only condition we wish to impose on the μ's. If we let the eigenvalue of $\hat{P}$ be K, the wavefunction $\Psi(x_1, x_2)$ can be rewritten as

$$\Psi(x_1, x_2) = e^{iKX}\psi(x). \tag{2.2}$$

In the center-of-momentum system, the BS equation reduces to the following equation in terms of the relative coordinate x:

$$[\hat{\mathbf{p}}^2 - (\hat{p}_0 - \nu + \omega)^2 + m^2][\mathbf{p}^2 - (\hat{p}_0 - \nu - \omega)^2 + m^2]\psi(x) = V(x)\psi(x), \tag{2.3}$$

where ω represents the energy of each of the two particles in the center-of-momentum system and

$$\nu = (\mu_2 - \mu_1)\omega. \tag{2.4}$$

The assumption has been made that $V(x_1, x_2)$ can be written as $V(x)$.

The Green's function $G(x, x')$ must satisfy the following differential equation:

$$[\hat{\mathbf{p}}^2 - (\hat{p}_0 - \nu + \omega)^2 + m^2][\hat{\mathbf{p}}^2 - (\hat{p}_0 - \nu - \omega)^2 + m^2]\, G(x, x') = \delta^4(x - x'). \tag{2.5}$$

The solutions to the inhomogeneous BS equation can then be written in the form

$$\psi(x) = \psi_0(x) + \int d^4x' G(x, x')V(x')\psi(x'), \tag{2.6}$$

where $\psi_0(x)$ is the solution to the homogeneous equation (plane wave). Let us introduce G_κ, the Fourier transform of $G(x, x')$. They are related by

$$G(x, x') = \frac{1}{(2\pi)^4}\int d^4\kappa e^{i\kappa(x-x')}G_\kappa, \tag{2.7}$$

where $\kappa = (\mathbf{k}, i\Omega)$, $d^4\kappa = d^3k\, d\Omega$, and the 4-vector κ is to be off-the-mass-shell. Using the usual form for the operators $\hat{\mathbf{p}}$ and $\hat{p}_0$ in the coordinate representation

$$\hat{\mathbf{p}} = -i\frac{\partial}{\partial \mathbf{x}} \quad \text{and} \quad \hat{p}_0 = i\frac{\partial}{\partial t}, \tag{2.8}$$

we find that G_κ has the form

$$G_\kappa = \{[\mathbf{k}^2 - (\Omega - \nu + \omega)^2 + m^2] \times [\mathbf{k}^2 - (\Omega - \nu - \omega)^2 + m^2]\}^{-1}. \tag{2.9}$$

On performing the integration over Ω and the angular variables of the 3-momenta $\mathbf{k}$, Eq. (2.7) becomes

$$G(x, x') = \frac{2}{(2\pi)^3}\frac{e^{-i\nu T}}{R}\int_0^\infty dk k \sin(kR) I(\omega_k), \tag{2.10}$$

where

$$I(\omega_k) = \int_{-\infty}^{+\infty} d\Omega e^{-i\Omega T}/[\omega_k^2 - (\Omega + \omega)^2] \times [\omega_k^2 - (\Omega - \omega)^2] \tag{2.11}$$

and

$$R = |\mathbf{x} - \mathbf{x}'|, \quad T = t - t', \quad k = |\mathbf{k}|, \quad \omega_k = (k^2 + m^2)^{\frac{1}{2}}. \tag{2.12}$$

The integral $I(\omega_k)$ as defined in Eq. (2.11) is unbounded if ω is real. The integrand contains four poles at

$$\Omega_1 = -(\omega_k + \omega), \quad \Omega_2 = \omega_k - \omega, \quad \Omega_3 = -(\omega_k - \omega), \quad \Omega_4 = \omega_k + \omega. \tag{2.13}$$

We must specify contours in the complex Ω plane which would bypass the poles. The residue to each of the four poles can be easily calculated and the Green's functions which result from the contribution of each pole are listed below:

$$\begin{aligned}
{}^{\pm}G_1 &= \frac{\pm i}{(4\pi)^2}\frac{e^{i(\omega-\nu)T}}{\omega R}\int_0^\infty dk \frac{k \sin(kR)e^{i\omega_k T}}{\omega_k(\omega_k + \omega)}, \\
{}^{\pm}G_2 &= \frac{\pm i}{(4\pi)^2}\frac{e^{i(\omega-\nu)T}}{\omega R}\int_0^\infty dk \frac{k \sin(kR)e^{-i\omega_k T}}{\omega_k(\omega_k - \omega)}, \\
{}^{\pm}G_3 &= \frac{\mp i}{(4\pi)^2}\frac{e^{-i(\omega+\nu)T}}{\omega R}\int_0^\infty dk \frac{k \sin(kR)e^{i\omega_k T}}{\omega_k(\omega_k - \omega)}, \\
{}^{\pm}G_4 &= \frac{\mp i}{(4\pi)^2}\frac{e^{-i(\omega+\nu)T}}{\omega R}\int_0^\infty dk \frac{k \sin(kR)e^{-i\omega_k T}}{\omega_k(\omega_k + \omega)}.
\end{aligned} \tag{2.14}$$

The superscript "+" or "−" on the Green's functions refers to the sign of T. If $T > 0$, we use ${}^{+}G_i$, and if $T < 0$, we use ${}^{-}G_i$. It is assumed that, in each case listed above, the correct domain of T would be chosen so that the semicircular integrals over the upper or lower half of the complex Ω plane will vanish. The integrands of the k-integrals still contain singular points, which we will handle later. These singular points in the complex k plane are to be expected since we are working with a fourth-order differential equation in four variables and, in general, a choice of contour in the off-the-mass-shell Ω plane is not sufficient to define uniquely our Green's function; we must also specify contours in the on-the-mass-shell k plane in order to obtain unique solutions. The evaluation of the k integrals will depend on the relative magnitudes of R and $|T|$ as well as the sign of T. We shall call the space–time interval $(x - x')$ spacelike, if $R > |T|$, and timelike, if $|T| > R$. The corresponding Green's function will be labeled G^σ if spacelike, and G_τ if timelike.

The specification of contours corresponds to the choice of different boundary conditions for the BS equation. We will consider four possible choices of contours in the complex Ω plane.

3. CAUSAL GREEN'S FUNCTION G_c

A. G_c for $T > 0$

The causal Green's function is defined by the contour shown in Fig. 1. For $T > 0$,

$$ {}^{+}G_c = {}^{+}G_2 + {}^{+}G_4, \tag{3.1} $$

which can be expressed using Eq. (2.14) as

$$ {}^{+}G_c = \frac{2i}{(4\pi)^2}\frac{e^{-ivT}}{\omega R}[i\sin(\omega T)\,{}^{+}F_1(q, R, T) + \omega\cos(\omega T)\,{}^{+}F_2(q, R, T)], \tag{3.2} $$

where

$$ F_1(q, R, \pm T) = \int_0^\infty dk\,\frac{k\sin(kR)e^{\mp i\omega_k T}}{k^2 - q^2} \tag{3.3} $$

and

$$ F_2(q, R, \pm T) = \int_0^\infty dk\,\frac{k\sin(kR)e^{\mp i\omega_k T}}{\omega_k(k^2 - q^2)}. \tag{3.4} $$

The superscript "+" or "−" on F_1 and F_2 refers to the sign of T that we must use in evaluating the integrals. If q is real, the integrands of F_1 and F_2 will have singular points at $k = \pm q$. We shall, as in the nonrelativistic scattering case, attach to q an infinitesimal positive imaginary part ($q \to q + i\epsilon$). Other choices of contours in the complex k plane are also possible; for example, we can let $q \to q - i\epsilon$. Since we are interested in outgoing scattered waves, we shall use the $+i\epsilon$ contour. F_1 and F_2 can now be written as

$$ F_1(q + i\epsilon, R, \pm T) = \int_0^\infty dk\,\frac{k\sin(kR)e^{\mp i\omega_k T}}{(k + q + i\epsilon)(k - q - i\epsilon)}, \tag{3.5} $$

$$ F_2(q + i\epsilon, R, \pm T) = \int_0^\infty dk\,\frac{k\sin(kR)e^{\mp i\omega_k T}}{\omega_k(k + q + i\epsilon)(k - q - i\epsilon)}. \tag{3.6} $$

It is helpful to note that F_1 and F_2 are related by

$$ \frac{dF_2}{dT} = -iF_1. \tag{3.7} $$

The evaluation of F_1 can be performed by routine contour methods, and the details are shown in Appendices A and B. The results depend on the relative magnitude of R and $|T|$ as well as the sign of T. We shall use the same notation as was indicated for the Green's function: F^s, if the interval is spacelike, and F^t, if the interval is timelike. From Appendix A, we have

$$ {}^{+}F_1^s(q + i\epsilon, R, T) = \tfrac{1}{2}\pi e^{iqR}e^{-i\omega T} + M(R, T), \tag{3.8} $$

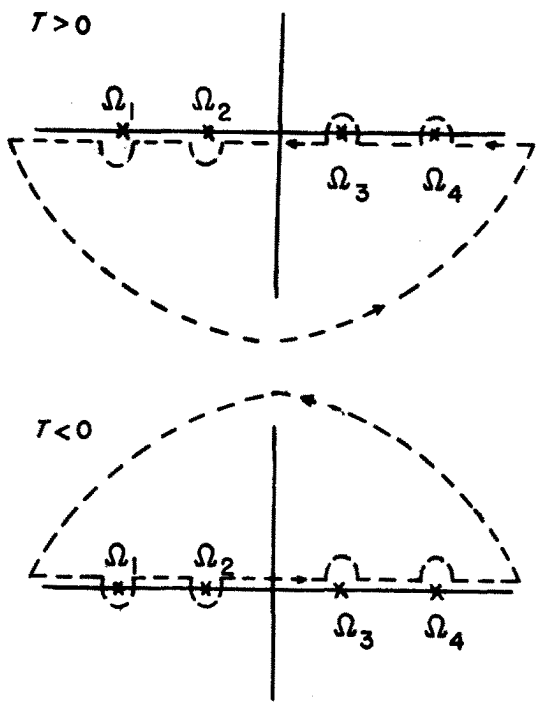


FIG. 1. Contour for the causal Green's function in the complex Ω plane.

where

$$ M(R, T) = -\int_0^m dk\,\frac{ke^{-kR}\sin[(m^2 - k^2)^{\frac{1}{2}}T]}{k^2 + q^2}. \tag{3.9} $$

From Appendix B, we have

$$ {}^{+}F_1^t(q + i\epsilon, R, T) = iW(R, -T) + iN(R, -T), \tag{3.10} $$

where

$$ W(R, \pm T) = \int_0^\infty d\omega_k\,\frac{\omega_k e^{\pm\omega_k T}\sinh[(\omega_k^2 + m^2)^{\frac{1}{2}}R]}{\omega_k^2 + \omega^2} \tag{3.11} $$

and

$$ N(R, \pm T) = \int_0^m dk\,\frac{k\exp[\pm i(m^2 - k^2T)^{\frac{1}{2}}]}{k^2 + q^2}\sinh(kR). \tag{3.12} $$

F_2 can be determined from F_1 by the integration of Eq. (3.7) with respect to T:

$$ F_2(q + i\epsilon, R, \pm T) = F_2(q + i\epsilon, R, 0) - i\int_0^T F_1(q + i\epsilon, R, \pm T')\,dT'. \tag{3.13} $$

Since G_c is not defined as yet for $T = 0$, we may arbitrarily set $F_2(q + i\epsilon, R, 0)$ equal to zero without loss of generality. This will be shown later to be equivalent to setting $G_c(q + i\epsilon, R, 0)$ equal to zero at $T = 0$ and will also satisfy the continuity requirement

of G_c at $T = 0$. Equation (3.13) becomes

$$F_2(q + i\epsilon, R, T) = -i\int_0^T dT' F_1(q + i\epsilon, R, T') \tag{3.14}$$

and, upon integration of ${}^+F_1^a$ and ${}^+F_1^r$, we have

$${}^+F_2^a(q + i\epsilon, R, T) = (\pi/2\omega)e^{iqR}(e^{-i\omega T} - 1) - iM'(R, T) \tag{3.15}$$

and

$${}^+F_2^r(q + i\epsilon, R, T) = W'(R, -T) + N'(R, -T), \tag{3.16}$$

where

$$M'(R, T) = \int_0^T dT' M(R, T'), \tag{3.17}$$

$$W'(R, \pm T) = \int_0^T dT' W(R, \pm T'), \tag{3.18}$$

and

$$N'(R, \pm T) = \int_0^T dT' N(R, \pm T'). \tag{3.19}$$

We can now express ${}^+G_c$ in terms of the above expressions. For spacelike intervals

$${}^+G_c^a(q + i\epsilon, R, T) = (i/8\pi^2)(e^{-ivT}/\omega R)\{\tfrac{1}{2}\pi[1 - \cos(\omega T)]e^{iqR} - i\omega\cos(\omega T)M'(R, T) + i\sin(\omega T)M(R, T)\} \tag{3.20}$$

and for timelike intervals

$${}^+G_c^r(q + i\epsilon, R, T) = \frac{i}{8\pi^2}\frac{e^{-ivT}}{\omega R}\{\omega\cos(\omega T)[W'(R, -T) + N'(R, -T)] - \sin(\omega T)[W(R, -T) + N(R, -T)]\}. \tag{3.21}$$

For large values of R, $M(R, T)$ and $M'(R, T)$ damp exponentially while $N(R, \pm T)$ and $N'(R, \pm T)$ increase exponentially. In Eq. (3.20) we can thus neglect the last two terms on the right-hand side of the equation for large values of R and write, for ${}^+G_c^a$,

$${}^+G_c^a(q + i\epsilon, R, T) \to \frac{i}{16\pi}e^{-ivT}\left(\frac{1 - \cos(\omega T)}{\omega}\right)\frac{e^{iqR}}{R}. \tag{3.22}$$

However, in Eq. (3.21), the right-hand side becomes unbounded and

$${}^+G_c^r(q + i\epsilon, R, T) \to \infty. \tag{3.23}$$

B. G_c for $T < 0$

For $T < 0$ from Fig. 1, we have

$${}^-G_c = {}^-G_1 + {}^-G_3. \tag{3.24}$$

Substituting from Eq. (2.14), we have

$${}^-G_c(q + i\epsilon, R, T) = [-2i/(4\pi)^2](e^{-ivT}/\omega R) \times [i\sin(\omega T)\,{}^-F_1(q + i\epsilon, R, -T) - \omega\cos(\omega T)\,{}^-F_2(q + i\epsilon, R, -T)], \tag{3.25}$$

where $+i\epsilon$ has been added to q for the same reason as given in Sec. 3A and F_1 and F_2 are given by Eqs. (3.3) and (3.4). From Appendices A and B, we find

$${}^-F_1^a(q + i\epsilon, R, -T) = \tfrac{1}{2}\pi e^{iqR}e^{i\omega T} - M(R, T) \tag{3.26}$$

and

$${}^-F_1^r(q + i\epsilon, R, -T) = iW(R, T) + iN(R, T). \tag{3.27}$$

Applying Eq. (3.14), we have

$${}^-F_2^a(q + i\epsilon, R, -T) = (\pi/2\omega)e^{iqR}(e^{i\omega T} - 1) - iM'(R, T) \tag{3.28}$$

and

$${}^-F_2^r(q + i\epsilon, R, -T) = -W'(R, T) - N'(R, T). \tag{3.29}$$

The causal Green's function for negative T becomes, on substituting into Eq. (3.25),

$${}^-G_c^a(q + i\epsilon, R, T) = (i/8\pi^2)(e^{-ivT}/\omega R)[\tfrac{1}{2}\pi(1 - \cos(\omega T))e^{iqR} - i\omega\cos(\omega T)M'(R, T) + i\sin(\omega T)M(R, T)] \tag{3.30}$$

for spacelike intervals and

$${}^-G_c^r(q + i\epsilon, R, T) = (i/8\pi^2)(e^{-ivT}/\omega R)\{\sin(\omega T)[W(R, T) + N(R, T)] - \omega\cos(\omega T)[W'(R, T) + N'(R, T)]\} \tag{3.31}$$

for timelike intervals. For large values of R, using the properties of $M(R, T)$ and $N(R, \pm T)$ mentioned in Sec. 3A, we have

$${}^-G_c^a(q + i\epsilon, R, T) \to \frac{i}{16\pi}e^{-ivT}\left(\frac{1 - \cos(\omega T)}{\omega}\right)\frac{e^{iqR}}{R} \tag{3.32}$$

and

$${}^-G_c^r(q + i\epsilon, R, T) \to \infty. \tag{3.33}$$

4. RETARDED AND ADVANCED GREEN'S FUNCTION

A. The Retarded Green's Function G_R

The retarded Green's function G_R is defined by the contour shown in Fig. 2. For $T < 0$, ${}^-G_R = 0$ for both spacelike and timelike intervals, while, for $T > 0$,

$${}^+G_R = {}^+G_1 + {}^+G_2 + {}^+G_3 + {}^+G_4. \tag{4.1}$$

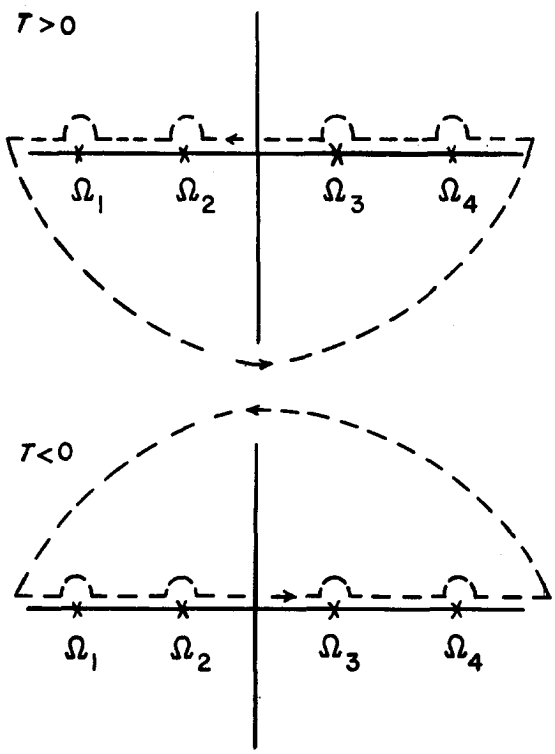


FIG. 2. Contour for the retarded Green's function in the complex Ω plane.

Substituting from Eq. (2.14), we have

$$
\begin{aligned}
{}^{+}G_R(q + i\epsilon, R, T) = {} & [2i/(4\pi)^2](e^{-ivT}/\omega R) \\
& \times \{i \sin(\omega T)[{}^{+}F_1(q + i\epsilon, R, T) \\
& + {}^{+}F_1(q + i\epsilon, R, -T)] \\
& + \omega \cos(\omega T)[{}^{+}F_2(q + i\epsilon, R, T) \\
& - {}^{+}F_2(q + i\epsilon, R, -T)]\}. \qquad (4.2)
\end{aligned}
$$

We may use Eq. (3.14) to determine F_2 from F_1, where defining $F_2(q + i\epsilon, R, 0) = 0$ again has the effect of setting $G_R = 0$ at $T = 0$ and making it continuous at $T = 0$. Inserting the possible expressions for ${}^{+}F_1(q + i\epsilon, R, \pm T)$ into Eq. (4.2), we have

$$
{}^{+}G_R^{\sigma}(q + i\epsilon, R, T) = 0 \qquad (4.3)
$$

for spacelike intervals, and

$$
\begin{aligned}
{}^{+}G_R^{r}(q + i\epsilon, R, T) & \\
= {} & (-i/8\pi^2)(e^{-ivT}/\omega R) \\
& \times \{\tfrac{1}{2}\pi[1 - \cos(\omega T)]e^{iqR} + i \sin(\omega T)M(R, T) \\
& - i\omega \cos(\omega T)M'(R, T) + \sin(\omega T) \\
& \times [W(R, -T) + N(R, -T)] \\
& - \omega \cos(\omega T)[W'(R, -T) + N'(R, -T)]\} \\
& \qquad (4.4)
\end{aligned}
$$

for timelike intervals. For large values of R, since $N(R, -T)$ and $N'(R, -T)$ are not bounded,

$$
{}^{+}G_R^{r}(q + i\epsilon, R, T) \to \infty. \qquad (4.5)
$$

B. Advanced Green's Function G_A

The advanced Green's function G_A is defined by the contour shown in Fig. 3. For $T > 0$, ${}^{+}G_A = 0$ for both spacelike and timelike intervals, while, for $T < 0$,

$$
{}^{-}G_A = {}^{-}G_1 + {}^{-}G_2 + {}^{-}G_3 + {}^{-}G_4. \qquad (4.6)
$$

We may use Eq. (4.2) to evaluate ${}^{-}G_A$ if we replace the "+" superscripts by the "−" superscripts. With similar methods to those used in part A of this section, we have

$$
{}^{-}G_A^{\sigma}(q + i\epsilon, R, T) = 0 \qquad (4.7)
$$

for spacelike intervals, and

$$
\begin{aligned}
{}^{-}G_A^{r}(q + i\epsilon, R, T) & \\
= {} & (-i/8\pi^2)(e^{-ivT}/\omega R) \\
& \times \{\tfrac{1}{2}\pi(1 - \cos(\omega T))e^{iqR} + i \sin(\omega T)M(R, T) \\
& - i\omega \cos(\omega T)M'(R, T) - \sin(\omega T) \\
& \times [W(R, T) + N(R, T)] \\
& + \omega \cos(\omega T)[W'(R, T) + N'(R, T)]\} \qquad (4.8)
\end{aligned}
$$

for timelike intervals. For large values of R, we once more have

$$
{}^{-}G_A^{r}(q + i\epsilon, R, T) \to \infty. \qquad (4.9)
$$

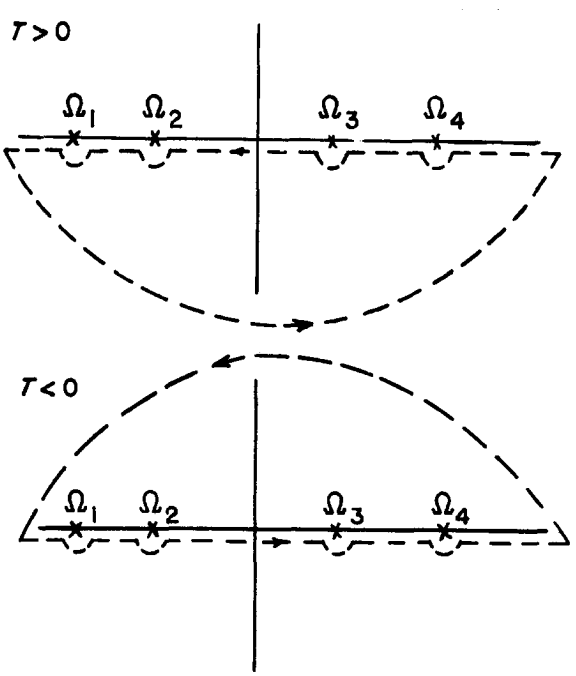


FIG. 3. Contour for the advanced Green's function in the complex Ω plane.

5. SCATTERING GREEN'S FUNCTION G_s

The scattering Green's function G_s is defined by the contour shown in Fig. 4. For $T > 0$, we have

$$^{+}G_s = {}^{+}G_1 + {}^{+}G_2, \tag{5.1}$$

which, by using Eq. (2.14), can be expressed as

$$^{+}G_s(q + i\epsilon, R, T) = [2i/(4\pi)^2](e^{i(\omega-\nu)T}/\omega R) \times [{}^{+}H_1(q + i\epsilon, R, T) - i\omega\, {}^{+}H_2(q + i\epsilon, R, T)], \tag{5.2}$$

where

$$^{\pm}H_1(q + i\epsilon, R, T) = \tfrac{1}{2}[{}^{\pm}F_1(q + i\epsilon, R, -T) + {}^{\pm}F_1(q + i\epsilon, R, T)] \tag{5.3}$$

and

$$^{\pm}H_2(q + i\epsilon, R, T) = (2i)^{-1}[{}^{\pm}F_2(q + i\epsilon, R, -T) - {}^{\pm}F_2(q + i\epsilon, R, T)]. \tag{5.4}$$

The F_1's are listed in Appendices A and B, and the F_2's can be calculated by the use of Eq. (3.13). For this case, we need not assign a value to $F_2(q + i\epsilon, R, 0)$ since it will cancel out in Eq. (5.4). The scattering Green's function becomes

$$^{+}G_s^a(q + i\epsilon, R, T) = \frac{i}{16\pi}\frac{e^{-i\nu T}}{\omega}\frac{e^{iqR}}{R} \tag{5.5}$$

for spacelike intervals (without the necessity for making any approximations for large values of R), and

$$\begin{aligned} {}^{+}G_s^r(q + i\epsilon, R, T) &= (i/16\pi^2)(e^{i(\omega-\nu)T}/\omega R) \\ &\times \{\tfrac{1}{2}\pi e^{iqR} - M(R, T) + i\omega M'(R, T) \\ &+ i[W(R, -T) + N(R, -T)] \\ &+ \omega[W'(R, -T) + N'(R, -T)]\} \end{aligned} \tag{5.6}$$

for timelike intervals. For large values of R, we again have

$$^{+}G_s^r(q + i\epsilon, R, T) \to \infty. \tag{5.7}$$

For $T < 0$, the scattering Green's function becomes

$$^{-}G_s = {}^{-}G_3 + {}^{-}G_4, \tag{5.8}$$

which can be expressed as

$$^{-}G_s(q + i\epsilon, R, T) = [2i/(4\pi)^2](e^{-i(\omega+\nu)T}/\omega R) \times [{}^{-}H_1(q + i\epsilon, R, T) + i\omega\, {}^{-}H_2(q + i\epsilon, R, T)]. \tag{5.9}$$

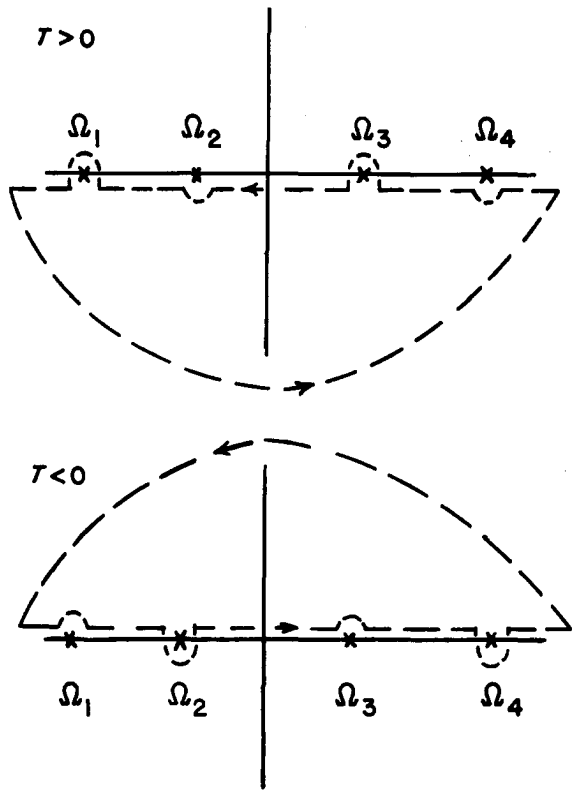


FIG. 4. Contour for the scattering Green's function in the complex Ω plane.

Making use of Eqs. (5.3), (5.4), (3.13), and Appendices A and B, we have

$$^{-}G_s^a(q + i\epsilon, R, T) = (i/16\pi)(e^{-i\nu T}/\omega)(e^{iqR}/R) \tag{5.10}$$

for spacelike intervals (without any approximations at large values of R), and

$$\begin{aligned} {}^{-}G_s^r(q + i\epsilon, R, T) &= (i/16\pi^2)(e^{-i(\omega+\nu)T}/\omega R) \\ &\times \{\tfrac{1}{2}\pi e^{iqR} + M(R, T) + i\omega M'(R, T) \\ &+ i[W(R, T) + N(R, T)] \\ &- \omega[W'(R, T) + N'(R, T)]\} \end{aligned} \tag{5.11}$$

for timelike intervals. For large values of R, we have once more

$$^{-}G_s^r(q + i\epsilon, R, T) \to \infty. \tag{5.12}$$

6. SUMMARY AND CONCLUSIONS

The results from Secs. 3–5 indicate that, for timelike intervals of $(x - x')$, no scattering solutions are possible for any of the Green's functions considered. The terms responsible for this behavior are the ones involving the integral $N(R, \pm T)$, defined in Eq. (3.12), which diverges at large values of R. For spacelike intervals, the Green's functions were all well behaved, however; only the scattering Green's

function G_s has the correct asymptotic behavior necessary to represent scattering solutions. The causal Green's function introduced in Sec. 3 represents, for spacelike intervals, waves traveling in both incoming and outgoing directions and cannot represent a scattering solution. The retarded and advanced Green's functions introduced in Sec. 4 are all equal to zero, for spacelike intervals; they, thus, have the required property for constructing a causal field theory, if we regard the BS equation as a field equation. But, again, they do not represent scattering solutions.

The scattering Green's function introduced in Sec. 5 has, for spacelike intervals, the exact form required for outgoing wave solutions. The solutions were, furthermore, obtained without the necessity of making any approximations for large values of R. The scattering amplitude can be determined by considering the full solution to the BS equation:

$$\psi(x) = \psi_0(x) + \int d^4x' G_s(x, x')V(x')\psi(x'). \quad (2.6)$$

If we can assume that $V(x')$ is such that the integral in Eq. (2.6) is zero when $(x - x')$ is timelike, we can write

$$\psi(x) = \psi_0(x) + \frac{i}{16\pi\omega}\int d^4x' e^{-ivT}\frac{e^{iqR}}{R}V(x')\psi(x'). \quad (6.1)$$

With the usual approximation of $|\mathbf{x}| > |\mathbf{x}'|$, we have

$$\psi(x) = \psi_0(x) + e^{+ivt}(e^{iq|\mathbf{x}|}/x) \times \left((i/16\pi\omega)\int d^4x' e^{-ivt'}e^{-iq'\cdot x'}V(x')\psi(x')\right). \quad (6.2)$$

The scattering amplitude $f(\mathbf{q} \to \mathbf{q}')$ can now be identified as

$$f(\mathbf{q} \to \mathbf{q}') = (i/16\pi\omega)\int d^4x' e^{ivt'}e^{-iq\cdot x'}V(x')\psi(x'). \quad (6.3)$$

The above scattering amplitude agrees with the results of Schwartz and Zemach,[6] but their results were obtained by using the causal Green's function, which we have shown cannot lead to a scattering state. The discrepancy can be found in an assertion used by Schwartz and Zemach in assuming that Q [$Q = (\beta^2 - q^2)^{\frac{1}{2}}$, where β is real and varies from $-\omega$ to $+\omega$] can be taken to lie in the fourth quadrant of the complex plane.[12] This can be shown to be incorrect on closer examination—if we let q be $q + i\epsilon$, Q lies in the third quadrant of the complex plane, and if we let q be $q - i\epsilon$, Q lies in the first quadrant of the complex plane.

The scattering Green's function is also the most natural one to be chosen from the point of view of nonrelativistic scattering theory (Lippmann–Schwinger equation), where the extension $q \to q + i\epsilon$ or $\omega \to \omega + i\epsilon$ led us to contours which gave outgoing scattering states. If we insert $\omega + i\epsilon$ for ω into Eqs. (2.11) and (2.13), the substitution would correspond to a contour in the complex Ω plane which would be equivalent to the contour for G_s shown in Fig. 4.

A further remark on the choice of contour in the k plane which corresponds to the extension $q \to q - i\epsilon$: This would not prevent the Green's functions, for timelike intervals, from becoming unbounded at large values of R, but, for spacelike intervals, it has the effect of replacing q by $-q$ in our expressions. For example, the scattering Green's function, for spacelike intervals, becomes

$$^{\pm}G_s^a(q - i\epsilon, R, T) = (i/16\pi)(e^{-ivT}/\omega)(e^{-iqR}/R), \quad (6.4)$$

which represents incoming waves, as is expected from our knowledge of the nonrelativistic Lippmann–Schwinger equation.

APPENDIX A

We shall consider in this appendix the evaluation of the function $F_1(q + i\epsilon, R, \pm T)$ first introduced in Sec. 3. It has the form

$$F_1(q + i\epsilon, R, \pm T) = \int_0^\infty dk \frac{k \sin(kR)e^{\mp i\omega_k T}}{(k + q + i\epsilon)(k - q - i\epsilon)}. \quad (3.5)$$

The method utilizes integrals of the form

$$L(\pm, \pm) = \int dk \frac{ke^{\pm i(kR \pm \omega_k T)}}{(k + q + i\epsilon)(k - q - i\epsilon)} \quad (A1)$$

to evaluate F_1. Consider the analytic properties of the integrand of $L(\pm, \pm)$. In the finite complex k plane, it has simple poles at $k = -q - i\epsilon$ and $k = q + i\epsilon$ plus a cut with branch points at $k = \pm im$.

The contour to be considered in the evaluation of $L(\pm, \pm)$ depends on the behavior of the exponential $e^{\pm i(kR \pm \omega_k T)}$ for large values of $|k|$. Let

$$k = k_R + ik_I$$

and

$$\omega_k = \omega_{kR} + i\omega_{kI}; \quad (A2)$$

then the exponential can be written as

$$e^{\pm i(kR \pm \omega_k T)} = e^{\pm i(k_R R \pm \omega_{kR} T)}|e^{\mp(k_I R \pm \omega_{kI} T)}. \quad (A3)$$

Since the first factor on the right hand side of Eq.

[12] *Note Added in Proof:* Long after this manuscript was submitted, B. C. McInnis and C. Schwartz [Phys. Rev. 117, 2621 (1969)] claimed to have "corrected" an error in Schwartz and Zemach (Ref. 6). They erroneously added only a pole for the k-singularities instead of two branch points connected by a cut. If the reader is not perfectly clear on this point, he should reread Appendices A and B.

TABLE I. Contours in the complex k plane.

$\alpha(+)$	T	Space–Time interval	Contour
>0	>0	σ[a]	UHP[c]
>0	>0	τ[b]	UHP
<0	>0	σ	LHP[d]
>0	>0	τ	LHP
>0	<0	σ	UHP
<0	<0	τ	UHP
<0	<0	σ	LHP
<0	<0	τ	LHP
$\alpha(-)$	T	Space–Time interval	Contour
>0	>0	σ	UHP
<0	>0	τ	UHP
<0	>0	σ	LHP
<0	>0	τ	LHP
>0	<0	σ	UHP
>0	<0	τ	UHP
<0	<0	σ	LHP
>0	<0	τ	LHP

[a] σ refers to spacelike intervals.
[b] τ refers to timelike intervals.
[c] UHP refers to the upper half plane.
[d] LHP refers to the lower half plane.

(A3) is just an oscillating function with unit magnitude, it cannot provide us with any information on the possible contours. Let us examine the second factor $e^{\mp\alpha(\pm)}$, where

$$\alpha(\pm) = (k_I R \pm \omega_{kI} T). \tag{A4}$$

In order that an integral over a semicircular contour vanish at large values of $|k|$, we must have $\mp\alpha(\pm) < 0$. For the possible choices of the exponential in Eq. (A3), the conditions on $\alpha(\pm)$ are

$$\begin{aligned} e^{+i(kR+\omega_k T)} &\to \alpha(+) > 0, \\ e^{-i(kR+\omega_k T)} &\to \alpha(+) < 0, \\ e^{+i(kR-\omega_k T)} &\to \alpha(-) > 0, \\ \text{and} \\ e^{-i(kR-\omega_k T)} &\to \alpha(-) < 0. \end{aligned} \tag{A5}$$

The sign of $\alpha(\pm)$ can be determined if we express ω_{kI} as a function of k_I and k_R. The relationship is

$$\omega_{kI} = 2^{-\frac{1}{2}}\{[(k_R^2 - k_I^2 + m^2)^2 + 4k_R^2 k_I^2]^{\frac{1}{2}} - [k_R^2 - k_I^2 + m^2]\}^{\frac{1}{2}}. \tag{A6}$$

Using Eq. (A6), we constructed Table I, which will inform us as to the possible contours in the complex k plane that must be taken if we are given the sign of T and the relative magnitudes of R and $|T|$ (spacelike or timelike intervals). Using Eqs. (A1), (A5), and Table I, we can now evaluate $F_1(q + i\epsilon, R, \pm T)$.

Let us consider

$$L(+, +) = \int dk \frac{k e^{i(kR+\omega_k T)}}{(k + q + i\epsilon)(k - q - i\epsilon)}. \tag{A7}$$

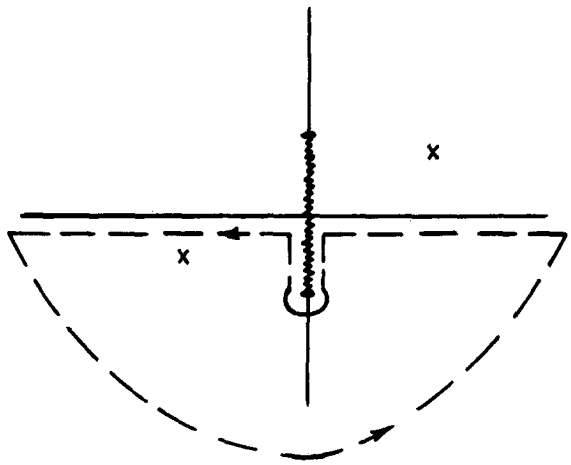


FIG. 5. Contour for $L(+, +)$ in the complex k plane.

From Eq. (A5), we must require that $\alpha(+) > 0$. From Table I, for $T > 0$ and spacelike or timelike intervals, and for $T < 0$ and spacelike intervals, we must use the upper half of the complex k plane. We next apply Cauchy's residue theorem to $L(+, +)$:

$$\int_C dk \frac{k e^{i(kR+\omega_k T)}}{(k + q + i\epsilon)(k - q - i\epsilon)} = 2\pi i(\text{Res at } k = q + i\epsilon), \tag{A8}$$

where C is shown in Fig. 5. On setting the integrals over the large and small semicircles equal to zero and combining the integrals along the left and right hand side of the real axis, we have

$$\int_0^\infty dk \frac{k e^{i\omega_k T} \sin(kR)}{(k + q + i\epsilon)(k - q - i\epsilon)} = \tfrac{1}{2}\pi e^{iqR} e^{i\omega T} - M(R, T), \tag{A9}$$

where the first term on the right-hand side gives the contribution of the residue at $k = q + i\epsilon$ and $M(R, T)$, defined in Eq. (3.9), represents the contribution by the cut. We can now write

$$\left.\begin{aligned} &{}^{+}F_1^{q}(q + i\epsilon, R, -T) \\ &{}^{+}F_1^{i}(q + i\epsilon, R, -T) \\ &{}^{-}F_1^{q}(q + i\epsilon, R, -T) \end{aligned}\right\} = \tfrac{1}{2}\pi e^{iqR} e^{i\omega T} - M(R, T). \tag{A10}$$

Let us next consider

$$L(-, +) = \int dk \frac{k e^{-i(kR+\omega_k T)}}{(k + q + i\epsilon)(k - q - i\epsilon)}. \tag{A11}$$

From Eq. (A5), we must require that $\alpha(+) < 0$. From Table I, for $T > 0$ and spacelike intervals, and for $T < 0$ and spacelike or timelike intervals, we must use the lower half of the complex k plane in the

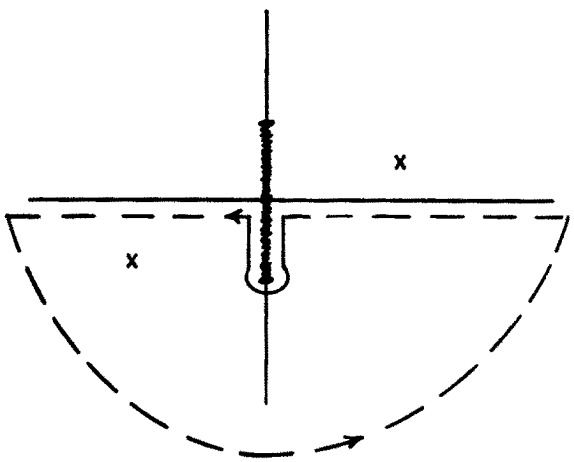


FIG. 6. Contour for $L(-, +)$ in the complex k plane.

construction of our contour. Applying the residue theorem to $L(-, +)$, we have

$$\int_C dk \frac{ke^{-i(kR+\omega_k T)}}{(k+q+i\epsilon)(k-q-i\epsilon)} = 2\pi i(\text{Res at } k=-q-i\epsilon), \quad \text{(A12)}$$

where C is shown in Fig. 6. Again the semicircular integrals vanish, and,on combining the integrals along the left and right hand side of the real axis, we have

$$\int_0^\infty dk \frac{ke^{-i\omega_k T}\sin(kR)}{(k+q+i\epsilon)(k-q-i\epsilon)} = \tfrac{1}{2}\pi e^{iqR}e^{-i\omega T} + M(R, T), \quad \text{(A13)}$$

where the first term on the right hand side of the equation represents the contribution of the residue at $k=-q-i\epsilon$ and $M(R, T)$, which is defined in Eq. (3.9), represents the contribution of the branch cut. We now have

$$\left.\begin{aligned} &{}^{+}F_1^{\sigma}(q+i\epsilon, R, T)\\ &{}^{-}F_1^{\sigma}(q+i\epsilon, R, T)\\ &{}^{-}F_1^{\tau}(q+i\epsilon, R, T)\end{aligned}\right\} = \tfrac{1}{2}\pi e^{iqR}e^{-i\omega T} + M(R, T). \quad \text{(A14)}$$

APPENDIX B

Using the techniques introduced in Appendix A, we were not able to determine an expression for ${}^{+}F_1^{\tau}(q+i\epsilon, R, T)$ and ${}^{-}F_1^{\tau}(q+i\epsilon, R, -T)$. Let us next consider the complex ω_k plane and express F_1 as an integral over ω_k:

$$F_1(q+i\epsilon, R, \pm T) = \int_m^\infty d\omega_k \frac{\omega_k e^{\mp i\omega_k T}\sin(kR)}{(\omega_k+\omega+i\epsilon)(\omega_k-\omega-i\epsilon)}, \quad \text{(B1)}$$

where $k=(\omega_k^2-m^2)^{\frac{1}{2}}$. Consider integrals of the form

$$J(\pm, \pm) = \int d\omega_k \frac{\omega_k e^{\pm i(kR\pm\omega_k T)}}{(\omega_k+\omega+i\epsilon)(\omega_k-\omega-i\epsilon)} \quad \text{(B2)}$$

with the integrand having simple poles at $\omega_k = -\omega-i\epsilon$, $\omega_k=\omega+i\epsilon$, and having a branch cut with branch points at $\omega_k=\pm m$. The contour to be considered in the evaluation of the integral will depend on the behavior of the exponential. Equation (A5) will again express the necessary asymptotic behavior for the exponentials. The sign of $\alpha(\pm)$ can be determined if we express k_I in terms of ω_{kR} and ω_{kI}. The relationship is

$$k_I = 2^{-\frac{1}{2}}\{[(\omega_{kR}^2-\omega_{kI}^2-m^2)^2+4\omega_{kR}^2\omega_{kI}^2]^{\frac{1}{2}} - [\omega_{kR}^2-\omega_{kI}^2-m^2]\}^{\frac{1}{2}}. \quad \text{(B3)}$$

Using Eq. (B3), we can construct Table II, which, for the complex ω_k plane, fulfills the same purpose as Table I did for the complex k plane. We are now ready to evaluate $F_1(q+i\epsilon, R, \pm T)$.

Let us consider

$$J(+, +) = \int d\omega_k \frac{\omega_k e^{+i(kR+\omega_k T)}}{(\omega_k+\omega+i\epsilon)(\omega_k-\omega-i\epsilon)}. \quad \text{(B4)}$$

From Eq. (A5), we must have $\alpha(+)>0$. From Table II, for $T<0$ and timelike intervals, we must use the lower half of the complex ω_k plane. Applying

TABLE II. Contours in the complex ω_k plane.

$\alpha(+)$	T	Space-Time interval	Contour
>0	>0	σ[a]	UHP[c]
>0	>0	τ[b]	UHP
>0	>0	σ	LHP[d]
<0	>0	τ	LHP
>0	<0	σ	UHP
<0	<0	τ	UHP
>0	<0	σ	LHP
>0	<0	τ	LHP
$\alpha(-)$	T	Space-Time interval	Contour
>0	>0	σ	UHP
<0	>0	τ	UHP
>0	>0	σ	LHP
>0	>0	τ	LHP
>0	<0	σ	UHP
>0	<0	τ	UHP
>0	<0	σ	LHP
<0	<0	τ	LHP

[a] σ refers to spacelike intervals.
[b] τ refers to timelike intervals.
[c] UHP refers to the upper half plane.
[d] LHP refers to the lower half plane.

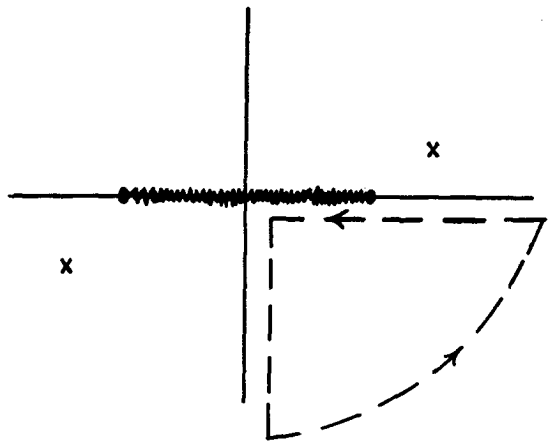


FIG. 7. Contour for $J(+,+)$ and $J(-,-)$ in the complex ω_k plane.

the residue theorem, we have

$$\int_C d\omega_k \frac{\omega_k e^{+i(kR+\omega_k T)}}{(\omega_k+\omega+i\epsilon)(\omega_k-\omega-i\epsilon)} = 0, \quad \text{(B5)}$$

where C is given in Fig. 7. The integral over the circular part of the contour again vanishes, and we have

$$\int_m^\infty d\omega_k \frac{\omega_k \exp[i(kR+\omega_k T)]}{(\omega_k+\omega+i\epsilon)(\omega_k-\omega-i\epsilon)} = \int_0^m dk \frac{k[\exp(-kR)]\exp i(m^2-k^2)^{\frac{1}{2}}T]}{k^2+q^2} + \int_0^\infty d\omega_k \frac{\omega_k[\exp(\omega_k T)]\exp[-(\omega_k^2+m^2)^{\frac{1}{2}}R]}{\omega_k^2+\omega^2}. \quad \text{(B6)}$$

Let us next consider

$$J(-,-) = \int d\omega_k \frac{\omega_k e^{-i(kR-\omega_k T)}}{(\omega_k+\omega+i\epsilon)(\omega_k-\omega-i\epsilon)}. \quad \text{(B7)}$$

From Eq. (A5), we must have $\alpha(-) < 0$. From Table II, for $T < 0$ and timelike intervals, we must use the lower half of the complex ω_k plane. Again applying Cauchy's theorem, we have

$$\int_C d\omega_k \frac{\omega_k e^{-i(kR-\omega_k T)}}{(\omega_k+\omega+i\epsilon)(\omega_k-\omega-i\epsilon)} = 0, \quad \text{(B8)}$$

where C is given in Fig. 7. The integral over the circular path vanishes, and we have

$$\int_m^\infty d\omega_k \frac{\omega_k \exp(-ikR)\exp(i\omega_k T)}{(\omega_k+\omega+i\epsilon)(\omega_k-\omega-i\epsilon)} = \int_0^m dk \frac{k\exp(kR)\exp[i(m^2-k^2)^{\frac{1}{2}}T]}{k^2+q^2} + \int_0^\infty d\omega_k \frac{\omega_k \exp(\omega_k T\exp)[(\omega_k^2+m^2)^{\frac{1}{2}}R]}{\omega_k^2+\omega^2}. \quad \text{(B9)}$$

Combining the results of Eqs. (B6) and (B9), we have

$$\int_m^\infty d\omega_k \frac{\omega_k e^{i\omega_k T}\sin(kR)}{(\omega_k+\omega+i\epsilon)(\omega_k-\omega-i\epsilon)} = {}^{-}F_1^r(q+i\epsilon, R, -T) = iW(R,T) + iN(R,T), \quad \text{(B10)}$$

where $N(R,T)$ and $W(R,T)$ are defined in Eqs. (3.12) and (3.11), respectively.

To derive an expression for ${}^{+}F_1^r(q+i\epsilon, R, T)$, we can apply similar techniques to $J(+,-)$ and $J(-,+)$. The result is

$${}^{+}F_1^r(q+i\epsilon, R, T) = iW(R,-T) + iN(R,-T), \quad \text{(B11)}$$

where $W(R,-T)$ and $N(R,-T)$ are again defined by Eqs. (3.11) and (3.12), respectively.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 2 FEBRUARY 1970

Green's Functions for Multidimensional Neutron Transport in a Slab

G. GARRETTSON* AND A. LEONARD
Department of Mechanical Engineering, Nuclear Division, Stanford University, Stanford, California 94305

(Received 23 April 1969)

The integral form of the one-speed, steady-state Boltzmann transport equation is solved for a point source in a homogeneous, isotropically scattering slab. In addition, solutions are obtained for line sources and plane sources in the slab, both normal and parallel to the slab faces. Using Fourier and Laplace transforms, the problem is reduced to that of solving a 1-dimensional integral equation with a difference kernel. This equation is transformed into a singular integral equation which is solved using standard methods. The Green's functions are subsequently obtained as generalized eigenfunction expansions over the spectrum of the 1-dimensional integral operator. This form yields a simple solution far from the source, and alternate expressions are obtained to facilitate evaluation near the source. In a thick slab the exact solutions are shown to reduce to simple closed expressions plus correction terms that decrease exponentially as the slab thickness increases. Most of the work previously done in multidimensional transport in slabs is shown to be easily reproduced using this theory in the thick-slab approximation. Also, virtually all other problems of this type can be solved using the theory presented here. In particular, the density from a pencil beam of particles normally incident to the slab is obtained.

1. INTRODUCTION

Until now, only a few problems in multidimensional transport theory have been treated analytically. Elliott[1] and Erdmann[2] give solutions for a point source in a half-space and two adjacent half-spaces, respectively, using the Wiener–Hopf technique, and Erdmann[3] iterates the half-space result to obtain an approximate solution to the point source in a slab. Starting with the integral transport equation, Smith[4] and Hunt[5] consider several problems of 2-dimensional transport in a 1-dimensional medium. Smith uses the Wiener–Hopf technique to solve for the isotropic scattering of radiation normally incident to a half-space and sinusoidally modulated in one transverse direction. Hunt considers radiation normally incident to a slab atmosphere and having radial symmetry with modulation $J_0(Br)$. He uses the replication property of the kernel to reduce the problem to that of solving the 1-dimensional integral equation which we treat in Sec. 4 by a different method.

Several attempts have been made to approximate the effect of finite transverse dimensions. Williams[6] and Kaper[7] use asymptotic theory ($e^{i\mathbf{B}\cdot\mathbf{r}}$ distribution) in the transverse direction and consider the resultant modified 1-dimensional transport equations. Williams treats the integral transport equation while Kaper considers the integro-differential equation. And in a semi-infinite slab ($z \in [0, \infty)$, $y \in [-y_0, y_0]$, and $x \in (-\infty, \infty)$), Smith and Hunt[8] solve the integral transport equation approximately as a series in powers of $e^{-\nu_0}$ by Laplace transforming in z and taking a Fourier cosine series in y.

To solve the transport equation in a finite prism with a point-isotropic source at its center, Boffi and Molinari[9] utilize the 3-dimensional Fourier transform. They obtain their solution as a spatial convolution of the point-source kernel with a function expressed as a triple summation of Legendre polynomials, whose coefficients are solutions to an infinite set of simultaneous algebraic equations.

A somewhat different approach to the problem of transport in 2- and 3-dimensional geometries has been developed by Gibbs.[10] Here the neutron density is shown generally to be expandable in a countable set of functions which satisfy a Helmholtz equation with continuous parameter. The expansion coefficients satisfy a set of coupled singular integral equations in the parameter. In his study of a particular quarter-space Milne problem, McCormick[11] uses Gibbs' approach to develop a coupled set of integral equations for the expansion coefficients.

Recently, Williams[12] has solved a simple 2-dimensional source problem in a 1-dimensional wide slab by Fourier transformation of the integral equation in the transverse direction. In the present paper, we determine the Green's function for all 2- or 3-dimensional source distributions in the 1-dimensional slab, i.e., we

* Present address: Physics Department, Naval Postgraduate School, Monterey, Calif. 93940.

[1] J. P. Elliott, Proc. Roy. Soc. (London) **A228**, 424 (1955).
[2] R. C. Erdmann, J. Math. Phys. **8**, 1040 (1967).
[3] L. B. Gratt and R. C. Erdmann, "The Wide Slab Flux from a Point Source of Neutrons," ANS Trans., 215 (June, 1969).
[4] M. G. Smith, Proc. Cambridge Phil. Soc. **60**, 909 (1964).
[5] G. E. Hunt, SIAM J. Appl. Math. **16**, 228 (1968).
[6] M. M. R. Williams, Nukleonik **9**, 305 (1967).
[7] Hans G. Kaper, J. Math. Phys. **10**, 286 (1969).
[8] M. G. Smith and G. E. Hunt, Proc. Cambridge Phil. Soc. **63**, 209 (1967).
[9] V. C. Boffi and V. Molinari, CNEN Report RT/FI(68) 43, Rome (1968).
[10] Alan G. Gibbs, J. Math. Phys. **10**, 875 (1969).
[11] S. I. Shreiner, N. J. McCormick, A. G. Gibbs, "Transport Solution of a Quarter-Space Problem," ANS Trans., 160 (June 1969).
[12] M. M. R. Williams, J. Math. Phys. **9**, 1873 (1968).

find the solution to a point uncollided source located at an arbitrary point in the slab. Using Fourier transforms in the transverse directions and a Laplace transform in the normal direction, we reduce the integral transport equation to a 1-dimensional integral equation with a difference kernel depending parametrically on the transform variables. A complete solution to this reduced equation is then obtained by the methods of Leonard and Mullikin.[13,14] Two alternate calculations of the Fourier inversion integral are given, one using contour integration and exploiting the singularities of the transform and the other evaluating analytically those parts of the transform that are not absolutely integrable. The former yields a simple solution for large transverse arguments, while the latter yields an expression which is easily evaluated numerically for small and intermediate arguments.

2. THE PROBLEM

Consider a homogeneous slab of thickness τ, which has a total cross section σ, and emits c secondaries per collision. The slab, infinite in both transverse directions, is surrounded by a vacuum or a pure absorber (see Fig. 1). Under the assumptions of steady-state, one-speed, and isotropic scattering, the integral equation for the neutron density $\rho(\mathbf{r})$ in the slab is

$$\rho(\mathbf{r}) = \frac{c\sigma}{4\pi}\int_0^\tau d\hat{x}\int_{-\infty}^{+\infty}\frac{e^{-\sigma|\mathbf{r}-\hat{\mathbf{r}}|}}{|\mathbf{r}-\hat{\mathbf{r}}|^2}\rho(\hat{\mathbf{r}})\,d\hat{y}\,d\hat{z} + S(\mathbf{r}), \quad (2.1)$$

with $0 \leq x \leq \tau$ and $-\infty < y, z < \infty$, where $\mathbf{r} = (x, y, z)$ and

$$|\mathbf{r}-\hat{\mathbf{r}}| = [(x-\hat{x})^2 + (y-\hat{y})^2 + (z-\hat{z})^2]^{\frac{1}{2}},$$

and where $S(\mathbf{r})$ is the uncollided neutron density.

Equation (2.1) is valid for any neutron-transport problem in the slab, including surface sources and/or volume sources. The uncollided neutron density $S(\mathbf{r})$ must first be obtained by considering the various neutron sources that are present and by using the methods in, e.g., Case, de Hoffman, and Placzek.[15] As an example, for a monodirectional pencil beam incident at $\mathbf{r} = 0$ in the direction $\mathbf{\Omega}_0(\theta_0, \varphi_0)$ (see Fig. 2), we have

$$S(\mathbf{r}) = e^{-\sigma x/\cos\theta_0}\delta(y - x\tan\theta_0\cos\varphi_0) \times \delta(z - x\tan\theta_0\sin\varphi_0). \quad (2.2)$$

To obtain the neutron density in a slab for any uncollided density $S(\mathbf{r})$, it is convenient to have the

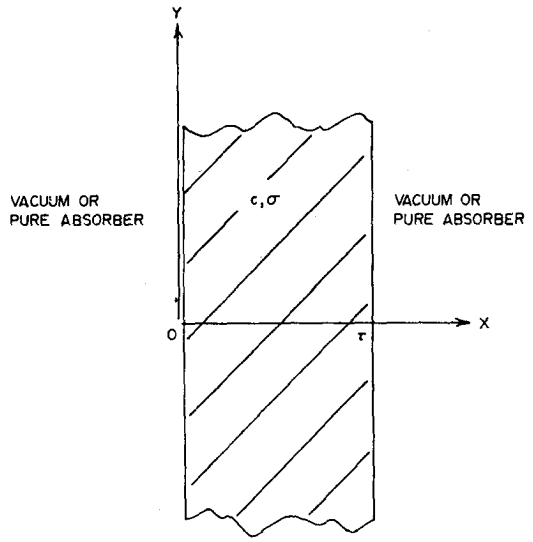


FIG. 1. A bare homogeneous slab.

Green's function $G(\mathbf{r}; \mathbf{r}')$, which satisfies

$$G(\mathbf{r}; \mathbf{r}') = \frac{c\sigma}{4\pi}\int_0^\tau d\hat{x}\int_{-\infty}^{+\infty}\frac{e^{-\sigma|\mathbf{r}-\hat{\mathbf{r}}|}}{|\mathbf{r}-\hat{\mathbf{r}}|^2}G(\hat{\mathbf{r}}; \mathbf{r}')\,d\hat{y}\,d\hat{z} + \delta^3(\mathbf{r}-\mathbf{r}'), \quad (2.3)$$

with $\mathbf{r}' = (x', y', z')$, $0 \leq x, x' \leq \tau$, and

$$-\infty < y', y, z', z < +\infty.$$

In Cartesian and cylindrical coordinates, the Dirac δ functions are

$$\delta^3(\mathbf{r}-\mathbf{r}') = \delta(x-x')\delta(y-y')\delta(z-z') = \delta^2(\tilde{\mathbf{r}}-\tilde{\mathbf{r}}')\delta(z-z'), \quad (2.4)$$

with

$$\delta^2(\tilde{\mathbf{r}}-\tilde{\mathbf{r}}') = [\delta(|\tilde{\mathbf{r}}-\tilde{\mathbf{r}}'|)/|\tilde{\mathbf{r}}-\tilde{\mathbf{r}}'|]\delta(\varphi-\varphi'),$$

where $\tilde{\mathbf{r}} = (\tilde{r}, \varphi)$ in cylindrical coordinates, $\tilde{\mathbf{r}} = (y, z)$ in Cartesian coordinates, and where $|\tilde{\mathbf{r}}-\tilde{\mathbf{r}}'| = [(y-y')^2 + (z-z')^2]^{\frac{1}{2}}$. Then the solution to (2.1) is given formally by

$$\rho(\mathbf{r}) = \int_0^\tau dx'\int_{-\infty}^{+\infty}G(\mathbf{r}; \mathbf{r}')S(\mathbf{r}')\,dy'\,dz', \quad (2.5)$$

and the angular neutron density can easily be obtained from $\rho(\mathbf{r})$ by quadrature.[15]

Although the point-uncollided density in (2.3) is mathematically convenient, it is not physically realizable. Therefore, some people prefer to work with a point-isotropic source. Under the assumptions of isotropic scattering, there is a simple relationship between the point-isotropic-source Green's function $G_{\text{iso}}(\mathbf{r}; \mathbf{r}')$ and the point-uncollided-source Green's function $G(\mathbf{r}; \mathbf{r}')$:

$$G(\mathbf{r}; \mathbf{r}') = c\sigma G_{\text{iso}}(\mathbf{r}; \mathbf{r}') + \delta^3(\mathbf{r}-\mathbf{r}'), \quad (2.6)$$

[13] A. Leonard and T. W. Mullikin, Trans. A.M.S. **116**, 465 (1965).
[14] A. Leonard and T. W. Mullikin, J. Math. Phys. **44**, 327 (1965).
[15] K. M. Case, F. de Hoffmann, and G. Placzek, *Introduction to the Theory of Neutron Diffusion* (U.S. Government Printing Office, Washington, D.C., 1953).

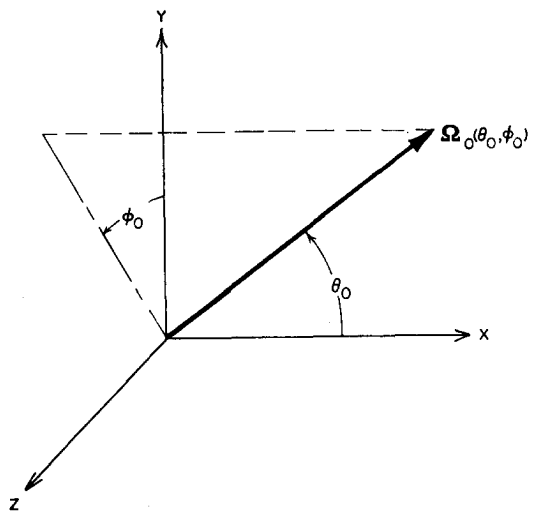


FIG. 2. The unit vector $\Omega(\theta_0, \phi_0)$.

where $G_{iso}(\mathbf{r}; \mathbf{r}')$ satisfies (2.1) with $S(\mathbf{r}) = e^{-\sigma|\mathbf{r}-\mathbf{r}'|}/4\pi|\mathbf{r}-\mathbf{r}'|^2$, which is the uncollided density at $\mathbf{r}$ from a unit point-isotropic source at $\mathbf{r}'$. The once-collided density at $\mathbf{r}$ from a point-uncollided source at $\mathbf{r}'$ is simply $c\sigma(e^{-\sigma|\mathbf{r}-\mathbf{r}'|}/4\pi|\mathbf{r}-\mathbf{r}'|^2)$.

3. TRANSFORMATION OF THE GREEN'S FUNCTION

One approach to solving (2.3) is to write a Neumann series solution

$$G(\mathbf{r}; \mathbf{r}') = (1 + \Delta + \Delta^2 + \cdots)\delta^3(\mathbf{r} - \mathbf{r}'), \quad (3.1)$$

where

$$\Delta[\cdot] = \frac{c\sigma}{4\pi}\int_0^\tau d\hat{x}\int_{-\infty}^{+\infty}\int \frac{e^{-\sigma|\mathbf{r}-\hat{\mathbf{r}}|}}{|\mathbf{r}-\hat{\mathbf{r}}|^2}[\cdot]\,d\hat{y}\,d\hat{z}. \quad (3.2)$$

However, if we take the norm of Δ to be

$$\|\Delta\| = \max_{\substack{0\le x\le\tau \\ -\infty<y,z<+\infty}} \frac{c\sigma}{4\pi}\int_0^\tau d\hat{x}\int_{-\infty}^{+\infty}\int \frac{e^{-\sigma|\mathbf{r}-\hat{\mathbf{r}}|}}{[\mathbf{r}-\hat{\mathbf{r}}]}\,d\hat{y}\,d\hat{z}, \quad (3.3)$$

then $\|\Delta\| = c[1 - E_2(\frac{1}{2}\sigma\tau)]$, where

$$E_n(q) = \int_1^\infty e^{-qu}\frac{du}{u^n}, \quad n = 1, 2, \cdots. \quad (3.4)$$

[See Ref. 15 for numerical tabulations of $E_n(q)$.] So, unless c or $\sigma\tau$ is very small, the Neumann series converges slowly. Therefore, it is desirable to solve (2.3) by other means.

To do so, first note that, if $g(x, y, z; x')$ satisfies

$$g(x, y, z; x') = \Delta[g](x, y, z; x') + \delta(x - x')\delta(y)\delta(z), \quad (3.5)$$

for $0 \le x, x' \le \tau$ and $-\infty < y, z < +\infty$, then

$$G(\mathbf{r}; \mathbf{r}') = g(x, y - y', z - z'; x'), \quad (3.6)$$

for $0 \le x, x' \le \tau$ and $-\infty < y, y', z, z' < +\infty$. To obtain $g(x, y, z; x')$, we first transform (3.5) into a 1-dimensional integral equation with a difference kernel, solve that equation, and then invert the transform.

Now we define the 2-dimensional Fourier transform in y and z (or, equivalently, the Hankel transform in $\tilde{r}$):

$$\mathfrak{G}(x, B; x') = \int_{-\infty}^{+\infty}\int g(x, y, z; x') \times \exp(-i\omega_y y - i\omega_z z)\,dy\,dz, \quad (3.7)$$

where $B = (\omega_y^2 + \omega_z^2)^{\frac{1}{2}}$. Then its inverse is

$$g(x, y, z; x') = \frac{1}{(2\pi)^2}\int_{-\infty}^{+\infty}\int \mathfrak{G}(x, B; x')\exp(i\omega_y y + i\omega_z z)\,d\omega_y\,d\omega_z$$

$$= \frac{1}{2\pi}\int_0^\infty \mathfrak{G}(x, B; x')J_0(B\tilde{r})B\,dB, \quad (3.8)$$

where $\tilde{r} = |\tilde{\mathbf{r}}| = (y^2 + z^2)^{\frac{1}{2}}$, and where J_0 is the ordinary Bessel function.[16] The B dependence of $\mathfrak{G}(x, B; x')$ will become evident from the equation it satisfies.

Taking the 2-dimensional Fourier transform of Eq. (3.5), we have

$$\mathfrak{G}(x, B; x') = \tfrac{1}{2}c\sigma\int_0^\tau K(|x - \hat{x}|; B^2)\mathfrak{G}(\hat{x}, B; x')\,d\hat{x} + \delta(x - x'), \quad (3.9)$$

where[4,5]

$$K(|x - \hat{x}|; B^2) = \int_\sigma^\infty \exp[-|x - \hat{x}|(B^2 + \eta^2)^{\frac{1}{2}}]\frac{d\eta}{(B^2 + \eta^2)^{\frac{1}{2}}}, \quad (3.10)$$

or

$$K(|x - \hat{x}|; B^2) = \int_0^{\alpha(B)} \frac{\exp(-|x - \hat{x}|/s)}{s(1 - s^2B^2)^{\frac{1}{2}}}\,ds, \quad (3.11)$$

in which $\alpha(B) = (B^2 + \sigma^2)^{-\frac{1}{2}}$. Note that $K(|x|; 0) = E_1(\sigma|x|)$, the familiar plane-source kernel. Defining the integral operator

$$\Delta_B[\cdot] = \tfrac{1}{2}c\sigma\int_0^\tau K(|x - \hat{x}|; B^2)[\cdot]\,d\hat{x} \quad (3.12)$$

will simplify the notation.

[16] *Handbook of Mathematical Functions*, M. Abramowitz and Irene Stegers, Eds. (U.S. Government Printing Office, Washington, D.C., 1964).

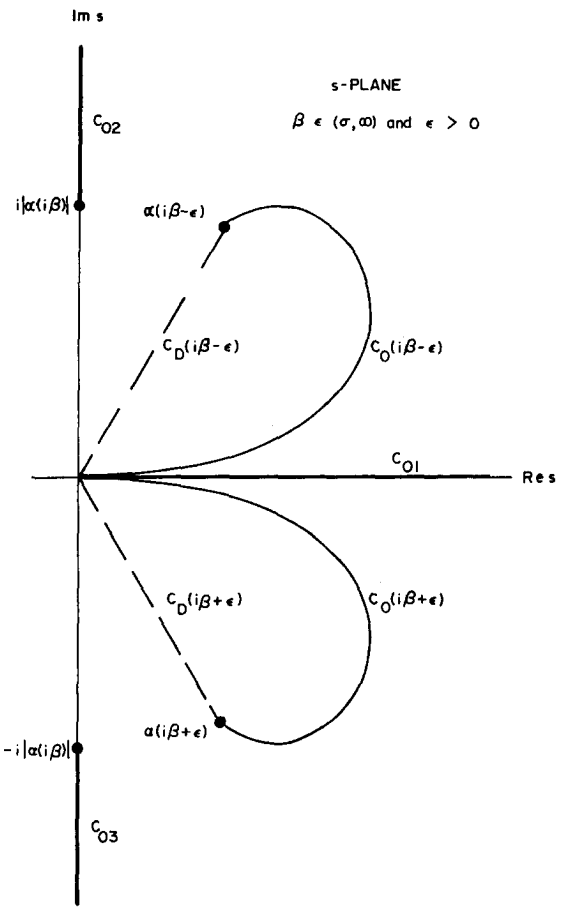


FIG. 3. The cut $C_0(B)$ and the deformed cut $C_D(B)$ for $B \notin C_+ \cup C_-$; the limit cuts of $C_0(B)$, $C_{01} \cup C_{02}$ and $C_{01} \cup C_{03}$, for $B \to i\beta \in C_+ \cup C_-$.

Both $\alpha(B)$ and $K(|x|; B^2)$ are analytic in the complex B plane cut by $C_+ \cup C_-$, where $C_\pm = \{\pm i\beta\colon \beta \in (\sigma, \infty)\}$. For complex B, Eq. (3.11), and all succeeding integrals with limits 0 to $\alpha(B)$, are to be interpreted as contour integrals along $C_0(B) = \{(B^2 + \eta^2)^{-\frac{1}{2}}\colon \eta \in (\infty, \sigma)\}$. For $B \notin C_+ \cup C_-$, $C_0(B)$ may be deformed to the straight-line contour between 0 and $\alpha(B)$, $C_D(B) = \{t \cdot \alpha(B)\colon t \in (0, 1)\}$, as shown in Fig. 3. However, in the limit $B \to i\beta \in C_+ \cup C_-$,

$$\lim_{\substack{\epsilon\to 0\\ \epsilon>0}} C_0(i\beta - \epsilon) = \lim_{\substack{\epsilon\to 0\\ \epsilon>0}} C_0(-i\beta + \epsilon) = C_{01} \cup C_{02}$$

and

$$\lim_{\substack{\epsilon\to 0\\ \epsilon>0}} C_0(i\beta + \epsilon) = \lim_{\substack{\epsilon\to 0\\ \epsilon>0}} C_0(-i\beta - \epsilon) = C_{01} \cup C_{03},$$

where $\beta \in (\sigma, \infty)$ and where $C_{01} = (0, \infty]$, $C_{02} = \{it\colon t \in [\infty, |\alpha(i\beta)|)\}$, and $C_{03} = \{it\colon -t \in [\infty, |\alpha(i\beta)|)\}$.

To finally obtain an integral equation that one can solve, we define the Laplace transform

$$F(x, B; \gamma) = \int_{-\infty}^{+\infty} \mathcal{G}(x, B; x')e^{-x'\gamma}\, dx' \quad (3.13)$$

and its inverse

$$\mathcal{G}(x, B; x') = \frac{1}{2\pi i}\int_{-i\infty}^{+i\infty} F(x, B; \gamma)e^{x'\gamma}\, d\gamma, \quad (3.14)$$

and then we transform (3.9) to obtain

$$F(x, B; \gamma) = \Delta_B[F](x, B; \gamma) + e^{-x\gamma}. \quad (3.15)$$

Using (3.14), (3.8), and (3.6), we can write the inversion formula for the Green's function in terms of the solution to (3.15):

$$G(\mathbf{r}; \mathbf{r}') = \frac{1}{4\pi^2 i}\int_0^\infty \int_{-i\infty}^{+i\infty} F(x, B; \gamma) \times e^{\gamma x'} J_0(B\,|\tilde{\mathbf{r}} - \tilde{\mathbf{r}}'|)B\, d\gamma\, dB. \quad (3.16)$$

Equation (3.15) will be solved treating B and γ as complex parameters. Then the Fourier and Laplace inversions (3.16) may be carried out in either order to yield the Green's function. In Sec. 7, we demonstrate that the result at each step of the inversion can be physically interpreted as the solution to a relevant multidimensional neutron-transport problem in a slab.

4. SOLUTION OF (3.15)

The Neumann series solution to (3.15),

$$F(x, B; \gamma) = (1 + \Delta_B + \Delta_B^2 + \cdots)e^{-x\gamma}, \quad (4.1)$$

converges slowly except for small c or $\sigma\tau$ or large B. However, one can use the methods of Leonard and Mullikin[13,14] to solve (3.15). By transforming (3.15) into a singular integral equation in γ, one obtains the solution to $F(x, B; \gamma)$ in terms of solutions to other Fredholm integral equations whose Neumann series solutions converge rapidly, at least for B not near $C_+ \cup C_-$.

For these purposes it is convenient to consider the auxiliary functions $f_{(\pm)}(x, B; \zeta)$ which satisfy

$$f_{(\pm)}(x, B; \zeta) = \Delta_B[f_{(\pm)}](x, B; \zeta) + S_{(\pm)}(x, \zeta), \quad (4.2)$$

with $S_{(\pm)}(x, \zeta) = e^{-x/\zeta} + (\pm)e^{-(\tau-x)/\zeta}$. Then,

$$F(x, B; \gamma) = \tfrac{1}{2}\{f_+(x, B; \gamma^{-1}) + f_-(x, B; \gamma^{-1})\}. \quad (4.3)$$

From (4.2) we have the formal solution

$$f_{(\pm)}(x, B; \zeta) = (I - \Delta_B)^{-1}[S_{(\pm)}](x, B; \zeta), \quad (4.4)$$

where $I[\cdot]$ is the identity operator and where $(I - \Delta_B)^{-1}[\cdot]$ is the inverse of the operator $I[\cdot] - \Delta_B[\cdot]$.

Operating on $S_{(\pm)}(x, \zeta)$ with $\Delta_B[\cdot]$, we obtain

$$\begin{aligned}\Delta_B[S_{(\pm)}](x, B; \zeta) &= \tfrac{1}{2}c\sigma\int_0^\tau K(|x-\hat{x}|; B^2)[e^{-\hat{x}/\zeta} + (\pm)e^{-(\tau-\hat{x})/\zeta}]\,d\hat{x} \\ &= \tfrac{1}{2}c\sigma\zeta\Big[S_{(\pm)}(x,\zeta)\int_0^{\alpha(B)}\Big(\frac{1}{s+\zeta} - \frac{1}{s-\zeta}\Big)\frac{ds}{(1-B^2s^2)^{\frac{1}{2}}} \\ &\quad + \int_0^{\alpha(B)}\frac{S_{(\pm)}(x,s)}{s-\zeta}\frac{ds}{(1-B^2s^2)^{\frac{1}{2}}} \\ &\quad - (\pm)e^{-\tau/\zeta}\int_0^{\alpha(B)}\frac{S_{(\pm)}(x,s)}{s+\zeta}\frac{ds}{(1-B^2s^2)^{\frac{1}{2}}}\Big]. \qquad (4.5)\end{aligned}$$

Thus, we can write the singular integral equation satisfied by $f_{(\pm)}$:

$$\begin{aligned}f_{(\pm)}(x, B; \zeta) &= (I-\Delta_B)^{-1}\{S_{(\pm)} - \Delta_B[S_{(\pm)}] + \Delta_B[S_{(\pm)}]\}(x, B; \zeta) \\ &= S_{(\pm)}(x,\zeta) + (I-\Delta_B)^{-1}\{\Delta_B[S_{(\pm)}]\}(x, B; \zeta) \\ &= S_{(\pm)}(x,\zeta) + f_{(\pm)}(x, B; \zeta)\tfrac{1}{2}c\xi(\zeta, B)\log\frac{\xi(\zeta, B)+1}{\xi(\zeta, B)-1} \\ &\quad + \tfrac{1}{2}c\sigma\zeta\int_0^{\alpha(B)}\frac{f_{(\pm)}(x, B; s)}{s-\zeta}\frac{ds}{(1-B^2s^2)^{\frac{1}{2}}} \\ &\quad - (\pm)e^{-\tau/\zeta}\tfrac{1}{2}c\sigma\zeta\int_0^{\alpha(B)}\frac{f_{(\pm)}(x, B; s)}{s+\zeta}\frac{ds}{(1-B^2s^2)^{\frac{1}{2}}}. \qquad (4.6)\end{aligned}$$

Hence, $F(x, B; \gamma)$ satisfies the following singular integral equation:

$$\begin{aligned}\Lambda\Big(\xi\Big(\frac{1}{\gamma}, B\Big)\Big)F(x, B; \gamma) &= e^{-x\gamma} + \tfrac{1}{2}c\sigma\int_0^{\alpha(B)}\frac{F(x, B; s^{-1})}{s\gamma - 1}\frac{ds}{(1-B^2s^2)^{\frac{1}{2}}} \\ &\quad - e^{-\tau\gamma}\tfrac{1}{2}c\sigma\int_0^{\alpha(B)}\frac{F(\tau - x, B; s^{-1})}{s\gamma + 1}\frac{ds}{(1-B^2s^2)^{\frac{1}{2}}}, \qquad (4.7)\end{aligned}$$

where

$$\Lambda(\xi) = 1 - \tfrac{1}{2}c\xi\log\frac{\xi+1}{\xi-1}, \qquad (4.8)$$

and

$$\xi(\zeta, B) = \sigma\zeta(1-B^2\zeta^2)^{-\frac{1}{2}}. \qquad (4.9)$$

$\xi(\zeta, B) \in (0, 1)$ for $\zeta \in C_0(B)$, for all B in the complex plane cut by $C_+ \cup C_-$, and $\Lambda(\xi(\zeta, B))$ is analytic in the ζ plane cut by $C_0(B) \cup C_0^-(B)$, where $C_0^-(B) = \{-\zeta: \zeta \in C_0(B)\}$, and has two zeros $\pm\zeta_0$.

In the ζ plane, $f_{(\pm)}(x, B; \zeta)$ is analytic for all $\zeta \neq 0$ and approaches a definite limit for $\zeta \to 0$ in the right half-plane. In the B plane, $f_{(\pm)}(x, B; \zeta)$ has cuts $C_+ \cup C_-$, and poles where $\Delta_B[\cdot]$ has an eigenvalue equal to one, so that the inverse of $(I - \Delta_B)[\cdot]$ does not exist.

Treating B and γ as complex parameters, Eqs. (4.6) or (4.7) can be solved using the methods of Muskhelishvili[17] to yield the solution

$$\begin{aligned}f_{(\pm)}(x, B; \zeta) &= \frac{h_{1(\pm)}(x, B; \zeta) + (\pm)h_{1(\pm)}(x, B; -\zeta)e^{-\tau/\zeta} - S_{(\pm)}(x,\zeta)}{\Lambda(\xi(\zeta, B))} \\ &\quad + C_{(\pm)}(x, B)\frac{h_{2(\pm)}(B; \zeta) + (\pm)h_{2(\pm)}(B; -\zeta)e^{-\tau/\zeta}}{\Lambda(\xi(\zeta, B))}, \qquad (4.10)\end{aligned}$$

where

$$-C_{(\pm)}(x, B) = \frac{h_{1(\pm)}(x, B; \zeta_0) + (\pm)h_{1(\pm)}(x, B; -\zeta_0)e^{-\tau/\zeta_0} - S_{(\pm)}(x, \zeta_0)}{h_{2(\pm)}(B; \zeta_0) + (\pm)h_{2(\pm)}(B; -\zeta_0)e^{-\tau/\zeta_0}}. \qquad (4.11)$$

The zeros of Λ in the ζ plane are $\pm\zeta_0$:

$$\zeta_0 = \zeta_0(B) = [B^2 + (\sigma/\nu_0)^2]^{-\frac{1}{2}} \qquad (4.12)$$

and

$$\begin{aligned}\Lambda(\xi(\pm\zeta_0, B)) = \Lambda(\pm\nu_0) &= 1 - c\nu_0\tanh^{-1}(1/\nu_0) \\ &= 0, \qquad (4.13)\end{aligned}$$

where $\nu_0 \in [1, \infty]$ for $c \in [0, 1]$ and $-i\nu_0 \in (0, \infty]$ for $c \in [1, \infty)$. $h_{1(\pm)}(x, B; \zeta)$ and $h_{2(\pm)}(B; \zeta)$ satisfy the Fredholm integral equations

$$h_{1(\pm)}(x, B; \zeta) = -(\pm)\mathfrak{L}[h_{1(\pm)}](x, B; \zeta) + S_{(\pm)}(x, \zeta) \qquad (4.14a)$$

and

$$h_{2(\pm)}(B; \zeta) = -(\pm)\mathfrak{L}[h_{2(\pm)}](B; \zeta) + e^{-\tau/\zeta}\zeta X(-\zeta, B), \qquad (4.14b)$$

where the integral operator $\mathfrak{L}[\cdot]$ is defined as follows:

$$\begin{aligned}&\mathfrak{L}[h_{1(\pm)}](x, B; \zeta) \\ &= \int_0^{\alpha(B)}\frac{\mathcal{K}(B, \zeta; s)}{s+\zeta}h_{1(\pm)}(x, B; s)\frac{ds}{(1-B^2s^2)^{\frac{1}{2}}}, \qquad (4.15)\end{aligned}$$

[17] N. I. Muskhelishvili, *Singular Integral Equations* (P. Noordhoff Ltd., Groningen, The Netherlands, 1953).

in which

$$\mathcal{K}(B, \zeta; s) = \tfrac{1}{2}c\sigma e^{-\tau/\zeta}\zeta X(-\zeta, B)\frac{X(-s, B)[\zeta_0^2 - s^2]}{\zeta_0^2 M(\xi(s, B))}, \tag{4.16}$$

$$M(\xi) = [\lambda(\xi)]^2 + [\tfrac{1}{2}c\pi\xi]^2, \tag{4.17}$$

$$X(\zeta, B) = \frac{\alpha(B)}{\alpha(B) - \zeta}\exp\left[\frac{1}{\pi}\int_0^{\alpha(B)}\left(\frac{1}{t - \zeta} - \frac{1}{t}\right)\theta(\xi(t, B))\,dt\right], \tag{4.18}$$

$$\theta(\xi) = \tan^{-1}(\tfrac{1}{2}c\pi\xi/\lambda(\xi)), \tag{4.19}$$

and

$$\lambda(\xi) = 1 - \tfrac{1}{2}c\xi\log[(1 + \xi)/(1 - \xi)]. \tag{4.20}$$

$X(\zeta, B)$ is constructed to be analytic in the ζ plane cut by $C_0(B)$, for all B in the B plane cut by $C_+ \cup C_-$, and to have no zeros or poles. $\theta(\xi) \in (0, \pi)$ for $\xi \in (0, 1)$, and $\xi(\zeta, B) \in (0, 1)$ for $\zeta \in C_0(B)$ [deformed to $C_D(B)$ for $B \notin C_+ \cup C_-$].

It is useful to note that $h_{1(\pm)}(x, B; \zeta)$ can be expressed in terms of $h_{2(\pm)}(B; \zeta)$ (see Ref. 14):

$$h_{1(\pm)}(x, B; \zeta) = S_{(\pm)}(x, \zeta) - (\pm)\frac{c\sigma}{2\zeta_0^2}\int_0^{\alpha(B)}\left(\frac{h_{2(\pm)}(B; t)\zeta h_{2(\mp)}(B; \zeta) - h_{2(\pm)}(B; \zeta)t h_{2(\mp)}(B; t)}{tM(\xi(t, B))(1 - B^2t^2)^{\frac{1}{2}}} \times e^{\tau/t}\frac{\zeta_0^2 - t^2}{\zeta^2 - t^2}S_{(\pm)}(x, t)\right)dt. \tag{4.21}$$

And it is often more convenient to write

$$h_{2(\pm)}(B; \zeta) = e^{-\tau/\zeta}\zeta X(-\zeta, B)H_{(\pm)}(B; \zeta). \tag{4.22}$$

Then, $H_{(\pm)}(B; \zeta)$ satisfies the Fredholm equation

$$H_{(\pm)}(B; \zeta) = -(\pm)L[H_{(\pm)}](B; \zeta) + 1, \tag{4.23a}$$

where

$$L[H_{(\pm)}](B; \zeta) = \int_0^{\alpha(B)}\frac{\mathcal{K}(B, s; s)}{s + \zeta}H_{(\pm)}(B; s)\frac{ds}{(1 - B^2s^2)^{\frac{1}{2}}}. \tag{4.23b}$$

In general, the Neumann series solutions to (4.14) and (4.23) converge rapidly. For

$$B \in C_p = (-\infty, +\infty) \cup \{i\beta: \beta \in (-\sigma, \sigma)\},$$

it can be shown that the norms

$$\|\mathfrak{L}\| = \max_{\zeta\in(0,\alpha(B))}\int_0^{\alpha(B)}\left|\frac{\mathcal{K}(B, \zeta; s)}{s + \zeta}\frac{1}{(1 - B^2s^2)^{\frac{1}{2}}}\right|ds \leq e^{-\tau/\alpha(B)} \tag{4.24a}$$

and

$$\|L\| = \max_{\zeta\in(0,\alpha(B))}\int_0^{\alpha(B)}\left|\frac{\mathcal{K}(B; s; s)}{s + \zeta}\frac{1}{(1 - B^2s^2)^{\frac{1}{2}}}\right|ds \leq \frac{e^{-\tau/\alpha(B)}}{X(-\alpha(B), B)}, \tag{4.24b}$$

so that for many applications the solutions to (4.12) and (4.23) are given with sufficient accuracy by the first few terms in the Neumann series. Furthermore, the truncated series need only be evaluated for $\zeta \in C_0(B)$, deformed to $C_D(B) = \{p \cdot \alpha(B): p \in (0, 1)\}$ for $B \notin C_+ \cup C_-$, and these results may be used in (4.14) and (4.23) to obtain $h_{1(\pm)}(x, B; \zeta)$, $h_{2(\pm)}(B; \zeta)$, and $H_{(\pm)}(B; \zeta)$ for all other ζ. And as shown above, it is only necessary to evaluate the Neumann series solution for $h_{2(\pm)}(B; \zeta)$ or for $H_{(\pm)}(B, \zeta)$; this result may be used in (4.21) to yield $h_{1(\pm)}(x, B; \zeta)$.

In particular, for $\tau \gg \alpha(B)$, $B \in C_p$, the Neumann series solution to (4.23) or to (4.14b) may be truncated to the first term, yielding

$$H_{(\pm)}(B; \zeta) = 1 + O(e^{-\tau/\alpha(B)}) \tag{4.25a}$$

or

$$h_{2(\pm)}(B; \zeta) = e^{-\tau/\zeta}\zeta X(-\zeta, B)[1 + O(e^{-\tau/\alpha(B)})], \tag{4.25b}$$

respectively. This result, substituted into (4.21), yields

$$h_{1(\pm)}(x, B; \zeta) = S_{(\pm)}(x, \zeta) - (\pm)\mathfrak{L}[S_{(\pm)}](x, B; \zeta) + O(e^{-\tau/\alpha(B)}), \tag{4.25c}$$

which is just the Neumann series solution to (4.14a) truncated to the second term. Substituting (4.25) into (4.10), (4.11), and (4.3) results in a relatively simple expression for $F(x, B; \gamma)$, which is easy to evaluate numerically and which is "exact" for most practical cases.

It is interesting to note that setting $B = 0$ gives the results and the corresponding functions for 1-dimensional integral transport theory in a slab.[13,14]

5. THE TRANSFORM INVERSIONS USING CONTOUR INTEGRATION: A CONVENIENT SOLUTION FOR $|\tilde{\mathbf{r}} - \tilde{\mathbf{r}}'| \gg 1/\sigma$

To evaluate $G(\mathbf{r}; \mathbf{r}')$, the results of Sec. 4 are substituted into the inversion formula (3.16) and the integrations over B and γ are performed. These integrals can both be evaluated using contour integration and the calculus of residues, yielding a relatively simple result for large radial arguments $|\tilde{\mathbf{r}} - \tilde{\mathbf{r}}'| \gg 1/\sigma$.

Suppose we define the inverse Fourier (Hankel) transform of $F(x, B; \gamma)$ as

$$\rho_3(x, y, z; \gamma) = \frac{1}{2\pi}\int_0^\infty F(x, B; \gamma)J_0(B\tilde{r})B\,dB, \quad (5.1)$$

where $\tilde{r} = (y^2 + z^2)^{\frac{1}{2}}$ and γ is treated as a complex parameter. Before evaluating (5.1) using contour integration, we need to examine further the analyticity of $F(x, B; \gamma)$ in the B plane.

A spectral analysis[18] of the operator $\Delta_B[\cdot]$ in the B plane, i.e., an investigation of the singularities of $(I - \Delta_B)^{-1}[\cdot]$, reveals a continuous spectrum

$$C_+ \cup C_- = \{\pm i\beta: \beta \in (\sigma, \infty)\}$$

and a point spectrum restricted to

$$C_p' = (-\infty, +\infty) \cup \{i\beta: \beta \in [-\sigma, \sigma]\}.$$

For $\tau > 0$ sufficiently small, the point spectrum is empty. In other words, $F(x, B; \gamma)$ is analytic in the B plane cut by $C_+ \cup C_-$ with poles in C_p', and there exists a $\tau_{\min} > 0$ such that, for $0 < \tau < \tau_{\min}$, there are no poles. In particular, $\tau_{\min}$ satisfies (5.2) (below) implicitly with $n = 1$ and $B_n = i\sigma$, and Williams[12] obtains approximate values of $\tau_{\min}$ for $0 \le c < 1$, by neglecting the last (nonlinear) term, $\mathfrak{U}(i\sigma, \tau_{\min})$.

Consider first the point spectrum. A detailed investigation of the denominator of $F(x, B, \gamma)$ reveals that $F(x, B, \gamma)$ has a finite number of poles for $0 < \tau < \infty$, namely, those $\pm B_n \in C_p'$, $n = 1, 2, \cdots, N$, for which the denominator of (4.11) is zero. Equivalently, these poles satisfy

$$\tau = T(B_n, n) - \mathfrak{U}(B_n, \tau), \quad n = 1, 2, \cdots, N, \quad (5.2)$$

where

$$T(B, n) = n\pi i\zeta_0(B) - 2\int_0^1 \frac{1 - \theta(t)/\pi}{1 - (t/\nu_0)^2}\frac{dt}{(\sigma^2 + B^2t^2)^{\frac{1}{2}}} \quad (5.3)$$

and

$$\mathfrak{U}(B, \tau) = \zeta_0(B)\log[H_{(\pm)}(B; -\zeta_0)/H_{(\pm)}(B; \zeta_0)], \quad (5.4)$$

in which $(\pm) = \text{sign}[(-1)^{n+1}]$, $n = 1, 2, \cdots, N$.

Specifically, for $0 \le c \le 1$, $F(x, B; \gamma)$ has poles at $\pm B_n$, $n = 1, 2, \cdots, N$: $B_n = i\beta_n$ and $\sigma/\nu_0 < \beta_1 < \beta_2 < \cdots < \beta_N \le \sigma$. For $c \ge 1$, $F(x, B; \gamma)$ has poles at $\pm B_n$, $n = 1, 2, \cdots, N$: $\sigma/|\nu_0| > B_1 > B_2 > \cdots > B_{n_0-1} \ge 0$, and $B_n = i\beta_n$, $n = n_0, n_0 + 1, \cdots, N$, with $0 \le \beta_{n_0} < \beta_{n_0+1} < \cdots < \beta_N \le \sigma$. And for a subcritical slab, $F(x, B; \gamma)$ has poles $\pm B_n$, $n = 1, 2, \cdots, N$: $B_n = i\beta_n$, $0 < \beta_1 < \beta_2 < \cdots < \beta_N \le \sigma$. Since $\mathfrak{U}(B, \tau) = O(e^{-\tau/\alpha(B)})$, $\mathfrak{U}$ is important in (5.2) only for those poles near $\pm i\sigma$. The remaining poles can quickly be determined by plotting $T(B, n)$ versus $B \in C_p$ for $n = 1, 2, \cdots$, and then by picking off those $\hat{B}_n \approx B_n$ for which $\tau = T(\hat{B}_n, n)$. For an illustration, the general shape of $T(B, n)$ is sketched in Fig. 4 for $\nu_0 = 5$. Note that the first term of $T(B, n)$ dominates for β near σ/ν_0, so that to a fair approximation, the first few poles of $F(x, B; \gamma)$ are given by

$$\beta_n \approx [(\sigma/\nu_0)^2 + (n\pi/\tau)^2]^{\frac{1}{2}}, \quad 1 \le n < N. \quad (5.5)$$

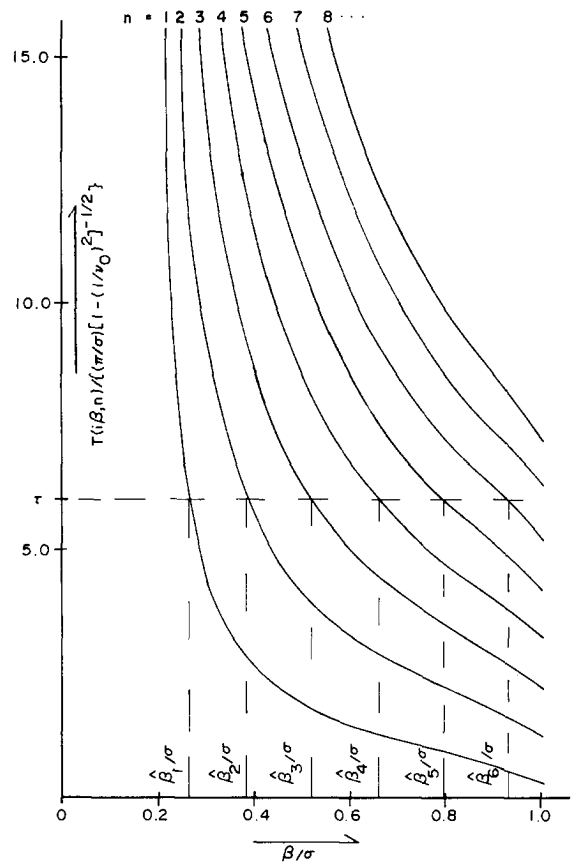


FIG. 4. A sketch of $T(i\beta, n)$ vs $\beta \in (\sigma/\nu_0, \sigma]$ for $\nu_0 = 5$, graphically determining the wide slab poles.

One further consideration is necessary before we can use contour integration to evaluate (5.1). Since

$$F(x, B, \gamma) \xrightarrow[B\to\infty]{} e^{-x\gamma} + O(e^{-\frac{1}{2}B\tau}/B^2), \quad 0 < x < \tau,$$

$$e^{-x\gamma} + O(1/B^2), \quad x = 0 \text{ or } \tau, \quad (5.6a)$$

and

$$J_0(B\tilde{r}) \xrightarrow[B\to\infty]{} O((B\tilde{r})^{-\frac{1}{2}}), \quad (5.6b)$$

part of $F(x, B; \gamma)J_0(B\tilde{r})B$ is not integrable on $B \in (0, \infty)$. This part of (5.1) exists in the sense of a

[18] J. Lehner and G. M. Wing, Commun Pure Appl. Math. 8, 217 (1955).

generalized function and must be evaluated as such. For example,

$$\frac{1}{2\pi}\int_0^\infty J_0(B\tilde{r})B\,dB = \delta(y)\delta(z). \tag{5.7}$$

Thus, it is convenient to define

$$F_0(x, B; \gamma) = F(x, B; \gamma) - [1 + (c\sigma/B)\tan^{-1}(B/\sigma)]e^{-x\gamma}. \tag{5.8}$$

Then,

$$F_0(x, B; \gamma)J_0(B\tilde{r})B \to O((B^3\tilde{r})^{-\frac{1}{2}}), \tag{5.9}$$

which is integrable on $B \in (0, \infty)$, and we can write

$$\rho_3(x, y, z; \gamma) = e^{-x\gamma}\delta(y)\delta(z) + \frac{c\sigma}{2\pi\tilde{r}}K_{i1}(\sigma\tilde{r})e^{-x\gamma} + \frac{1}{2\pi}\int_0^\infty F_0(x, B; \gamma)J_0(B\tilde{r})B\,dB, \tag{5.10}$$

where K_{i1} is the Bickley function[19]:

$$\frac{1}{\tilde{r}}K_{i1}(\sigma\tilde{r}) = \int_\sigma^\infty K_0(\beta\tilde{r})\,d\beta = \int_0^\infty \tan^{-1}\left(\frac{B}{\sigma}\right)J_0(B\tilde{r})\,dB, \tag{5.11}$$

in which K_0 is the modified Bessel function.[16] Since $(c\sigma/B)\tan^{-1}(B/\sigma)$ is analytic in the B plane cut by $C_+ \cup C_-$, $F_0(x, B; \gamma)$ is analytic in the B plane cut by $C_+ \cup C_-$, with the same poles as $F(x, B; \gamma)$.

To evaluate the integral in (5.10), consider an odd representation of the ordinary Bessel function,[16]

$$J_0(B\tilde{r}) = \frac{1}{\pi i}\int_1^\infty \frac{e^{iB\tilde{r}t}}{(t^2-1)^{\frac{1}{2}}}\,dt - \frac{1}{\pi i}\int_{-\infty}^{-1}\frac{e^{iB\tilde{r}t}}{(t^2-1)^{\frac{1}{2}}}\,dt, \tag{5.12}$$

for $B\tilde{r} \in (0, \infty)$, and treat the integrals as contour integrations in the complex t plane. Deforming the contours $t \in (1, +\infty)$ and $t \in (-\infty, -1)$ to lift them slightly above the real axis, we can substitute (5.12) into the last term of (5.10) and use Fubini's theorem to invert the order of integration:

$$\frac{1}{2\pi}\int_0^\infty F_0(x, B; \gamma)J_0(B\tilde{r})B\,dB = \frac{1}{2\pi^2 i}\int_1^\infty \frac{dt}{(t^2-1)^{\frac{1}{2}}}\left(\int_0^\infty F_0(x, B; \gamma)e^{iB\tilde{r}t}B\,dB - \int_0^\infty F_0(x, B; \gamma)e^{-iB\tilde{r}t}B\,dB\right). \tag{5.13}$$

Then, using Cauchy's theorem and the contours

[19] W. G. Bickley and J. Naylor, Phil. Mag. **20**, 343 (1935).

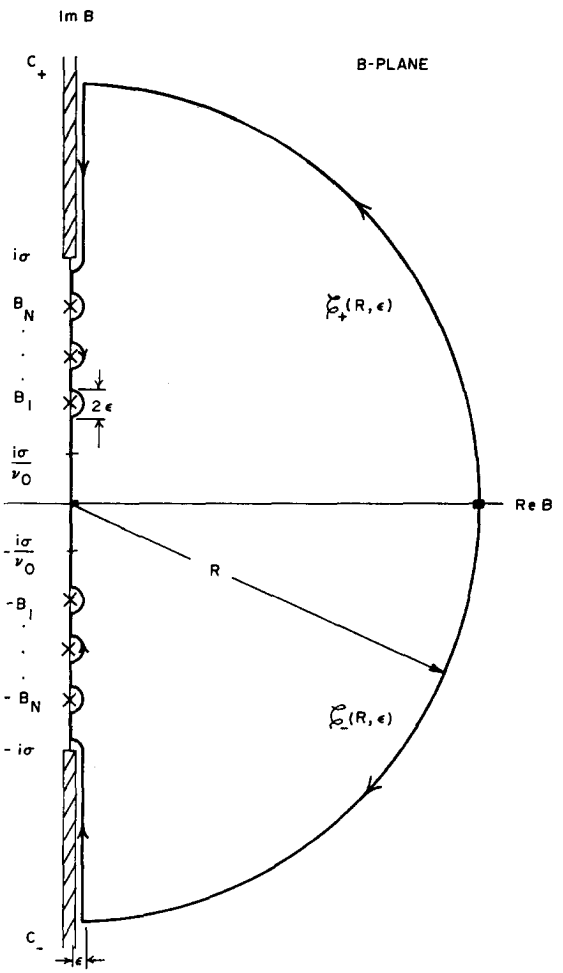


FIG. 5. The Hankel transform inversion contours for $F(x, B; \gamma)$.

$C_+(R, \epsilon)$ and $C_-(R, \epsilon)$ illustrated in Fig. 5, we obtain

$$\frac{1}{2\pi}\int_0^\infty F_0(x, B; \gamma)J_0(B\tilde{r})B\,dB = \lim_{\substack{R\to\infty\\ \epsilon\to 0}}\left[\frac{-1}{2\pi^2 i}\int_1^\infty \frac{dt}{(t^2-1)^{\frac{1}{2}}}\left(\int_{C_+(R,\epsilon)} F_0(x, B; \gamma)e^{iB\tilde{r}t}B\,dB - \int_{C_-(R,\epsilon)} F_0(x, B; \gamma)e^{-iB\tilde{r}t}B\,dB\right)\right]$$
$$= \frac{1}{\pi}\sum_{n=1}^N K_0(\beta_n\tilde{r})B_n\mathcal{R}_n(x; \gamma) + \frac{1}{2\pi^2 i}\int_\sigma^\infty \mathcal{F}(x, \beta, \gamma)K_0(\beta\tilde{r})\beta\,d\beta - \frac{c\sigma}{2\pi\tilde{r}}K_{i1}(\sigma\tilde{r})e^{-x\gamma}, \tag{5.14}$$

where

$$\mathcal{F}(x, \beta; \gamma) = \lim_{\epsilon\to 0}[F(x, i\beta - \epsilon; \gamma) - F(x, i\beta + \epsilon; \gamma)]$$

for $\beta \in (\sigma, \infty)$ and where

$$\mathcal{R}_n(x, \gamma) = \lim_{B\to B_n}[(B - B_n)F(x, B; \gamma)]$$

is the residue of $F(x, B; \gamma)$ at the pole $B_n = i\beta_n$, $n = 1, 2, \cdots, N$. Substituting (5.14) into (5.10) and using the results of Sec. 4 to evaluate $\mathcal{R}_n(x; \gamma)$, we obtain

$$\rho_3(x, y, z; \gamma) = e^{-x\gamma}\delta(y)\delta(z) + \frac{1}{2\pi^2 i}\int_\sigma^\infty \mathcal{F}(x, \beta; \gamma)K_0(\beta\tilde{r})\beta\, d\beta + \sum_{n=1}^{N}\frac{\mathcal{N}_n(\gamma)}{\mathcal{D}_n}K_0(\beta_n\tilde{r})\Phi_n(x), \quad (5.15)$$

where

$$\mathcal{N}_n(\gamma) = (-1)^n \times \frac{h_{2(\pm)}(B_n; \gamma^{-1}) + (\pm)h_{2(\pm)}(B_n; -\gamma^{-1})e^{-\gamma\tau}}{\Lambda(\xi(\gamma^{-1}, B_n))}, \quad (5.16)$$

$$\mathcal{D}_n = (\pi/\lambda_n)\,|H_{(\pm)}(B_n; \zeta_0(B_n))|^2 \cdot [D_n^\infty + D_n^\tau], \quad (5.17)$$

in which

$$\lambda_n = i/\zeta_0(B_n) = [\beta_n^2 - (\sigma/\nu_0)^2]^{\frac{1}{2}}, \quad (5.18)$$

$$(\pm) = \text{sign}\,[(-1)^{n+1}], \quad (5.19)$$

$$D_n^\infty = \frac{n\pi}{\lambda_n^2} + 2\lambda_n\int_0^1 \frac{1 - \theta(t)/\pi}{1 - (t/\nu_0)^2}\frac{t^2\, dt}{(\sigma^2 + B_n^2 t^2)^{\frac{3}{2}}}, \quad (5.20)$$

$$D_n^\tau = \frac{1}{\beta_n}\frac{1}{\zeta_0(B_n)}\frac{\partial}{\partial B}[\mathfrak{U}(B, \tau)]_{B=B_n} = \frac{\tau - T(B_n, n)}{\lambda_n} + \frac{\exp\{i[\tau - T(B_n, n)]\lambda_n\}}{\beta_n} \times \frac{\partial}{\partial B}\left(\frac{H_{(\pm)}(B; -\zeta_0(B))}{H_{(\pm)}(B; \zeta_0(B))}\right)_{B=B_n}, \quad (5.21)$$

and where

$$\Phi_n(x) = a_n\cos[\lambda_n(\tfrac{1}{2}\tau - x)] + b_n\sin[\lambda_n(\tfrac{1}{2}\tau - x)] - \int_0^{\alpha(B_n)} A_n(\nu)[e^{-x/\nu} - (-1)^n e^{-(\tau - x)/\nu}]\, d\nu, \quad (5.22)$$

in which

$$a_n = Q_n(-1)^{\frac{1}{2}(n-1)}, \quad n \text{ odd}, \quad n = 1, 2, \cdots, N,$$
$$a_n = 0, \quad n \text{ even}, \quad (5.23)$$
$$b_n = 0, \quad n \text{ odd}, \quad n = 1, 2, \cdots, N,$$
$$b_n = Q_n(-1)^{\frac{1}{2}(n-2)}, \quad n \text{ even}, \quad (5.24)$$

$$Q_n = [(2u)^{\frac{1}{2}}/\lambda_n]\,|H_{(\pm)}(B_n; \zeta_0(B_n))|, \quad (5.25)$$

$$u = (\sigma^2/\nu_0^2)(\nu_0^2 - 1)/[1 - (1 - c)\nu_0^2], \quad (5.26)$$

and

$$A_n(\nu) = \tfrac{1}{2}c\sigma\frac{X(-\nu, B_n)}{M(\xi(\nu, B_n))}\frac{H_{(\pm)}(B_n, \nu)}{(1 - B_n^2\nu^2)^{\frac{1}{2}}}. \quad (5.27)$$

To finally obtain $G(\mathbf{r}; \mathbf{r}')$, we need only perform the Laplace inversion

$$G(\mathbf{r}; \mathbf{r}') = \frac{1}{2\pi i}\int_{-i\infty}^{+i\infty}\rho_3(x, y - y', z - z'; \gamma)e^{x'\gamma}\, d\gamma, \quad (5.28)$$

for $0 \le x, x' \le \tau$ and $-\infty < y, y', z, z' < +\infty$, using contour integration and the calculus of residues. Recall from Sec. 4 that, if $\zeta = 1/\gamma$, then $X(1/\gamma, B)$ is analytic in the γ plane cut by $\tilde{C}_0(B) = \{(B^2 + \eta^2)^{\frac{1}{2}}: \eta \in (\sigma, \infty)\}$ and has no zeros or poles. On the other hand,

$$\Lambda\left(\xi\left(\frac{1}{\gamma}, B\right)\right) = 1 - \tfrac{1}{2}c\sigma\int_{\tilde{C}_0(B)}\left(\frac{1}{s - \gamma} + \frac{1}{s + \gamma}\right)\frac{ds}{(s^2 - B^2)^{\frac{1}{2}}} \quad (5.29)$$

is analytic in the γ plane cut by $\tilde{C}_0(B) \cup \tilde{C}_0^-(B)$, where

$$\tilde{C}_0^-(B) = \{-(B^2 + \eta^2)^{\frac{1}{2}}: \eta \in (\sigma, \infty)\},$$

with zeros at $\pm\gamma_0(B)$, where

$$\gamma_0(B) = 1/\zeta_0(B) = [B^2 + (\sigma/\nu_0)^2]^{\frac{1}{2}}.$$

Since $H_\pm(B; 1/\gamma)$ is analytic in the γ plane cut by $\tilde{C}_0^-(B)$, substituting (4.22) into (5.16) reveals that $\mathcal{N}_n(\gamma)$ is analytic in the γ plane cut by $\tilde{C}_0(B) \cup \tilde{C}_0^-(B)$, with poles at $\pm\gamma_0(B)$.

Then, using Cauchy's theorem and the contours Γ_+ and Γ_- illustrated in Fig. 6,

$$\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\mathcal{N}_n(\gamma)e^{x'\gamma}\, d\gamma = -\frac{(-1)^n}{2\pi i}\int_{\Gamma_+}\frac{X(-\gamma^{-1}, B_n)H_{(\pm)}(B_n; \gamma^{-1})e^{-(\tau - x')\gamma}\, d\gamma}{\gamma\Lambda(\xi(\gamma^{-1}, B_n))} - \frac{1}{2\pi i}\int_{\Gamma_-}\frac{X(-\gamma^{-1}, B_n)H_{(\pm)}(B_n; -\gamma^{-1})e^{x'\gamma}\, d\gamma}{\gamma\Lambda(\xi(\gamma^{-1}, B_n))} = \Phi_n(x'). \quad (5.30)$$

The use of (5.30) and (3.14) in (5.28) and (5.15) results in the following expression for $G(\mathbf{r}; \mathbf{r}')$:

$$G(\mathbf{r}; \mathbf{r}') = \delta^3(\mathbf{r} - \mathbf{r}') + \frac{1}{2\pi^2 i}\int_\sigma^\infty \mathcal{E}(x, x'; \beta)K_0(\beta\,|\tilde{\mathbf{r}} - \tilde{\mathbf{r}}'|)\beta\, d\beta + \sum_{n=1}^{N}\frac{K_0(\beta_n\,|\tilde{\mathbf{r}} - \tilde{\mathbf{r}}'|)}{\mathcal{D}_n}\Phi_n(x)\Phi_n(x'), \quad (5.31)$$

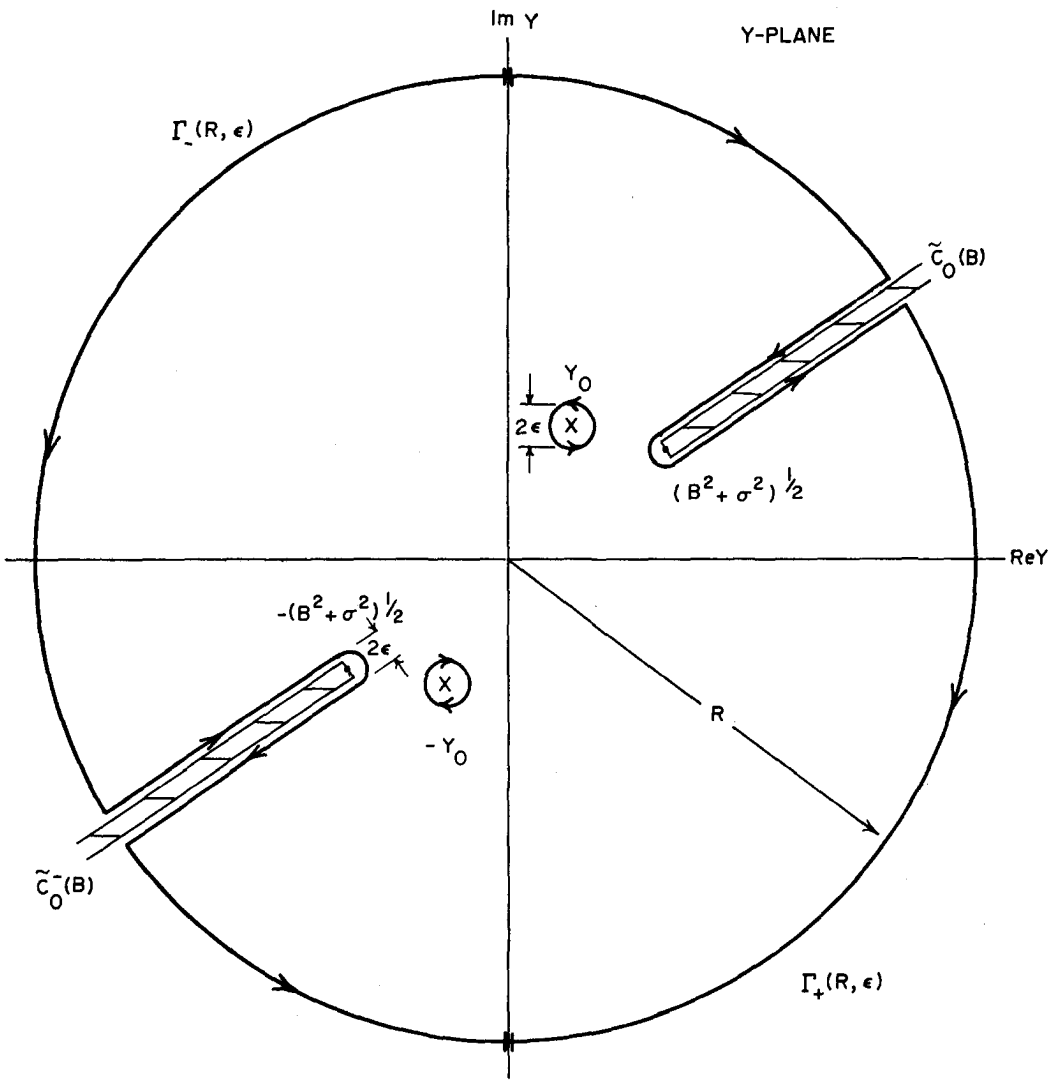


FIG. 6. The Laplace transform inversion contours for terms comprising $F(x, B; \gamma)$.

where

$$\mathcal{E}(x, x'; \beta) = \lim_{\substack{\epsilon \to 0 \\ \epsilon > 0}} [\mathcal{G}(x, i\beta - \epsilon; x') - \mathcal{G}(x, i\beta + \epsilon; x')],$$

where $\mathcal{D}_n$ is given by (5.17) and where $\Phi_n(x)$ is given by (5.22).

It is interesting to note that $\Phi_n(x)$ is an eigenfunction of the 1-dimensional integral operator $\Delta_{B_n}[\cdot]$:

$$\Phi_n(x) = \Delta_{B_n}[\Phi_n](x) = \tfrac{1}{2}c\sigma \int_0^\tau K(|x - \hat{x}|; B_n^2)\Phi_n(\hat{x})\, d\hat{x}. \tag{5.32}$$

Thus, (5.31) represents a general eigenfunction expansion over the point spectrum and over the continuous spectrum of $\Delta_B[\cdot]$.

Since

$$K_0(\beta\tilde{r}) \xrightarrow[\tilde{r}\to\infty]{} O(e^{-\beta\tilde{r}}\tilde{r}^{-\frac{1}{2}}), \tag{5.33}$$

the third term of (5.31) dominates for $|\tilde{\mathbf{r}} - \tilde{\mathbf{r}}'| \gg 1/\sigma$, and we obtain the following relatively simple expression:

$$G(\mathbf{r};\mathbf{r}') = \sum_{n=1}^{N} \frac{K_0(\beta_n|\tilde{\mathbf{r}} - \tilde{\mathbf{r}}'|)}{\mathcal{D}_n}\Phi_n(x)\Phi_n(x') + O(e^{-\sigma|\tilde{\mathbf{r}}-\tilde{\mathbf{r}}'|}), \tag{5.34}$$

for $|\tilde{\mathbf{r}} - \tilde{\mathbf{r}}'| \gg 1/\sigma$. Furthermore, in the case of a "wide slab," the elements of (5.34) reduce to particularly simple closed forms. For $\tau \gg \alpha(i\beta_n)$, the approximation (4.25a) substituted into expressions (5.2) and (5.17)–(5.27) yields the following:

$$\beta_n = \hat{\beta}_n + O(e^{-\tau/\alpha(i\hat{\beta}_n)}), \tag{5.35a}$$

where (see Fig. 4)

$$\tau = T(i\hat{\beta}_n, n); \tag{5.35b}$$

$$\lambda_n = \hat{\lambda}_n + O(e^{-\tau/\alpha(i\hat{\beta}_n)}), \tag{5.36a}$$

where

$$\hat{\lambda}_n = [\hat{\beta}_n^2 - (\sigma/\nu_0)^2]^{\frac{1}{2}}; \tag{5.36b}$$

$$\mathfrak{D}_n = (\pi/\hat{\lambda}_n)\hat{D}_n^\infty + O(e^{-\tau/a(i\hat{\beta}_n)}), \tag{5.37a}$$

where

$$\hat{D}_n^\infty = \frac{n\pi}{\hat{\lambda}_n^2} + 2\hat{\lambda}_n \int_0^1 \frac{1-\theta(t)/\pi}{1-(t/\nu_0)^2} \frac{t^2\,dt}{(\sigma^2 - \hat{\beta}_n^2 t^2)^{\frac{3}{2}}}; \tag{5.37b}$$

and

$$\Phi_n(x) = \hat{\Phi}_n(x) + O(e^{-\tau/a(i\hat{\beta}_n)}), \tag{5.38a}$$

where

$$\hat{\Phi}_n(x) = \frac{(2u)^{\frac{1}{2}}}{\hat{\lambda}_n}\hat{\psi}_n(x) - \tfrac{1}{2}c\sigma \int_0^{a(i\hat{\beta}_n)} \frac{X(-\nu; i\hat{\beta}_n)}{M(\xi(\nu, i\hat{\beta}_n))} \times [e^{-x/\nu} - (-1)^n e^{-(\tau-x)/\nu}] \frac{d\nu}{(1+\hat{\beta}_n^2\nu^2)^{\frac{1}{2}}}, \tag{5.38b}$$

in which

$$\hat{\psi}_n(x) = (-1)^{\frac{1}{2}(n-1)} \cos[\hat{\lambda}_n(\tfrac{1}{2}\tau - x)], \quad n \text{ odd}, \quad n = 1, 2, \cdots, N,$$
$$= (-1)^{\frac{1}{2}(n-2)} \sin[\hat{\lambda}_n(\tfrac{1}{2}\tau - x)], \quad n \text{ even}. \tag{5.38c}$$

To note the limitations of simple diffusion theory, compare the transport-theory solution (5.31) to the solution $G_{\text{diff}}(\mathbf{r}; \mathbf{r}')$, obtained by using the diffusion approximation in the transport equation:

$$G_{\text{diff}}(\mathbf{r}; \mathbf{r}') = \sum_{m=1}^{\infty} \frac{K_0(d_m|\tilde{\mathbf{r}} - \tilde{\mathbf{r}}'|)}{3\pi\sigma_a(\tau + 2l)} \Psi_m(x)\Psi_m(x'), \tag{5.39}$$

where $d_m = \{3\sigma_a\sigma + [m\pi/(\tau + 2l)]^2\}^{\frac{1}{2}}$, σ_a is the absorption cross section, $l \approx 0.7/\sigma$ is the extrapolation distance, and $\Psi_m(x) = \sin[m\pi(x + l)/(\tau + 2l)]$ is an eigenfunction of the 1-dimensional diffusion operator,

$$\frac{d^2\Psi_m(x)}{dx^2} = -\left(\frac{m\pi}{\tau + 2l}\right)^2 \Psi_m(x). \tag{5.40}$$

Although both (5.31) and (5.39) are eigenfunction expansions, diffusion theory predicts an infinite point spectrum for all slab widths, while transport theory yields a continuous spectrum plus a finite point spectrum which is empty for thin slabs. For $0 < \tau < \tau_{\min}$, only the continuum expansion [second term of (5.31)] remains. And, as is typical in 1-dimensional problems, the transport-theory solution includes an "end correction" [third term of (5.22) that is important near the boundaries $x = 0$ or τ] which is not found in diffusion theory.

Finally, let us compare the dominant buckling modes of transport theory with those of diffusion theory in the normal direction [λ_1 vs $\pi/(\tau + 2l)$] and in the radial direction (β_1 vs d_1). Using $\beta_1 \approx [(\sigma/\nu_0)^2 + (n\pi/\tau)^2]^{\frac{1}{2}}$, Eq. (5.5), and $\nu_0^2 \approx 1/3(1 - c) = \sigma/3\sigma_a$ (for c near 1), we see that the fundamental diffusion and transport modes are approximately equal only in a highly scattering, very wide slab.

6. THE TRANSFORM INVERSIONS BY DIRECT METHODS: A CONVENIENT SOLUTION FOR $|\tilde{\mathbf{r}} - \tilde{\mathbf{r}}'| \lesssim 1/\sigma$

For $B \in C_+ \cup C_-$, the Neumann series solution to (4.14) apparently converges rather slowly, so that $\mathcal{E}(x, x'; \beta)$ is difficult to determine accurately. Hence, there is little advantage in using (5.31) instead of (3.1) to evaluate $G(\mathbf{r}; \mathbf{r}')$ for $|\tilde{\mathbf{r}} - \tilde{\mathbf{r}}'| \lesssim 1/\sigma$. However, we can develop an expression, which is more suitable than either (3.1) or (5.31), for evaluating $G(\mathbf{r}; \mathbf{r}')$ for small and intermediate radial arguments. To accomplish this, we first perform the Laplace inversion in γ using contour integration. Then, we analytically invert that part of the Fourier inversion integral which decays slowly as $B \to \infty$, leaving the rapidly converging part for numerical integration.

To evaluate (3.14), we first note that, although $F(x, B; \gamma)$ is analytic in the γ plane for all $|\gamma| < \infty$ and approaches a definite limit for $\gamma \to \infty$ in the right half-plane, the individual terms in the expressions (4.7) or (4.10) are not analytic for all γ. Therefore, one can integrate $F(x, B; \gamma)$ term by term, using (4.7), or (4.3) and (4.10), and the techniques of contour integration and calculus of residues.

In particular, consider (4.7), the singular integral equation satisfied by $F(x, B; \gamma)$, and divide through by $\Lambda(\xi(\gamma^{-1}, B))$:

$$F(x, B; \gamma) = \frac{e^{-x\gamma}}{\Lambda(\xi(\gamma^{-1}, B))} + \tfrac{1}{2}c\sigma \int_{\tilde{C}_0(B)} \frac{F(x, B; s)}{\Lambda(\xi(\gamma^{-1}, B))(\gamma - s)} \frac{ds}{(s^2 - B^2)^{\frac{1}{2}}} - e^{-\tau\gamma}\tfrac{1}{2}c\sigma \int_{\tilde{C}_0(B)} \frac{F(\tau - x, B; s)}{\Lambda(\xi(\gamma^{-1}, B))(\gamma + s)} \frac{ds}{(s^2 - B^2)^{\frac{1}{2}}}. \tag{6.1}$$

Referring to (5.29), each term of (6.1) is analytic in the γ plane cut by $\tilde{C}_0(B) \cup \tilde{C}_n^-(B)$ and has poles at $\pm\gamma_0(B)$.

Now substitute (6.1) into (3.14) and refer to Fig. 6. In the first term, we use the contour $\Gamma_+(R, \epsilon)$ for $\tau \geq x > x' \geq 0$ and use $\Gamma_-(R, \epsilon)$ for $\tau \geq x' > x \geq 0$. In the second term, we use $\Gamma_-(R, \epsilon)$ and, in the third term, we use $\Gamma_+(R, \epsilon)$. Letting $R \to \infty$ and $\epsilon \to 0$, we obtain the following result:

$$\mathcal{G}(x, B; x') = \mathcal{G}_\infty(x, B; x') - \mathcal{G}_0(x, B; x') - \mathcal{G}_0(\tau - x, B; \tau - x') \tag{6.2}$$

where [see Eqs. (5.26) and (4.12)–(4.20)]

$$\mathfrak{G}_\infty(x, B; x') = \delta(x - x') + P(|x - x'|, B; 0) \quad (6.3)$$

and

$$\mathfrak{G}_0(x, B; x') = \tfrac{1}{2}c\sigma\int_0^{\alpha(B)} F\left(x, B; \frac{1}{s}\right)P(x', B; s)\frac{ds}{(1 - B^2s^2)^{\frac{1}{2}}}, \quad (6.4)$$

in which

$$P(x', B; s) = \frac{u\zeta_0^2(B)e^{-x'/\zeta_0(B)}}{s + \zeta_0(B)} + \tfrac{1}{2}c\sigma\int_0^{\alpha(B)} \frac{e^{-x'/t}}{M(\xi(t, B))(s + t)}\frac{dt}{(1 - B^2t^2)^{\frac{1}{2}}}. \quad (6.5)$$

One can arrive at an equivalent expression by substituting (4.3), (4.10), and (4.11) into (3.14) and by using the contours illustrated in Fig. 6.

Finally, the use of (6.2) in (3.6) and (3.8) yields the following expression for $G(\mathbf{r}; \mathbf{r}')$:

$$G(x, y, z; x', y', z') = G_\infty(x, y, z; x', y', z') - G_0(x, y, z; x', y', z') - G_0(\tau - x, y, z; \tau - x', y', z'), \quad (6.6)$$

where[20] we have evaluated

$$G_\infty(\mathbf{r}; \mathbf{r}') = \frac{1}{2\pi}\int_0^\infty \mathfrak{G}_\infty(x, B; x')J_0(B|\tilde{\mathbf{r}} - \tilde{\mathbf{r}}'|)B\,dB = \delta^3(\mathbf{r} - \mathbf{r}') + \frac{u}{2\pi}\frac{\exp(-\sigma|\mathbf{r} - \mathbf{r}'|\nu_0^{-1})}{|\mathbf{r} - \mathbf{r}'|} + \frac{c\sigma^2}{4\pi}\frac{1}{|\mathbf{r} - \mathbf{r}'|}\int_0^1 \frac{\exp(-\sigma|\mathbf{r} - \mathbf{r}'|t^{-1})}{M(t)}\frac{dt}{t^2}, \quad (6.7)$$

and where

$$G_0(\mathbf{r}; \mathbf{r}') = \frac{1}{2\pi}\int_0^\infty \mathfrak{G}_0(x, B; x')J_0(B|\tilde{\mathbf{r}} - \tilde{\mathbf{r}}'|)B\,dB = \frac{c\sigma}{4\pi}\int_0^\infty J_0(B|\tilde{\mathbf{r}} - \tilde{\mathbf{r}}'|)B\left\{\int_\sigma^\infty F(x, B; (s^2 + B^2)^{\frac{1}{2}}) \times \left[\frac{u}{[B^2 + (\sigma/\nu_0)^2]^{\frac{1}{2}}}\frac{\exp\{-x'[B^2 + (\sigma/\nu_0)^2]^{\frac{1}{2}}\}}{(B^2 + s^2)^{\frac{1}{2}} + [B^2 + (\sigma/\nu_0)^2]^{\frac{1}{2}}} + \tfrac{1}{2}c\sigma\int_\sigma^\infty \frac{\exp[-x'(B^2 + t^2)^{\frac{1}{2}}]}{(B^2 + s^2)^{\frac{1}{2}} + (B^2 + t^2)^{\frac{1}{2}}}\frac{dt}{M(\sigma/t)(B^2 + t^2)^{\frac{1}{2}}}\right] \times \frac{ds}{(B^2 + s^2)^{\frac{1}{2}}}\right\}dB. \quad (6.8)$$

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

$G_\infty(\mathbf{r}; \mathbf{r}')$ is the point-uncollided-source Green's function for an infinite medium [$G_\infty(\mathbf{r}; \mathbf{r}')$ satisfies (2.3) with the "0 to τ" limits of integration in $\hat{x}$ extended to "$-\infty$ to $+\infty$"], while the G_0 are the boundary correction terms. For small[21] $|\tilde{r} - \tilde{r}'|$ $G_0(x, y, z; x', y', z')$ and $G_0(\tau - x, y, z; \tau - x', y', z')$ are significant in (6.6) for x and/or x' near 0 and τ, respectively, because by (4.1),

$$F(x, B; (s^2 + B^2)^{\frac{1}{2}}) \propto O(\exp[-x(B^2 + s^2)^{\frac{1}{2}}]), \quad (6.9)$$

for $B \in (0, \infty)$ and $s \in (\sigma, \infty)$, and because of the exponential terms $\exp\{-x'[B^2 + (\sigma/\nu_0)^2]^{\frac{1}{2}}\}$ and

$$\exp[-x'(B^2 + t^2)^{\frac{1}{2}}]$$

in (6.8).

For x' and x not both zero, the integrals in (6.8) can be shown to converge absolutely, and that expression can be used to evaluate $G_0(\mathbf{r}; \mathbf{r}')$ numerically. Naturally, the convergence is more rapid the farther x and/or x' are from 0. For $x = x' = 0$, it is not evident that the integrals in (6.8) are absolutely convergent, so it may be necessary to substitute (5.8) into (6.8). The resulting integrals that do not converge absolutely can be evaluated analytically, while the integrals involving $F_0(x, B; (s^2 + B^2)^{\frac{1}{2}})$ converge absolutely and can be evaluated numerically. With respect to the boundary $x, x' = \tau$, the same observations are true in evaluating $G_0(\tau - x, y, z; \tau - x', y', z')$.

For the purpose of numerical integration, it is useful to note that the approximations (4.25) can be used to put the expression for $F(x, B; (s^2 + B^2)^{\frac{1}{2}})$ into simple closed form. Furthermore, even for thin slabs, these approximations are accurate for B sufficiently large so that $B \in \mathfrak{B} = \{B' \in (0, \infty): \tau \gg \alpha(B')\}$.

Therefore, for $|\mathbf{r} - \mathbf{r}'| \ll 1/\sigma$, or when $0 < \tau < \tau_{\min}$, expressions (6.6)–(6.8) are generally much more suitable than either (5.31) or (3.1) for numerically evaluating $G(\mathbf{r}; \mathbf{r}')$. In the special case that c or τ is very small, (3.1) might be more efficient.

7. SOME ADDITIONAL RESULTS: THE PHYSICAL SIGNIFICANCE OF ρ_3, F, AND $\mathfrak{G}$

To obtain the solution to (2.1) for any specific uncollided density $S(\mathbf{r})$, we need only use the results of (5.34) and (6.6) in (2.5). However, in the process of solving for $G(\mathbf{r}; \mathbf{r}')$, we have also obtained the solutions to (2.1) for some physically significant uncollided sources.

A. Normally Incident Beams

In particular, referring to (5.11) and (3.15), we see that $\rho_3(x, y - y', z - z'; \gamma)$ satisfies (2.1) with $S(\mathbf{r}) = \delta(y - y')\delta(z - z')e^{-x\gamma}$, where γ is a free parameter.

[21] As $|\tilde{r} - \tilde{r}'|$ increases, so does $(G_\infty - G)/G$ for all $x, x' \in [0, \tau]$.

For example, $\rho_3(x, y-y', z-z'; \sigma)$ is the neutron density at (x, y, z) from a pencil beam normally incident to the slab at $(0, y', z')$, and is therefore the Green's function for all problems involving beams normally incident to a slab. Likewise, referring to (2.6), $[\rho_3(x, y-y', z-z'; 0) - \delta(y-y')\delta(z-z')]/c\sigma$ is the neutron density at (x, y, z) from an isotropic line source normally incident to the slab at $y = y'$ and $z = z'$ and in the slab $x \in [0, \tau]$.

By appropriately adjusting the free parameter γ, one can think of a variety of other problems for which $\rho_3(x, y, z; \gamma)$ is the solution. In particular, the solution to an interesting "pseudo-two-group" transport problem in a slab can be written in terms of ρ_3. That is, suppose one considers a particle, such as a neutron, that has no charge and that interacts weakly with other free particles of its kind. Consider the transport of this particle in a homogeneous slab (Fig. 1), and suppose the particle physics is described by the following model:

(1) The velocity of all uncollided particles is a constant v_0, while the velocity of all particles which have had at least one collision is a constant v.

(2) The mean free path for a first collision is $1/\gamma_0$, while the mean free path for all succeeding collisions is $1/\sigma$.

(3) For every first collision there are c_0 secondaries emitted isotropically, while for every subsequent collision there are c secondaries emitted isotropically.

Then, the steady-state particle density $N(x, y-y', z-z')$, from a pencil beam of these particles normally incident to the slab at $(0, y', z')$, is simply

$$N(x, y-y', z-z') = \frac{v_0 c_0 \gamma_0}{vc\sigma}[\rho_3(x, y-y', z-z'; \gamma_0) - \delta(y-y')\delta(z-z')e^{-x\gamma_0}] + \delta(y-y')\delta(x-x')e^{-x\gamma_0}. \quad (7.1)$$

This solution might have some application in high-energy neutron, or gamma, shielding problems.

To evaluate $\rho_3(x, y, z; \gamma)$ for any γ, refer to the equations of Sec. 5. For $|\tilde{\mathbf{r}} - \tilde{\mathbf{r}}'| \gg 1/\sigma$, we can use (5.15) to obtain

$$\rho_3(x, y-y', z-z'; \gamma) = \sum_{n=1}^{N} \frac{\mathcal{N}_n(\gamma)}{\mathfrak{D}_n} K_0(\beta_n |\tilde{\mathbf{r}} - \tilde{\mathbf{r}}'|)\Phi_n(x) + O(e^{-\sigma|\tilde{\mathbf{r}}-\tilde{\mathbf{r}}'|}), \quad (7.2)$$

where $\mathcal{N}_n(\gamma)$, $\mathfrak{D}_n$, and $\Phi_n(x)$ are given by (5.16)–(5.27). When $\tau \gg \alpha(i\beta_n)$, the approximations (4.25) and (5.35)–(5.38) can be used to reduce (7.2) to a simple, closed form.

To evaluate $\rho_3(x, y-y', z-z'; \gamma)$ for $|\tilde{\mathbf{r}} - \tilde{\mathbf{r}}'| \lesssim 1/\sigma$ or when $0 < \tau < \tau_{\min}$, one can use (5.10) and evaluate the last term numerically, using approximations (4.25) for $B \in \mathfrak{B}$. However, if $0 \leq c < 1$ and $\operatorname{Re}\gamma > -\sigma/\nu_0$, it is more convenient to define

$$F_1(x, B; \gamma) = F(x, B; \gamma) - \frac{\exp(-x\gamma)}{\Lambda(\xi(\gamma^{-1}, B))} - \frac{u\zeta_0(B)\exp(-x/\zeta_0(B))}{\gamma - [\zeta_0(B)]^{-1}} - \frac{c\sigma}{2\gamma}\int_0^{\alpha(B)} \frac{\exp(-x/s)}{s - 1/\gamma} \frac{ds}{M(\xi(s, B))(1 - B^2s^2)^{\frac{1}{2}}}. \quad (7.3)$$

Then,

$$\rho_3(x, y-y', z-z'; \gamma) = \rho_{3\infty}(x, y-y', z-z'; \gamma) + \frac{1}{2\pi}\int_0^{\infty} F_1(x, B; \gamma) J_0(B|\tilde{\mathbf{r}} - \tilde{\mathbf{r}}'|) B\, dB, \quad (7.4)$$

where

$$\rho_{3\infty}(x, y-y', z-z'; \gamma) = w(x)\delta(y-y')\delta(z-z')e^{-x\gamma} + \frac{u}{2\pi} q_\gamma(x, |\mathbf{r} - \mathbf{r}'|; \nu_0) + \frac{c\sigma^2}{4\pi}\int_0^1 \frac{q_\gamma(x, |\mathbf{r} - \mathbf{r}'|; t)}{t^2 M(t)}\, dt, \quad (7.5)$$

for $-\infty < x, y, y', z, z' < +\infty$, in which

$$q_\gamma(x, \tilde{r}; t) = \exp(-x\gamma)\int_{-x}^{\infty} \frac{\exp[-\sigma t^{-1}(s^2 + \tilde{r}^2)^{\frac{1}{2}} - \gamma s]\, ds}{(s^2 + \tilde{r}^2)^{\frac{1}{2}}} \quad (7.6)$$

and

$$w(x) = 0, \quad x < 0, \qquad = 1, \quad x > 0. \quad (7.7)$$

Since

$$F_1(x, B; \gamma) \xrightarrow[B\to\infty]{} \begin{cases} O(\exp(-\frac{1}{2}\tau B)B^{-2}), & 0 < x < \tau, \\ O(B^{-2}), & x = 0 \text{ or } \tau, \end{cases} \quad (7.8)$$

the last term of (7.4) is much easier to integrate numerically than is the last term of (5.10). However, we cannot use (7.4) for all γ and c. $\rho_{3\infty}(x, y-y', z-z'; \gamma)$ is the neutron density at (x, y, z) in an

infinite medium from an uncollided source

$$w(x)\delta(y-y')\delta(z-z')e^{-\alpha\gamma},$$

$$\rho_{3\infty}(x, y-y', z-z'; \gamma) = \frac{c\sigma}{4\pi}\iiint_{-\infty}^{+\infty}\frac{\exp(-\sigma|\mathbf{r}-\hat{\mathbf{r}}|)}{|\mathbf{r}-\hat{\mathbf{r}}|^2} \times \rho_{3\infty}(\hat{x}, \hat{y}-y'; \hat{z}-z'; \gamma)\, d^3\hat{r} + w(x)\delta(y-y')\delta(z-z')e^{-\alpha\gamma}, \quad (7.9)$$

and makes sense, physically, only for $0 \leq c < 1$ and $\operatorname{Re}\gamma > -\sigma/\nu_0$.

In radiation transport theory, $\rho_3(x, y, z; \sigma)$ is the solution to the "flashlight problem." Hunt solved a particular problem of this class where a beam of radiation normally incident on a slab atmosphere is modulated by the Bessel function $J_0(B\tilde{r})$. Then the radiation "source function" $J(\mathbf{r})$ satisfies (2.1) with $S(\mathbf{r}) = 2\pi J_0(B\tilde{r})e^{-x}$, where B is a free parameter. In terms of the theory of Sec. 4, his result is simply $J(\mathbf{r}) = 2\pi J_0(B\tilde{r})F(x, B; 1)$, where c is interpreted as the albedo for a single scattering.[5]

To physically interpret $F(x, B; \gamma)$, refer to (3.15), (3.12), and (4.2). $F(x, 0; \sigma)$ represents the neutron density in a slab from a uniform unit beam normally incident to the left face [satisfying (2.1) with $S(\mathbf{r}) = e^{-\sigma x}$], while $f_+(x, 0; \sigma)$ represents the neutron density in a slab from uniform unit beams normally incident to both faces.

B. Asymptotic Theory in the Transverse Directions

For $B \in (0, \infty)$, $F(x, B; \sigma)$ can be interpreted as the "asymptotic-theory" solution for neutron transport in a finite prism. (Asymptotic theory: Assume the form of the solution in the two transverse dimensions to be $e^{i\mathbf{B}\cdot\mathbf{r}}$ and solve the resulting modified transport equation exactly in the third dimension.) That is, consider the integral equation for the neutron density $\rho(x, y, z)$ in a finite prism (see Fig. 7) with a cosine-modulated beam normally incident to the left face:

$$\rho(x, y, z) = \frac{c\sigma}{4\pi}\int_0^{\tau} d\hat{x}\int_{-\frac{1}{2}a}^{\frac{1}{2}a} d\hat{y}\int_{-\frac{1}{2}b}^{\frac{1}{2}b} d\hat{z}\,\rho(\hat{x}, \hat{y}, \hat{z})\frac{\exp(-\sigma|\mathbf{r}-\hat{\mathbf{r}}|)}{|\mathbf{r}-\hat{\mathbf{r}}|^2} + e^{-\sigma x}\cos\left(\frac{\pi}{a}y\right)\cos\left(\frac{\pi}{b}z\right), \quad (7.10)$$

for $x \in [0, \tau]$, $y \in [-\frac{1}{2}a, \frac{1}{2}a]$, and $z \in [-\frac{1}{2}b, \frac{1}{2}b]$. Assuming a solution of the form

$$\rho(x, y, z) = \cos(\pi a^{-1}y)\cos(\pi b^{-1}z)n(x; a, b) \quad (7.11)$$

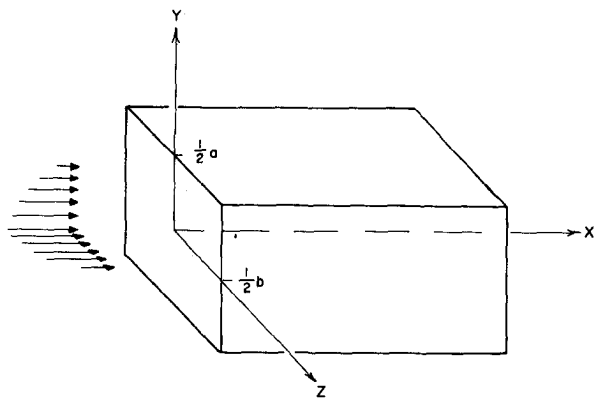


FIG. 7. A rectangular parallelopiped with a cosine-modulated beam incident to the left face.

and extending the limits of integration in (7.10) to $(-\infty, +\infty)$ in y and z, it is easy to show that $n(x; a, b)$ satisfies (3.15) with $\gamma = \sigma$ and with

$$B = [(\pi/a)^2 + (\pi/b)^2]^{\frac{1}{2}} = B_g,$$

the "geometric buckling." Thus,

$$\rho(x, y, z) \approx \rho_{\text{asymptotic}}(x, y, z) = \cos(\pi a^{-1}y)\cos(\pi b^{-1}z)F(x, B_g; \sigma), \quad (7.12)$$

and many of the results obtained by Kaper,[7] Williams,[6] and Smith[4] can easily be reproduced using the theory cited in Sec. 4.

For example, with $c > 1$ and fixing a and b, the critical length τ_0 of the prism in asymptotic theory is given by the smallest $\tau > 0$ for which the denominator of $F(x, B; \sigma)$ is zero. Using (5.2) and (5.3),

$$\tau_0 = \frac{\pi}{[(\sigma/|\nu_0|)^2 - B_g^2]^{\frac{1}{2}}} - 2\int_0^1 \frac{1-\theta(t)/\pi}{1+(t/|\nu_0|)^2}\frac{dt}{(\sigma^2+B_g^2t^2)^{\frac{1}{2}}} - \mathfrak{U}(B_g, \tau_0), \quad (7.13)$$

where $\mathfrak{U}(B_g, \tau_0) = O(\exp[-\tau_0/\alpha(B_g)])$. For a first estimate or for $\tau_0 \gg \alpha(B_g)$ (i.e., if a and b are small), the first two terms of (7.13) suffice and can be easily evaluated.[12] Furthermore, referring to (5.22)–(5.27), the critical neutron density is proportional to $\cos(\pi a^{-1}y)\cos(\pi b^{-1}z)\Phi_1(x)$ with B_1 replaced by B_g.

C. Plane-Source Densities

To physically interpret $\mathfrak{G}(x, B; x')$, refer to Eq. (3.9). $\mathfrak{G}(x, 0; x')$ is the neutron density at x in a slab from an uncollided plane source at $x' \in [0, \tau]$ and is, therefore, the Green's function for 1-dimensional neutron transport in a slab. Using (2.6), $[\mathfrak{G}(x, 0; x') - \delta(x-x')]/c\sigma$ is the neutron density at x from a plane-isotropic

source at x'. Also, referring to Fig. 7,

$$\rho_{\text{asymptotic}}(x, y, z; x') = \cos(\pi a^{-1}y)\cos(\pi b^{-1}z) \times \mathfrak{G}(x, B_g; x')$$

is the asymptotic-theory solution for neutron transport in a finite prism $(\tau \cdot a \cdot b)$ with a cosine-modulated, plane-uncollided source at $x' \in [0, \tau]$.

$$\rho_{\text{asymptotic}}(x, y, z; x')$$

satisfies (2.1) with

$$S(r) = \cos(\pi a^{-1}y)\cos(\pi b^{-1}z)\delta(x - x').$$

To evaluate $\mathfrak{G}(x, B_g; x')$, it is convenient to use (6.2)–(6.5) with approximations (4.25) for $B_g \in \mathfrak{B}$. In (6.2), $\mathfrak{G}_\infty(x, 0; x')$ represents the infinite medium plane-source Green's function, while $\mathfrak{G}_0(x, 0; x')$ and $\mathfrak{G}_0(\tau - x, 0; \tau - x')$ are the boundary correction terms important near $x = 0$ and $x = \tau$, respectively. Note that (6.4) can be rewritten as in (6.8) and evaluated numerically, the integrand converging exponentially for x and/or x' not on the boundary.

8. 2-DIMENSIONAL NEUTRON TRANSPORT IN A SLAB

The results of 2-dimensional transport theory in a slab are immediately obtained by setting $B = \omega$ and by considering only 1-dimensional Fourier transforms. Section 4 is unchanged. The results of 2-dimensional asymptotic theory can be obtained from Sec. 7 by letting a or b go to infinity.

If $\rho_2(x, y; \gamma)$ satisfies (2.1) with $S(\mathbf{r}) = \delta(y)e^{-x\gamma}$, then $\rho_2(x, y - y'; \gamma)$ is the neutron density at (x, y) in a slab from a "sheet" of neutrons normally incident to the slab at $y = y'$. $\rho_2(x, y - y'; 0)$ is the neutron density at (x, y) in a slab from a plane-uncollided source normal to the slab at $y = y'$.

To evaluate $\rho_2(x, y - y'; \gamma)$ for $|y - y'| \gg 1/\sigma$ and $\tau \geq \tau_{\min}$, we use an equation analogous to (5.15) and (7.2):

$$\rho_2(x, y - y'; \gamma) = \delta(y)e^{-x\gamma} + \frac{1}{2\pi i}\int_\sigma^\infty \mathcal{F}(x, \beta; \gamma)e^{-\beta|y-y'|}\,d\beta + i\sum_{n=1}^{N} e^{-\beta_n|y-y'|}\mathcal{R}_n(x; \gamma) \quad (8.1a)$$

$$= \sum_{n=1}^{N} \frac{\pi}{\beta_n}\frac{\mathcal{N}_n(\gamma)}{\mathcal{D}_n}\Phi_n(x)e^{-\beta_n|y-y'|} + O(e^{-\tau|y-y|}), \quad (8.1b)$$

for $|y - y'| \gg 1/\sigma$. To evaluate $\rho_2(x, y - y'; \gamma)$ for $|y - y'| \not\gg 1/\sigma$ or when $0 < \tau < \tau_{\min}$, we use equations analogous to (7.3)–(7.6) when $\operatorname{Re}\gamma > -\sigma/\nu_0$ and when $0 \leq c < 1$:

$$\rho_2(x, y; \gamma) = \rho_{2\infty}(x, y; \gamma) + \frac{1}{\pi}\int_0^\infty F_1(x, \omega; \gamma)\cos(\omega y)\,d\omega, \quad (8.2)$$

where

$$\rho_{2\infty}(x, y; \gamma) = \delta(y)e^{-x\gamma} + \frac{u}{\pi}j_\gamma(x, y; \nu_0) + \frac{c\sigma^2}{2\pi}\int_0^1 \frac{j_\gamma(x, y; t)}{t^2 M(t)}\,dt, \quad (8.3)$$

in which

$$j_\gamma(x, y; t) = e^{-x\gamma}\int_{-x}^{\infty} K_0\left(\frac{\sigma}{t}(s^2 + y^2)^{\frac{1}{2}}\right)e^{-\gamma s}\,ds. \quad (8.4)$$

In case $\operatorname{Re}\gamma \leq -\sigma/\nu_0$ or $c \geq 1$, the equation analogous to (5.10) is [see (3.4)]

$$\rho_2(x, y; \gamma) = \delta(y)e^{-x\gamma} + \tfrac{1}{2}c\sigma E_1(\sigma|y|)e^{-x\gamma} + \frac{1}{\pi}\int_0^\infty F_0(x, \omega; \gamma)\cos(\omega y)\,d\omega. \quad (8.5)$$

It is interesting to note that $[\rho_2(x, y; 0) - \delta(y)]/c\sigma$ is the neutron density at (x, y) from an isotropic plane source normal to the slab at $y = 0$ and in the slab $x \in [0, \tau]$. This is the solution to a problem considered recently by Williams,[12] and he obtains a result similar to (8.1) in the "wide slab" approximation [(4.25) and (5.35)–(5.38) in (8.1)].

If $G_2(x, y; x', y')$ satisfies (2.1) with

$$S(\mathbf{r}) = \delta(x - x')\delta(y - y'),$$

then it is the neutron density at (x, y) from an uncollided line source at $x = x', y = y'$, and $z \in (-\infty, +\infty)$ in the slab. That is, $G_2(x, y; x', y')$ is the Green's function for 2-dimensional neutron transport in a slab.

To evaluate $G_2(x, y; x', y')$ for $|y - y'| \gg 1/\sigma$ and $\tau \geq \tau_{\min}$, the equation equivalent to (5.34) is

$$G_2(x, y; x', y') = \sum_{n=1}^{N}\frac{\pi}{\beta_n}\frac{\exp(-\beta_n|y - y'|)}{\mathcal{D}_n}\Phi_n(x)\Phi_n(x') + O(\exp(-\sigma|y - y'|)) \quad (8.6)$$

for $|y - y'| \gg 1/\sigma$. To evaluate $G_2(x, y; x', y')$ for $|y - y'| \not\gg 1/\sigma$ or when $0 < \tau < \tau_{\min}$, the equations equivalent to (6.6)–(6.8) are

$$G_2(x, y; x', y') = G_{2\infty}(x, y; x', y') - G_{20}(x, y; x', y') - G_{20}(\tau - x, y; \tau - x', y'), \quad (8.7)$$

TABLE I. A summary of Green's functions obtained for multidimensional neutron transport in a slab.

Functional notation	Description	Equation satisfied	Convenient expressions $\lvert\tilde{\mathbf{r}}-\tilde{\mathbf{r}}'\rvert \gg 1/\sigma$ and $\tau \geq \tau_{\min}$ and $\lvert y-y'\rvert \gg 1/\sigma$ and $\tau \geq \tau_{\min}$	Convenient expressions $\lvert\tilde{r}-\tilde{r}'\rvert \not\ll 1/\sigma$ or $\tau < \tau_{\min}$ and $\lvert y-y'\rvert \not\ll 1/\sigma$ or $\tau < \tau_{\min}$
$G(\mathbf{r};\mathbf{r}')$	3D Green's function (uncollided) point-source density	(2.3)	(5.34) [with (5.35)–(5.38) for $\tau \gg \alpha(B_n)$]	(6.6)–(6.8) [with (4.25) for $B \in \mathfrak{B}$][a]
$G_2(x,y;x',y')$	2D Green's function (uncollided) line-source density	(2.1) with $S(\mathbf{r}) = \delta(x-x')\delta(y-y')$	(8.6) [with (5.35)–(5.38) for $\tau \gg \alpha(B_n)$]	(8.7)–(8.9) [with (4.25) for $B \in \mathfrak{B}$]
$\mathfrak{G}(x,B;x')$	1D Green's function uncollided-plane-source density ($B = 0$) B is a free parameter	(3.9)	(6.2)–(6.5) [with (4.25) for $\tau \gg \alpha(B)$]	
$\rho_3(x,y-y',z-z';\gamma)$	Normal beam Green's function γ is a free parameter: i.e., $\gamma = 0$: normal line source $\gamma = \sigma$: normal pencil beam $\gamma = \gamma_0$: "pseudo-two-group" normal beam	(2.1) with $S(\mathbf{r}) = \delta(y-y')\delta(z-z')e^{-x\gamma}$	(7.2) [with (4.25) and (5.35)–(5.38) for $\tau \gg \alpha(B_n)$]	(i) (7.3)–(7.7) for $\operatorname{Re}\gamma > -\sigma/\nu_0$ and $0 \leq c < 1$ (ii) (5.10) for $\operatorname{Re}\gamma \leq -\sigma/\nu_0$ or $c \geq 1$ [with (4.25) for $B \in \mathfrak{B}$]
$\rho_2(x,y-y';\gamma)$	2D normal beam Green's function γ is a free parameter; i.e., $\gamma = 0$: normal plane source $\gamma = \sigma$: normal sheet $\gamma = \gamma_0$: "pseudo-two-group" normal sheet	(2.1) with $S(\mathbf{r}) = \delta(y-y')e^{-x\gamma}$	(8.1) [with (4.25) and (5.35)–(5.38) for $\tau \gg \alpha(B_n)$]	(i) (8.2)–(8.4) for $\operatorname{Re}\gamma > -\sigma/\nu_0$ and $0 \leq c < 1$ (ii) (8.5) for $\operatorname{Re}\gamma \leq -\sigma/\nu_0$ or $c \geq 1$ [with (4.25) for $B \in \mathfrak{B}$]

[a] $\mathfrak{B} = \{B' \in (0,\infty): \tau \gg \alpha(B')\}$.

where

$$G_{2\infty}(x,y;x',y') = \delta(x-x')\delta(y-y') + \frac{u}{\pi}K_0\left(\frac{\sigma}{\nu_0}[(x-x')^2+(y-y')^2]^{\frac{1}{2}}\right) + \frac{c\sigma^2}{2\pi}\int_0^1 \frac{K_0\left(\frac{\sigma}{t}[(x-x')^2+(y-y')^2]^{\frac{1}{2}}\right)}{t^2M(t)}\,dt \quad (8.8)$$

and

$$G_{20}(x,y;x',y') = \frac{1}{\pi}\int_0^\infty \mathfrak{G}_0(x,\omega;x')\cos[\omega(y-y')]\,d\omega. \quad (8.9)$$

$G_{2\infty}$ is the infinite medium 2-dimensional Green's function, while the G_{20} are boundary correction terms.

In using the equations of this section, the same remarks made in Secs. 5–7 [concerning the rate of convergence of integrals and the use of approximations (4.25)] apply here.

9. CONCLUSIONS

In Secs. 2–8, we have obtained Green's functions for several classes of multidimensional neutron transport problems in a slab, under the assumptions of steady-state, one-speed, and isotropic scattering.

In the general case, the point-source Green's function $G(\mathbf{r};\mathbf{r}')$ can be used to obtain the solution to the transport equation (2.1) for any uncollided neutron density $S(\mathbf{r})$. For example, the neutron density $\rho(\mathbf{r};\mathbf{\Omega}_0)$, from a monodirectional pencil beam incident to the slab at any angle, can be obtained by integrating $G(\mathbf{r};\mathbf{r}')$ along the beam path [see Eqs. (2.2), (2.5), and Fig. 2]:

$$\rho(\mathbf{r};\mathbf{\Omega}_0) = \int_0 G(x,y,z;x',x'\tan\theta_0\cos\varphi_0, x'\tan\theta_0\sin\varphi_0)\exp\left(\frac{-\sigma x'}{\cos\theta_0}\right)dx'. \quad (9.1)$$

And the neutron density from any beam can be obtained by integrating $\rho(\mathbf{r};\mathbf{\Omega}_0)$ over the angular and spatial distribution of the beam. Note that the "pseudo-two-group" problem can be done for nonnormal beams by using (9.1) in (7.1).

For problems with a higher degree of symmetry, e.g., normally incident neutron beam, it is simpler to use another more appropriate Green's function having the same symmetry. Table I gives a summary of the most convenient equations for evaluating these functions under various circumstances.