JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 5, NUMBER 8 AUGUST 1964

Inversion of Cyclic Matrices*

JEAN-LOUIS CALAIS AND KLAUS APPEL
Quantum Chemistry Group, University of Uppsala, Uppsala, Sweden
(Received 15 November 1963)

The inversion of the overlap matrix in LCAO calculations of crystals is discussed for the general case, when an arbitrary number of neighbors is taken into account. Of the several alternative methods that are possible in one dimension, one is singled out—which can be used as well in three dimensions—to calculate the inverse to any desired accuracy. Particular attention is paid to the numerical aspects of the problem.

1. INTRODUCTION

RECENTLY a number of papers have appeared,[1-3] which discuss the inverse and other functions of cyclic matrices. The importance of such calculations in solid-state physics need hardly be stressed here, (see, e.g., Ref. 4).

The methods described in Refs. 1 and 2 are simple and elegant for the special cases treated there—neighbors of first and second order only. Unfortunately it seems to be difficult to use them directly in a more general case. In Ref. 3 the general case is treated in principle.

Since there is generally a rather big gap between general mathematical formulas and numerical calculations, we want to sum up the discussion and particularly stress (1) the connections between the various methods, (2) the final step to the numerical calculations.

For a general cyclic matrix,

$$\Delta = (1, S_1, S_2, \cdots, S_2, S_1)_{\text{cyclic}} = 1 + S, \qquad (1)$$

of order n, we know explicitly both the eigenvalues

$$d_l = \sum_{p=0}^{n-1} S_p e^{2\pi i p l/n}, \quad l = 0, 1, 2, \cdots (n-1), \qquad (2)$$

and the unitary matrix U, which diagonalizes (1):

$$U_{jl} = \frac{1}{n^{\frac{1}{2}}} e^{2\pi i j l/n}; \quad j, l = 0, 1, \cdots (n-1); \qquad (3)$$

$$U^{\dagger}\Delta U = d; \qquad UU^{\dagger} = U^{\dagger}U = 1, \qquad (4)$$

where d is a diagonal matrix with elements d_0, $d_1, \cdots d_{n-1}$. Therefore we also have a closed expression for any function of Δ,

$$F(\Delta) = UF(d)U^{\dagger} = UF(U^{\dagger}\Delta U)U^{\dagger}, \qquad (5)$$

i.e.,

$$F(\Delta)_{\mu\nu} = \frac{1}{n}\sum_{l=0}^{n-1} F(d_l) e^{2\pi i(\mu-\nu)l/n}. \qquad (6)$$

By means of these formulas one can avoid the convergence problems associated with the expansions of $F(\Delta)$ in powers of S, which appear[4], e.g., for $F(x) = x^{-1}$ or $F(x) = x^{-\frac{1}{2}}$.

In Ref. 1 it is shown how one can obtain closed expressions for Δ^{-1} and $\Delta^{-\frac{1}{2}}$ in the two simplest cases:

* The research reported in this paper was sponsored in part by the King Gustaf VI Adolf's 70-Years Fund for Swedish Culture, Knut and Alice Wallenberg's Foundation, The Swedish Natural Science Research Council, and in part by the Aeronautical Research Laboratory, OAR, through the European Office, Aerospace Research, U. S. Air Force.

[1] P. O. Löwdin, R. Pauncz, and J. de Heer, J. Math. Phys. 1, 461 (1960).

[2] T. L. Gilbert, J. Math. Phys. 3, 107 (1962).

[3] P. B. Abraham and G. Weiss, J. Math. Phys. 3, 340 (1962).

[4] P. O. Löwdin, Advan. Phys. 5, 1 (1956), particularly pp. 46 ff and 89 ff.

$$\mathbf{\Delta} = (1, S_1, 0, \cdots 0, S_1), \tag{7}$$

$$\mathbf{\Delta} = (1, S_1, S_2, 0, \cdots 0, S_2, S_1). \tag{8}$$

This is very important, since ordinary procedures for matrix inversion are not applicable when n, the order of $\mathbf{\Delta}$, is very large. It would definitely be desirable with such closed formulas also in the general case (1). The difficulties in such a generalization are discussed in the next section.

Although the methods discussed here are valid for any cyclic matrix, we confine our interest to the case where $\mathbf{\Delta}$ is an overlap matrix in an LCAO calculation for a crystal. Then the elements of $\mathbf{\Delta}$ are so-called overlap integrals between atomic orbitals centered at different nuclei in a crystal lattice,

$$S_\mu = \int \phi^*(\mathbf{r})\phi(\mathbf{r} - \mathbf{r}_\mu)\, d\tau. \tag{9}$$

The matrix $\mathbf{\Delta}^{-1}$ appears in, e.g., the expression for the cohesive energy. Particularly for metals, where the overlap integrals (9) are quite large, it is essential to include $\mathbf{\Delta}^{-1}$ in the energy calculations.

The eigenvalues (2), of an overlap matrix may be interpreted as normalization integrals[4]; i.e., they are always positive. In other words, the matrix $\mathbf{\Delta}$ is positive-definite. This introduces restrictions on the absolute and relative magnitudes of the S_i's. Particularly in numerical calculations it is important to remember that, although the complete overlap matrix associated with a lattice is positive-definite, it can happen that the truncated part of it which is used in a numerical application is not (or "nearly not") positive definite. For particular cases (only S_1, or only S_1 and S_2 included), one can express these restrictions in a nice closed form; in the general case less explicit methods must be used.

Although we are naturally most interested in real, three-dimensional crystals, it is valuable to study at first a one-dimensional model. This is done in Refs. 1, 2, and 3. The three-dimensional case is treated in Ref. 3.

Gilbert[2] introduces the overlap function

$$\mathbb{S}(z) = \sum_{l=-\infty}^{\infty} S_l z^l, \tag{10}$$

which is of essential importance for understanding the properties of $\mathbf{\Delta}^{-1}$. If $S_{-l} = S_l$, we can rewrite (10):

$$\mathbb{S}(z) = 1 + \sum_{l=1}^{\infty} S_l(z^l + z^{-l}) = \sum_{l=0}^{\infty} S_l C_l(y), \tag{11}$$

where

$$S_0 = \tfrac{1}{2}; \qquad y = z + z^{-1}, \tag{12}$$

and the C_n's are Chebyshev polynomials of the first kind, defined by

$$C_0(y) = 2; \qquad C_1(y) = y; \qquad C_{n+1}(y) = yC_n(y) - C_{n-1}(y), \tag{13}$$

or

$$C_n(y) = 2\cos n\vartheta = 2\cos n(\arccos \tfrac{1}{2}y); \qquad y = 2\cos\vartheta; \qquad z = e^{i\vartheta}. \tag{14}$$

For the one-dimensional model Gilbert expresses the elements of $\mathbf{\Delta}^{-1}$ as contour integrals, which are easily evaluated. In his formulation the problem essentially reduces to that of finding the zeros of a polynomial. We see later, how this same problem reappears disguised in other methods.

Gilbert also discusses how the restrictions on the S_k's are related to the analytic properties of the overlap function (10). It would definitely be desirable to proceed along these lines; a detailed study of the overlap function associated with a three-dimensional lattice would give us a much better understanding of the problem and might perhaps even lead to simple, closed formulas for $\mathbf{\Delta}^{-1}$ in the general case (1), similar to those derived in the special cases in Ref. 1.

Gilbert's method seems to be restricted, at least for practical calculations, to the one-dimensional case. For two and three dimensions he proposes an iterative version of Löwdin's power series method.[4]

Abraham and Weiss[3] use formula (6) to calculate various functions of $\mathbf{\Delta}$. For the special case of the inverse they obtain a closed formula, which contains the zeros of the overlap function, i.e., their result is essentially the same as Gilbert's. They also discuss how the procedure is extended to two and three dimensions, although they do not give any examples of this.

It is the purpose of the present note (a) to discuss the difficulties in generalizing the methods of Ref. 1, from the special cases, (7), (8), to the general case (1); (b) to show the connection between the method used by Abraham and Weiss and those of Löwdin, Pauncz, and de Heer; (c) to adapt formula (6) to practical numerical calculations both for one and three dimensions.

2. THE ONE-DIMENSIONAL CASE

The most important result in the paper of Löwdin, Pauncz, and de Heer[1] (from now on abbreviated LPH) is probably the fact that, at least for some cases, it is possible to obtain closed explicit formulas for the elements of $\mathbf{\Delta}^{-1}$ and $\mathbf{\Delta}^{-\frac{1}{2}}$, for any order of the matrix $\mathbf{\Delta}$. They succeeded in deriving these

formulas in three different ways by using various properties of the Chebyshev polynomials of both kinds.

Here we derive another expression for $\mathbf{\Delta}^{-1}$ in the special case (8) (S_1 and S_2 only), and combine that with the result of LPH for case (7) (S_1 only), to get an expression for $\mathbf{\Delta}^{-1}$ in the general case (1). We also derive another expression for $\mathbf{\Delta}^{-1}$ in the general case and show its connection with (5).

The so-called topological matrices, $\mathbf{M}_i$, (for a definition, see Ref. 1), satisfy the defining relations of the Chebyshev polynomials of the first kind (13), so that we have the important result

$$\mathbf{M}_p = C_p(\mathbf{M}_1). \tag{15}$$

Because of that relation we can work with functions of a real or complex variable, instead of functions of matrices. We can, e.g., define an overlap function for the case (1), when we include overlap to order n:

$$s_n(z) = \sum_{l=-[\frac{1}{2}n]}^{[\frac{1}{2}n]-1} S_l z^l. \tag{16}$$

Combining (15) with (11) we can write

$$s_n(z) = \sum_{l=0}^{[\frac{1}{2}n]} S_l C_l(y) = s_n(y), \tag{17}$$

$$s_n(\mathbf{M}_1) = \sum_{l=0}^{[\frac{1}{2}n]} S_l C_l(\mathbf{M}_1) = \sum_{l=0}^{[\frac{1}{2}n]} S_l \mathbf{M}_l = \mathbf{\Delta}. \tag{18}$$

One way to construct a function $F(\mathbf{\Delta})$, would be the following. Find an expansion of $F[s_n(y)]$ in terms of the Chebyshev polynomials:

$$F[s_n(y)] = \sum_p a_p C_p(y). \tag{19}$$

Then, due to (15), we get directly

$$F(\mathbf{\Delta}) = F[s_n(\mathbf{M}_1)] = \sum_p a_p C_p(\mathbf{M}_1)$$

$$= \sum_p a_p \mathbf{M}_p = (2a_0, a_1, a_2, a_3, \cdots)_{\text{cyclic}}. \tag{20}$$

After these preliminaries we consider the special case (8):

$$\mathbf{\Delta} = (1, S_1, S_2, 0, \cdots 0, S_2, S_1)_{\text{cyclic}}$$

$$= \mathbf{1} + S_1\mathbf{M}_1 + S_2\mathbf{M}_2. \tag{8}$$

To obtain $\mathbf{\Delta}^{-1}$, LPH wrote $\mathbf{\Delta}$ as a product and calculated the inverse of each one of the two commuting factors. Here we instead follow a suggestion by Löwdin[5] and decompose $\mathbf{\Delta}^{-1}$ in "partial fractions":

$$\mathbf{\Delta}^{-1} = S_2^{-1}(a_1 \cdot \mathbf{1} + \mathbf{M}_1)^{-1}(a_2 \cdot \mathbf{1} + \mathbf{M}_1)^{-1}$$

$$= \frac{1}{S_2(a_2 - a_1)} [(a_1 \cdot \mathbf{1} + \mathbf{M}_1)^{-1} - (a_2 \cdot \mathbf{1} + \mathbf{M}_1)^{-1}], \tag{21}$$

[5] P. O. Löwdin (private communication).

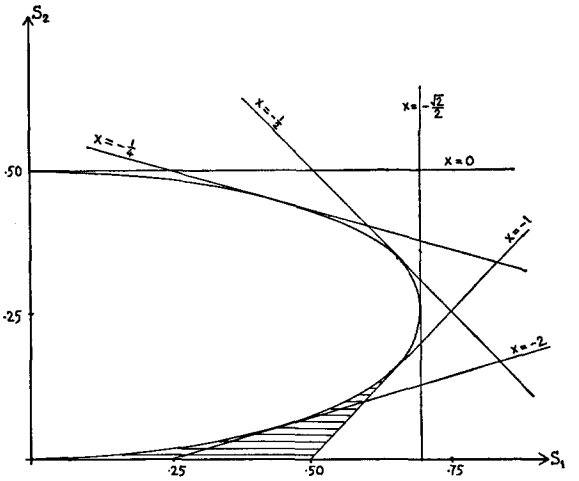


FIG. 1

where

$$a_1 = a_2^* = 2(\cos u \cosh v - i \sin u \sinh v);$$

$$\cos u = \frac{1}{4S_2^{\frac{1}{2}}} [(1 + 2S_1 + 2S_2)^{\frac{1}{2}} - (1 - 2S_1 + 2S_2)^{\frac{1}{2}}],$$

$$\cosh v = \frac{1}{4S_2^{\frac{1}{2}}} [(1 + 2S_1 + 2S_2)^{\frac{1}{2}} + (1 - 2S_1 + 2S_2)^{\frac{1}{2}}] \tag{22}$$

[It should be noticed, that (22) differs from LPH's formula (72) by a factor 2 in the denominator.] On the same page of LPH there are also two other misprints: formula LPH (69) should read

$$\mathbf{\Delta} = \mathbf{1} + S_1\mathbf{M}_1 + S_2(\mathbf{M}_1^2 - 2 \cdot \mathbf{1})$$

$$= S_2(a_1 \cdot \mathbf{1} + \mathbf{M}_1)(a_2 \cdot \mathbf{1} + \mathbf{M}_1),$$

and below formula (78) the two δ's should be

$$\delta_0 = S_1^{-\frac{1}{2}}\tfrac{1}{2}A_0; \qquad \delta_l = (-1)^l A_l S_l^{-\frac{1}{2}}.$$

LPH have given a closed formula for $(a \cdot \mathbf{1} + \mathbf{M}_1)^{-1}$ for $a > 2$. In that case the elements of the inverse decrease exponentially, whereas they oscillate for $a \leq 2$. By extending their method to complex a's, we get, after some algebraic manipulations, for our case (8),

$$\Delta_{0\mu}^{-1} = \frac{\operatorname{Im}\{(a_1 \cdot \mathbf{1} + \mathbf{M}_1)^{-1}_{0\mu}\}}{2S_2 \sin u \sinh v}$$

$$= (-1)^\mu \frac{e^{-\mu v}}{4S_2(\cosh^2 v - \cos^2 u)}$$

$$\times (\sin \mu u \cot u + \cos \mu u \coth v). \tag{23}$$

This formula is valid if the point (S_1, S_2) is inside

the half-ellipse

$$\frac{S_1^2}{\frac{1}{2}} + \frac{(S_2 - \frac{1}{4})^2}{\frac{1}{16}} = 1 \tag{24}$$

in the (S_1, S_2) plane; (see Fig. 1). There is another region in the (S_1, S_2) plane, (shaded in the figure), where $\mathbf{\Delta}$ is also positive-definite. In that region we can use LPH's formula (16) for each one of the two inverse matrices in (21).[6]

With (23) and LPH (16), we can in principle calculate $\mathbf{\Delta}^{-1}$ in the general case, (1), $(n = 2\nu)$:

$$\mathbf{\Delta} = (1, S_1, S_2, S_3, \cdots S_3, S_2, S_1) = \sum_{r=0}^{\nu} S_r \mathsf{M}_r, \tag{25}$$

where we define S_0 as $\frac{1}{2}$. Again the problem is reduced to finding all the real and complex zeros of the overlap function

$$s_\nu(y) = \sum_{r=0}^{\nu} S_r C_r(y). \tag{26}$$

Then $1/s_\nu(y)$ can be expanded in partial fractions, analogously to (21). For each real root we use LPH (16), and for each complex pair, (23).

There are two reasons to consider also another method for the general case (25). First of all, a fast computer, which is desirable to get all the zeros of the overlap function, is not always available. Secondly, we want to connect the methods, which use Chebyshev polynomials, with the general formula (6).

To that end we use the fact that the Chebyshev polynomials of the first kind, defined by (13), (14), form an orthogonal set in the interval $(-2, 2)$, i.e., they satisfy the relations

$$\int_{-2}^{2} \frac{C_k(x)C_l(x)}{(1 - \frac{1}{4}x^2)^{\frac{1}{2}}}\, dx = \begin{cases} 0 & \text{for } k \neq l, \\ 4\pi & \text{for } k = l \neq 0, \\ 8\pi & \text{for } k = l = 0. \end{cases} \tag{27}$$

The inverse of the overlap function (26) can be expanded in these polynomials:

$$\frac{1}{s_\nu(y)} = \sum_{k=0}^{\infty} \alpha_k C_k(y) \tag{28}$$

on one condition; namely that $s_\nu(y)$ does not have any zeros in the interval $(-2, 2)$. We see how the restrictions on the S_i's now appear in a different way. The coefficients in the expansion (28) are given directly by means of (27):

$$\alpha_0 = \frac{1}{8\pi} \int_{-2}^{2} \frac{C_0(y)}{s_\nu(y)(1 - \frac{1}{4}y^2)^{\frac{1}{2}}}\, dy = \frac{1}{2\pi} \int_0^\pi \frac{1}{\sigma_\nu(\vartheta)}\, d\vartheta,$$

$$\alpha_k = \frac{1}{4\pi} \int_{-2}^{2} \frac{C_k(y)}{s_\nu(y)(1 - \frac{1}{4}y^2)^{\frac{1}{2}}}\, dy = \frac{1}{\pi} \int_0^\pi \frac{\cos k\vartheta}{\sigma_\nu(\vartheta)}\, d\vartheta, \tag{29}$$

where $s_\nu(y) = \sigma_\nu(\vartheta)$, $y = 2\cos\vartheta$. The integrands can be transformed to rational functions, so that the α_k's can be calculated exactly. In practice, however, we prefer to evaluate them by means of numerical integration (see Sec. 4).

It now remains to connect this with (5). Comparing (2), (14), and (17), we see that the eigenvalues of $\mathbf{\Delta}$ are given as

$$d_k = (\mathbf{U}^\dagger \mathbf{\Delta} \mathbf{U})_{kk} = \sum_{\mu=0}^{\nu} 2S_\mu \cos\frac{2\pi k\mu}{n} = \sigma_\nu\left(\frac{2\pi k}{n}\right). \tag{30}$$

Here, as before, we set $S_0 = \frac{1}{2}$. Then, using (5), we get

$$\begin{aligned} \Delta_{0k}^{-1} &= (\mathbf{U}\mathsf{d}^{-1}\mathbf{U}^\dagger)_{0k} \\ &= n^{-1} \sum_{l,m} 1 \cdot \left[\sigma_\nu\left(\frac{2\pi l}{n}\right)\right]^{-1} \delta_{lm} \cdot e^{-2\pi i m k/n} \\ &= \frac{1}{n} \sum_m \frac{e^{-2\pi i m k/n}}{\sigma_\nu(2\pi m/n)}. \end{aligned} \tag{31}$$

When n is a very large number, this sum can be calculated as an integral $(2\pi m/n = \vartheta)$:

$$\begin{aligned} \Delta_{0k}^{-1} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{-ik\vartheta}}{\sigma_\nu(\vartheta)}\, d\vartheta \\ &= \begin{cases} \dfrac{1}{2\pi} \displaystyle\int_0^\pi \frac{1}{\sigma_\nu(\vartheta)}\, d\vartheta & \text{for } k = 0, \\ \dfrac{1}{\pi} \displaystyle\int_0^\pi \frac{\cos k\vartheta}{\sigma_\nu(\vartheta)}\, d\vartheta & \text{for } k \neq 0, \end{cases} \end{aligned} \tag{32}$$

which means that we have derived (29) in another way.

It is easily checked that in the special case (8), for which the integral (32) can be calculated in closed form, we get the form (23). We must remember that both here and in (23) we consider the asymptotic case. The general case (finite n) is obtained by converting the infinite expansion (28) (or rather the corresponding one for $\mathbf{\Delta}^{-1}$), into a finite series, according to the prescriptions given by LPH in their "method C".

In principle, the number of integrals, (29), to be calculated, can be restricted to q, if $q + 1$ overlap integrals, $S_0, S_1, \cdots S_q$ are included. The remaining coefficients $\alpha_{q+1}, \alpha_{q+2} \cdots$ can be obtained from a

[6] For a detailed discussion, see J.-L. Calais and K. Appel, Technical Note 96, Quantum Chemistry Group, Uppsala, Sweden (1963) (unpublished).

recursion relation

$$\alpha_r + S_1(\alpha_{r-1} + \alpha_{r+1}) + S_2(\alpha_{r-2} + \alpha_{r+2}) + \cdots + S_q(\alpha_{r-q} + \alpha_{r+q}) = 0, \quad r = 1, 2, 3, \cdots, \tag{33}$$

obtained from the basic equation $\mathbf{A}\mathbf{A}^{-1} = 1$.[7] Unfortunately this relation is numerically very unstable. In practice, therefore, it is preferable to calculate the α_k's needed directly by integration.

3. THE THREE-DIMENSIONAL CASE

The overlap matrix associated with a set of atomic orbitals, centered at different sites in a three-dimensional lattice, is cyclic just as the corresponding matrix for a one-dimensional chain. One has to be a little careful, however, in adapting the methods previously described, to the three-dimensional case.

It seems to be difficult to generalize the method involving the zeros of a polynomial. One can certainly number the atoms in the crystal "in a one-dimensional way," but then the matrix loses its cyclic property.

In order to take advantage of all the symmetry of the lattice, the matrix elements have to be arranged in such a way, that the cyclic property in all three dimensions appears. One way of achieving that, is the following. First, consider one atomic plane in the crystal. The atoms in that plane are labeled 0, 1, 2, $\cdots$, $N-1$ in the first row, N, $N+1$, $\cdots$ $2N-1$ in the second row, etc., up to the last row: $N(N-1)$, $N(N-1)+1$, $\cdots$, N^2-1. To each row corresponds a set of topological matrices, M_k, of the type considered in Sec. 2. The whole plane corresponds to a supermatrix with these M_k's as parts. The two-dimensional cyclic property is seen in the M_k's, which are cyclic, and in the fact that the supermatrix is cyclic with respect to these M_k's. After this it is easy to go from two to three dimensions: we just form another supermatrix, the blocks of which are the "plane supermatrices".

Analogous to (18) we can now write $\mathbf{A}$ as a function of three basic supermatrices. And as in (15) any supermatrix can be expressed as a three-dimensional Chebyshev polynomial of these basic supermatrices—something like

$$\mathfrak{M}_{klm} = C_k(\mathfrak{M}_{100})C_l(\mathfrak{M}_{010})C_m(\mathfrak{M}_{001}). \tag{34}$$

The elements of $\mathbf{A}^{-1}$ could be calculated as triple integrals—analogous to (29):

$$\alpha_{klm} = \frac{1}{8\pi^3}\int_0^\pi\int_0^\pi\int_0^\pi \frac{\cos k\vartheta \cos l\varphi \cos m\chi}{\sigma(\vartheta,\varphi,\chi)}\, d\vartheta\, d\varphi\, d\chi. \tag{35}$$

We have seen, that in one dimension the corresponding result can be derived in two ways: either we consider the α_k's as coefficients of the C_k's in the expansion of $[s_r(y)]^{-1}$, (28), or we carry out the two unitary transformations

$$\mathbf{A}^{-1} = \mathsf{U}(\mathsf{U}^\dagger\mathbf{A}\mathsf{U})^{-1}\mathsf{U}^\dagger, \tag{36}$$

as in (30), (31). Since in the practical calculations, we only consider a limited number of neighbors around a particular atom, the labeling described above for the supermatrices is not very practical. We therefore finally give a derivation of $\mathbf{A}^{-1}$, using (36) and leading to integrals of type (35).

We assume as usual that the crystal contains G^3 atoms and that it fulfills the Born–von Kármán condition. Analogous to (30) and (31) we get

$$d(\mathbf{k}) = \sum_{\mathbf{m}}^{G^3} S(\mathbf{m})e^{2\pi i\mathbf{k}\cdot\mathbf{m}}; \tag{37}$$

$$\Delta^{-1}(0,\boldsymbol{\lambda}) = \frac{1}{G^3}\sum_{\mathbf{k}}\frac{e^{-2\pi i\mathbf{k}\cdot\boldsymbol{\lambda}}}{d(\mathbf{k})}. \tag{38}$$

Here as usual, $\mathbf{k}$ is a vector in the reciprocal space. (We use the notations of Ref. 4, Sec. 4.) Since the $\mathbf{k}$'s form a quasicontinuous set, we can write $\Delta^{-1}(0,\boldsymbol{\lambda})$ as an integral over the first Brillouin zone,

$$\Delta^{-1}(0,\boldsymbol{\lambda}) = V_{0a}\int \frac{e^{-2\pi i\mathbf{k}\cdot\boldsymbol{\lambda}}}{d(\mathbf{k})}\, d\mathbf{k}. \tag{39}$$

If $S(\mathbf{m}) = S(-\mathbf{m})$ we can write the denominator as

$$d(\mathbf{k}) = 1 + 2\sum_{\mathbf{m}} S(\mathbf{m})\cos(2\pi\mathbf{k}\cdot\mathbf{m}). \tag{40}$$

This function, $d(\mathbf{k})$ is even under inversion, ($\mathbf{k} \to -\mathbf{k}$), whereas the imaginary part of the numerator, $\sin 2\pi\mathbf{k}\cdot\boldsymbol{\lambda}$, is odd under inversion. Therefore the integral can be written as

$$\Delta^{-1}(0,\boldsymbol{\lambda}) = V_{0a}\int \frac{\cos 2\pi\mathbf{k}\cdot\boldsymbol{\lambda}}{d(\mathbf{k})}\, d\mathbf{k}. \tag{41}$$

If, further, $S(\mathbf{m})$ depends only on the distance m, we can simplify (41) even more. This is the case, when S denotes the overlap integral between two atomic s functions. If $S(\mathbf{m}) = S(m)$, we can add all the cos terms corresponding to the same m. This gives rise to different types of terms, depending on the class of neighbor $\mathbf{m}$ represents. The table below shows the reduction for the typical cases in a cubic crystal. The letters a, b, c denote the components

[7] Compare Ref. 4 p. 91.

TABLE I. Reduction for typical cubic-crystal cases.

Class	m^2	Corresponding factor of $2S(m)$ in $d(\mathbf{k})$
1	$a^2+b^2+c^2$	4 cos ax cos by cos cz + 5 analogous terms, obtained by permuting (a, b, c)
2	$a^2+a^2+b^2$	4 cos ax cos ay cos bz + 4 cos ax cos by cos az + 4 cos bx cos ay cos az
3	$a^2+a^2+a^2$	4 cos ax cos ay cos az
4	a^2+b^2+0	2 cos ax cos by + 5 analogous terms
5	a^2+a^2+0	2 cos ax cos ay + 2 cos ax cos az + 2 cos ay cos az
6	a^2+0+0	cos ax + cos ay + cos az

of $\mathbf{m}$: $\mathbf{m} = a\mathbf{a}_1 + b\mathbf{a}_2 + c\mathbf{a}_3$; each case represents a particular distance $m = (a^2+b^2+c^2)^{\frac{1}{2}}$; ($2\pi k_1 = x$, etc.).

In this case, $S(\mathbf{m}) = S(m)$, $d(\mathbf{k})$ is isotropic, and therefore we also see that, e.g.,

$$\Delta^{-1}(\mathbf{0}, a\mathbf{a}_1) = \Delta^{-1}(\mathbf{0}, a\mathbf{a}_2) = \Delta^{-1}(\mathbf{0}, a\mathbf{a}_3). \quad (42)$$

Similar relations, of course hold for all equivalent points $\boldsymbol{\lambda}$.

The integration region to be used for (41) is the first Brillouin zone. For many lattice types the corresponding integration limits are rather complicated. We can however, simplify the situation considerably, by integrating over an equivalent volume in reciprocal space. If $\mathbf{K}$ is a lattice vector in reciprocal space, i.e., such that $\mathbf{K}\cdot\boldsymbol{\lambda}$ = integer, when $\boldsymbol{\lambda}$ is a lattice vector in ordinary space, we have

$$\cos 2\pi(\mathbf{k}+\mathbf{K})\cdot\boldsymbol{\lambda} = \cos 2\pi\mathbf{k}\cdot\boldsymbol{\lambda}, \quad (43)$$

and therefore

$$\frac{\cos 2\pi(\mathbf{k}+\mathbf{K})\cdot\boldsymbol{\lambda}}{d(\mathbf{k}+\mathbf{K})} = \frac{\cos 2\pi\mathbf{k}\cdot\boldsymbol{\lambda}}{d(\mathbf{k})}. \quad (44)$$

As an example, we transform (41) for the three cubic lattices.

A. Simple Cubic Lattice

If the unit cell has the edge a, we have $V_{0a} = a^3$. The first Brillouin zone is a cube with edge $1/a$. With the transformation

$$2\pi a\mathbf{k} = \boldsymbol{\vartheta}; \quad d(\mathbf{k}) = \sigma(\boldsymbol{\vartheta}), \quad \boldsymbol{\lambda} = a(\lambda_x, \lambda_y, \lambda_z), \quad (45)$$

we get

$$\Delta^{-1}(\mathbf{0}, \boldsymbol{\lambda}) = \frac{1}{8\pi^3}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\frac{\cos(\vartheta_x\lambda_x+\vartheta_y\lambda_y+\vartheta_z\lambda_z)}{\sigma(\boldsymbol{\vartheta})}\,d\boldsymbol{\vartheta}$$

$$= \frac{1}{\pi^3}\int_0^{\pi}\int_0^{\pi}\int_0^{\pi}\frac{\cos(\vartheta_x\lambda_x+\vartheta_y\lambda_y+\vartheta_z\lambda_z)}{\sigma(\boldsymbol{\vartheta})}\,d\boldsymbol{\vartheta}. \quad (46)$$

B. Body-Centered Cubic Lattice

The first Brillouin zone is a dodecahedron with volume $2/a^3$, when the nearest-neighbor distance is $a\,\frac{1}{2}\sqrt{3}$.[8] $V_{0a} = \frac{1}{2}a^3$. On account of (44), we can as well integrate over the two cubes,

$$-\frac{1}{a} < k_x \le \frac{1}{a}; \quad 0 < k_y \le \frac{1}{a}; \quad 0 < k_z \le \frac{1}{a}. \quad (47)$$

With the transformation

$$\pi a\mathbf{k} = \boldsymbol{\vartheta}; \quad d(\mathbf{k}) = \sigma(\boldsymbol{\vartheta}); \quad \boldsymbol{\lambda} = \tfrac{1}{2}a(\lambda_x, \lambda_y, \lambda_z), \quad (48)$$

we get

$$\Delta^{-1}(\mathbf{0}, \boldsymbol{\lambda}) = \frac{1}{2\pi^3}\int_{-\pi}^{\pi}\int_0^{\pi}\int_0^{\pi}\frac{\cos\left(\frac{2}{a}\boldsymbol{\vartheta}\cdot\boldsymbol{\lambda}\right)}{\sigma(\boldsymbol{\vartheta})}\,d\boldsymbol{\vartheta}$$

$$= \frac{1}{\pi^3}\int_0^{\pi}\int_0^{\pi}\int_0^{\pi}\frac{\cos\vartheta_x\lambda_x\cos\vartheta_y\lambda_y\cos\vartheta_z\lambda_z}{\sigma(\boldsymbol{\vartheta})}\,d\boldsymbol{\vartheta}. \quad (49)$$

The second form in (49) is correct in the case $S(\mathbf{m}) = S(m)$.

C. Face-Centered Cubic Lattice

In this case, the first Brillouin zone is a dodecahedron with squares and hexagons as faces. If the nearest-neighbor distance is $a\,\frac{1}{2}\sqrt{2}$, its volume is $4/a^3$; (i.e., $V_{0a} = \frac{1}{4}a^3$); with (48) we get

$$\Delta^{-1}(\mathbf{0}, \boldsymbol{\lambda}) = \frac{1}{4\pi^3}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_0^{\pi}\frac{\cos\left(\frac{2}{a}\boldsymbol{\vartheta}\cdot\boldsymbol{\lambda}\right)}{\sigma(\boldsymbol{\vartheta})}\,d\boldsymbol{\vartheta} \quad (50)$$

$$= \frac{1}{\pi^3}\int_0^{\pi}\int_0^{\pi}\int_0^{\pi}\frac{\cos\vartheta_x\lambda_x\cos\vartheta_y\lambda_y\cos\vartheta_z\lambda_z}{\sigma(\boldsymbol{\vartheta})}\,d\boldsymbol{\vartheta}. \quad (51)$$

Again, this second form is correct for $S(\mathbf{m}) = S(m)$.

Finally, we should notice that this method can very easily be adapted to calculate $\boldsymbol{\Delta}^{-\frac{1}{2}}$:

$$\Delta^{-\frac{1}{2}}(\mathbf{0}, \boldsymbol{\lambda}) = V_{0a}\int\frac{\cos(2\pi\mathbf{k}\cdot\boldsymbol{\lambda})}{[d(\mathbf{k})]^{\frac{1}{2}}}\,d\mathbf{k}. \quad (52)$$

In this connection it is appropriate to recall that in 1937 Wannier[9] calculated $\boldsymbol{\Delta}^{-\frac{1}{2}}$ in essentially this way, and gave a closed expression for it in the special case that $\Delta(\mathbf{m}, \mathbf{n})$ is a simple exponential.

4. NUMERICAL PROCEDURES

Since the method involving the zeros of a polynomial does not seem to work for three dimensions, we here consider only the other method, (which can be derived in two ways), where we have to calculate integrals of type (32) or (41). The one-dimensional integral (32) can be integrated exactly

[8] See, e.g., N. F. Mott and H. Jones, *The Theory of the Properties of Metals and Alloys* (Dover Publications, Inc., New York, 1958), p. 156 ff.

[9] G. H. Wannier, Phys. Rev. **52**, 191 (1937).

TABLE II.
Elements of $\mathbf{A}$ and $\mathbf{A}^{-1}$ in the body-centered cubic lattice for varying numbers of neighbors and integration points.

r	$S(r)$	(a) $\frac{13}{8}$	(b) $\frac{19}{8}$	(c) $\frac{19}{9}$	(d) $\frac{18}{10}$	(e) $\frac{32}{11}$
000	1.0000000	1.639993	1.639282	1.639283	1.639283	1.639314
111	0.2890106	−0.274157	−0.274037	−0.274037	−0.274037	−0.274043
200	0.2105272	−0.067761	−0.067327	−0.067327	−0.067327	−0.067359
220	0.0716946	0.067930	0.068107	0.068107	0.068107	0.068105
311	0.0361180	0.020354	0.019883	0.019882	0.019882	0.019922
222	0.0291945	0.032055	0.031619	0.031619	0.031619	0.031638
400	0.0132246	−0.001048	−0.001606	−0.001603	−0.001603	−0.001595
331	0.0076727	−0.012493	−0.012666	−0.012668	−0.012668	−0.012706
420	0.0064486	−0.008484	−0.008097	−0.008094	−0.008094	−0.008103
422	0.0033229	−0.004038	−0.002915	−0.002912	−0.002912	−0.002983
333	0.0020804	−0.003265	−0.002777	−0.002779	−0.002779	−0.002715
511	0.0020804	−0.000789	0.000141	0.000134	0.000133	0.000041
440	0.0009972	0.001072	0.003385	0.003387	0.003387	0.003499
531	0.0006563		0.001885	0.001885	0.001884	0.001894
442	0.0005728		0.001957	0.001958	0.001958	0.002024
600	0.0005728		−0.000114	−0.000108	−0.000106	0.000191
620	0.0003375		0.000220	0.000212	0.000215	0.000393
533	0.0002303		0.000134	0.000137	0.000136	0.000266
622	0.0002033		−0.000146	−0.000160	−0.000157	−0.000074
444	0.0001248					0.000096
551	0.0000876					−0.000787
711	0.0000876					−0.000090
640	0.0000780					−0.000616
642	0.0000495					−0.000350
553	0.0000355					−0.000273
731	0.0000355					−0.000126
800	0.0000207					0.000002
733	0.0000152					0.000044
644	0.0000137					0.000000
820	0.0000137					0.000001
660	0.0000091					0.000213
822	0.0000091					0.000031

by means of the transformation $\tan \frac{1}{2}\vartheta = z$. That is, however, equivalent to calculating all the zeros of a polynomial, and therefore rather cumbersome. In the three-dimensional case (41), we cannot integrate exactly. In both cases it is, therefore, natural to use numerical integration.

We seem to have the choice between two integration formulas for this kind of integrals: Chebyshev–Radau formula,[10]

$$\int_0^\pi f(\cos\vartheta)\cos\lambda\vartheta\, d\vartheta = H\sum_{i=1}^{n}\sum_{k=0}^{2\lambda-1}(-1)^k f\left(\cos\frac{a_i + k\pi}{\lambda}\right), \qquad (53)$$

and a particular Gauss–Mehler formula[10]

$$\int_0^\pi f(\cos\vartheta)\, d\vartheta = \frac{\pi}{n}\sum_{j=1}^{n} f\left(\cos\frac{2j-1}{2n}\pi\right). \qquad (54)$$

The weight H, and the abscissas a_i of (53) are given in Kopal[10] for $n = 1, 2, 3$, by Greenwood *et al.*[11] for $n = 1 - 4$, and by us[12] for $n = 5, 6$.

The Gauss–Mehler formula has the advantage of greater simplicity, whereas one would expect that a smaller number of points would suffice for a comparable accuracy in the Chebyshev–Radau formula. Actually it turned out that a triple Gauss–Mehler integration over $K + 3$ or $K + 4$ points, where $\cos K\vartheta$ is the highest term occurring in the intergrand, has given sufficient accuracy, whereas at least $2K$ points are necessary in Chebyshev–Radau integration.

The use of a standard computer subroutine for Gauss–Mehler integration would involve a large duplication of work, since the denominators $\sigma(\boldsymbol{\vartheta})$ are common in the integrands for various $\mathbf{k}$. In programming the present calculation for the Alwac III-E computer of the Quantum Chemistry Group, all integrations were therefore performed simultaneously, computing the contributions to all integrals from one point in $\boldsymbol{\vartheta}$ space before proceeding to the next point.

Initially, a recursion formula was used to form $\cos K\vartheta$ for various multiples of ϑ. It turned out, however, that this gave poor accuracy when repeated several steps for small ϑ, and instead the cosine routine was called independently for each K.

To check the convergence of this method, both

[10] Z. Kopal, *Numerical Analysis* (Chapman and Hall Ltd., London, 1955), pp. 384 and 426.
[11] R. E. Greenwood, P. D. M. Carnahan, and J. W. Nolley, Math. Tables Aids Comput. **13**, 37 (1959).
[12] K. Appel, J.-L. Calais, and F. Sasaki (unpublished).

with respect to order of integration and to including neighbors at larger distances, the program was applied to a body-centered cubic lattice with overlap integrals

$$S(\mathbf{r}) = (1 + \eta r + \tfrac{1}{3}\eta^2 r^2)e^{-\eta r}, \qquad (55)$$

corresponding to hydrogen 1s functions, with $\eta r = 3.32$ for nearest neighbors.[13] The results are shown in Table II. The first column of Table II shows the components of the vector $\mathbf{r}$ in the ordinary lattice, arranged according to increasing distance from the origin. In the second column, we give the corresponding overlap integrals, calculated by means of (55). The remaining columns, (a), (b), (c), (d), and (e) give the elements of $\mathbf{\Delta}^{-1}$ in various approximations. The first number gives the number of types of neighbors included, and the second one the number of points used in the Mehler integration. E.g., $\frac{19}{9}$ in column (c), means that neighbors up to order 19 are included and that the corresponding integrals are calculated with a 9-point formula.

[13] W. J. Carr, Jr., Phys. Rev. 128, 120 (1962).

The three columns (b), (c), (d) give an idea of the rate of convergence when the number of points of the integration formula is increased. It is also worthwhile to notice, how the elements of $\mathbf{\Delta}^{-1}$ are influenced when more and more neighbors are taken into account [columns (a), (d), (e)].

CONCLUSION

Although we still do not have closed formulas in the general case, (arbitrarily many S_k's included), the procedures described at least make it possible to treat that case numerically in a simple way. Even with a relatively slow computer, the time needed is quite reasonable.

ACKNOWLEDGMENTS

It is a pleasure to thank professor P. O. Löwdin for suggesting this problem as well as for valuable comments. Special thanks are also due to *fil. lic.* Jan Linderberg for many interesting discussions.

Ising-Model Spin Correlations on the Triangular Lattice

JOHN STEPHENSON
Wheatstone Physics Laboratory, King's College, London, England.
(Received 28 January 1964)

A Pfaffian representation, of the partition function of the triangular lattice is used to derive expressions for various two, four, and six spin correlations in terms of Pfaffians. The pair correlations along a diagonal are expressed as a Toeplitz determinant whose limiting form yields the spontaneous magnetization. At the ferromagnetic critical point the correlations decay as $1/r^{\frac{1}{4}}$ with approximately radial symmetry. At the antiferromagnetic zero point the ground state is highly degenerate—it has finite entropy—and on a given sublattice the pair correlations along a row decay as $\epsilon/r^{\frac{1}{4}}$, where $\epsilon = +\epsilon_0$ on the sublattice containing the origin spin and $\epsilon \simeq -\frac{1}{2}\epsilon_0$ on the other two sublattices. Finally, the perpendicular susceptibility, $\chi_\perp$, which depends on a finite number of correlations, is calculated; its ferromagnetic behavior is similar to that of the perpendicular susceptibilities of the quadratic and honeycomb lattices, but for an antiferromagnet $\chi_\perp$ diverges as $1/T$ at low temperatures.

1. INTRODUCTION

ONE of the most interesting aspects of a two-dimensional Ising ferromagnet in zero magnetic field[1,2] is that the disordered spin system occurring at high temperatures undergoes a cooperative transition at a certain critical temperature (the Curie point) and enters an ordered state having spontaneous magnetic moment. The detailed thermodynamic properties of the quadratic lattice were first elucidated by Onsager[3]; and Onsager and Kaufman[4,5] showed how the short-range correlations could be evaluated. More recently, Montroll, Potts, and Ward[6] have rederived the formulas for the correlations by simpler methods: they employ the Pfaffian representation of the partition function introduced by Kasteleyn.[7] They also obtain the well-known result for the spontaneous magnetization[8,9] which is related to the long-range order parameter. The formula for the magnetization can be written in the form

$$M = (1 - k_1^2)^{\frac{1}{8}}, \tag{1.1}$$

where for the quadratic lattice with ferromagnetic interaction energies $-J_1$ and $-J_2$ in the horizontal and vertical directions, respectively, the modulus k_1 is given by

$$k_1 = [\sinh(2J_1/kT)\sinh(2J_2/kT)]^{-1}. \tag{1.2}$$

By suitable redefinition of k_1, the formula (1.1) can be shown to hold for the triangular and honeycomb lattices too.[10,11] The exact solutions for these two lattices were published independently by a number of authors (Houtappel,[12] Husimi and Syozi,[13] Wannier,[14] and Potts[15]).

In this paper[16] we follow the method of Kasteleyn[7] to represent the partition function of the triangular lattice as a Pfaffian, which is evaluated in the limit of an infinite lattice. By analogy with the MPW theory, we express the pair correlations along an axis of symmetry of the lattice as a Toeplitz determinant [Eqs. (6.10)–(6.13)], and thence derive the formula corresponding to (1.1) for the magnetization [Eq. (7.11)]. In Sec. 8 a numerical investigation of the pair correlations at the ferromagnetic critical point reveals that, as the separation of the spins increases, the correlations appear to decay with radial symmetry as $1/r^{\frac{1}{4}}$. Then the arguments given by Fisher[17] enable us to conclude that the parallel susceptibility, $\chi_\parallel$, should diverge as $(1 - T_c/T)^{-7/4}$. At the antiferromagnetic zero point the spins prefer, on the average, the state of minimum energy which contributes most to the zero-point entropy.[14] For a given sublattice the row correlations appear to decay

[1] For recent reviews see C. Domb, Advan. Phys. 9, Nos. 34, 35 (1960), and also M. E. Fisher, J. Math. Phys. 4, 278 (1963).
[2] E. Ising, Z. Phys. 31, 253 (1925).
[3] L. Onsager, Phys. Rev. 65, 117 (1944).
[4] B. Kaufman, Phys. Rev. 76, 1232 (1949).
[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), hereafter referred to as MPW.
[7] P. W. Kasteleyn, J. Math. Phys. 4, 287 (1963).
[8] L. Onsager, Nuovo Cimento Suppl. 6, 261 (1949).
[9] C. N. Yang, Phys. Rev. 85, 808 (1952). See also C. H. Chang, Phys. Rev. 88, 1422 (1952).
[10] R. B. Potts, Phys. Rev. 88, 352 (1952).
[11] H. S. Green, Z. Phys. 171, 129 (1963).
[12] R. M. F. Houtappel, Physica 16, 425 (1950).
[13] K. Husimi and I. Syozi, Progr. Theoret. Phys. (Kyoto) 5, 177, 341 (1950).
[14] G. H. Wannier, Phys. Rev. 79, 357 (1950).
[15] R. B. Potts, Proc. Phys. Soc. London, A68, 145 (1955). See also G. F. Newell, Phys. Rev. 79, 876 (1950) and H. N. V. Temperley, Proc. Roy. Soc. (London) A202, 202 (1950).
[16] In order to facilitate a comparison of the results, Secs. 2–7 of the present paper follow almost identically in development MPW, Ref. 6. The notation, however, is closer to that of Domb (Ref. 1) in that $z = \exp(-2K)$, and $v = \tanh K$ is essentially the *dual* transformation variable. w is reserved for the *inversion* transformation.
[17] M. E. Fisher, Physica 25, 521 (1959).

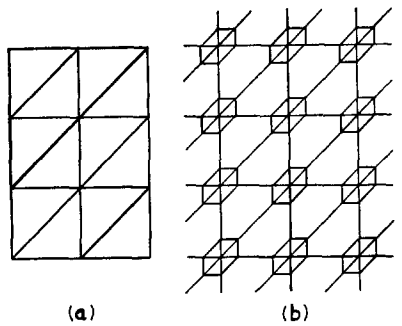


FIG. 1. The triangular lattice L and the decorated dimer lattice L'.

as $\epsilon_0 r^{-\frac{1}{4}} \times [1 + \epsilon_1/r + \epsilon_2/r^2 + \cdots]$. The remaining sections are concerned with the perpendicular susceptibility, $\chi_\perp$, which has been discussed by Fisher for the quadratic, honeycomb, and Bethe lattices.[18] For the triangular lattice $\chi_\perp$ can be calculated explicitly in terms of two, four, and six spin correlation functions (Sec. 10). For an antiferromagnet the behavior of $\chi_\perp$ is similar to that of $\chi_\parallel$ for which series expansions are available.[19,20] The triangular lattice does not admit a unique ordered antiferromagnetic ground state,[14] and the susceptibilities are infinite at zero temperature.

2. THE PARTITION FUNCTION

The triangular lattice can be considered as a rectangular lattice with a single second-neighbor interaction (Fig. 1a). For a lattice of N spins, each of which interacts with its six nearest neighbors (N is taken to be large compared with $N^{\frac{1}{2}}$ so that edge effects may be neglected), the partition function is given by

$$Z_N = \sum_{\sigma=\pm 1} \prod_{\text{nn}} \exp\,(K_1\sigma_{\alpha,\beta}\sigma_{\alpha+1,\beta} + K_2\sigma_{\alpha,\beta}\sigma_{\alpha,\beta+1} + K_3\sigma_{\alpha,\beta}\sigma_{\alpha+1,\beta+1}), \tag{2.1}$$

where $\sigma_{\alpha,\beta} = \pm 1$ indicates the state of the spin at the lattice site (α, β) and

$$K_i = J_i/kT, \qquad i = 1, 2, 3, \tag{2.2}$$

where $-J_1$, $-J_2$, $-J_3$ are the interaction energies between horizontal, vertical, and diagonal neighboring pairs of spins, respectively. For a ferromagnetic interaction the J_i are positive and a parallel spin state is energetically favored. The product is taken over all nearest-neighbor pairs and the sum over the 2^N states of the N spins. If we now define

$$z_i = \exp\,(-2K_i), \tag{2.3}$$

and

$$v_i = \tanh K_i = (1 - z_i)/(1 + z_i), \tag{2.4}$$

then the partition function may be written

$$Z_N = (\cosh K_1 \cosh K_2 \cosh K_3)^N \times \sum_{\sigma=\pm 1} \prod_{\text{nn}} (1 + v_1\sigma_{\alpha,\beta}\sigma_{\alpha+1,\beta}) \times (1 + v_2\sigma_{\alpha,\beta}\sigma_{\alpha,\beta+1})(1 + v_3\sigma_{\alpha,\beta}\sigma_{\alpha+1,\beta+1}). \tag{2.5}$$

In the expansion of the product, all terms containing an odd number of σ variables will vanish when the summation is performed. The remaining terms can be interpreted graphically as representing polygons with an even number of lines meeting at each vertex. Extracting a factor 2^N, which occurs after performing the product and summation, we can write

$$Z_N = (2 \cosh K_1 \cosh K_2 \cosh K_3)^N \times [1 + \sum_{r,s,t} v_1^r v_2^s v_3^t p(r, s, t)], \tag{2.6}$$

where $p(r, s, t)$ is the number of polygons defined above which can be constructed with r horizontal, s vertical, and t diagonal nearest-neighbor links. The calculation of the partition function is thus expressed as a "counting" problem. Kac and Ward[21] obtained a determinantal solution to this combinatorial problem for the quadratic lattice, and Potts[10] applied the same method to the triangular lattice. Recently Hurst and Green[22] have employed Pfaffians to perform the counting directly, whereas Kasteleyn[7] has shown that the enumeration of polygons on the quadratic lattice is equivalent to that of counting dimers on a decorated quadratic lattice (the "bathroom tile" lattice), which is also achieved by use of Pfaffians. Now we outline the application of Kasteleyn's method to the triangular lattice.

We have to construct a decorated triangular lattice L' such that the counting of all possible closed-packed dimer configurations on it is equivalent to the counting of the Ising polygons on the plane triangular lattice L. The decorated lattice L' must allow 0, 2, 4, or 6 lines to meet at each vertex, so each lattice site of L must be "exploded" and replaced by a hexagonal decoration (Figs. 1 and 2) in which each vertex is joined to the remaining five vertices. The vertices of each hexagon will be

[18] M. E. Fisher, J. Math. Phys. 4, 124 (1963), and Physica 26, 618 (1960).
[19] M. F. Sykes and I. J. Zucker, Phys. Rev. 124, 410 (1961).
[20] M. F. Sykes, J. Math. Phys. 2, 52 (1961); C. Domb and M. F. Sykes, J. Math. Phys. 2, 63 (1961) and Proc. Roy. Soc. (London) A240, 214 (1957).
[21] M. Kac and J. C. Ward, Phys. Rev. 88, 1332 (1952).
[22] C. A. Hurst and H. S. Green, J. Chem. Phys. 33, 1059 (1960). See also: A. M. Dykhne and Yu. B. Rumer, Usp. Fiz. Nauk 75, 101 (1961) [English transl.: Soviet Phys.—Usp. 75, 698 (1962)]. For the application of this method to more general planar lattices, see: C. A. Hurst, J. Chem. Phys. 38, 2558 (1963).

labeled clockwise U, S, R, D, T, L, and each hexagon will be identified by its Cartesian coordinates (p, q) on the lattice. Each possible dimer configuration on L' corresponds uniquely to a set of polygons on L, and each set of polygons on L corresponds to a dimer configuration on L', but the latter correspondence is not unique. The possible dimer configurations on L' corresponding to zero, two, four, and six lines meeting at a site are drawn (and their multiplicities given) in Fig. 3. Kasteleyn's topological theorem[23] states that if the (p_{ij})th element of the Pfaffian **P** is positive when the orientation (the arrow) is from site i to site j, then all transition diagrams with an *odd* orientation parity will be counted with the same sign in the expansion of the Pfaffian and those with an *even* orientation parity will be counted with the opposite sign. A close inspection of the orientation indicated in Figs. 1 and 2 will reveal that all the transition cycles corresponding to the diagrams in Fig. 3 having zero or two crossing points will be counted $+1$, and those with one or three crossing points will be counted -1. It transpires that each possible meet of bonds at each site is counted correctly: once and with a positive sign.

Now the elements of the Pfaffian can be written down immediately, but it is more convenient to proceed directly to the skew-symmetric matrix **A** whose determinant is equal to the square of the Pfaffian, so that

$$Z_N^2 = (2 \cosh K_1 \cosh K_2 \cosh K_3)^{2N} \, |\mathbf{A}|. \qquad (2.7)$$

The $6N \times 6N$ matrix **A** has a doubly cyclic structure

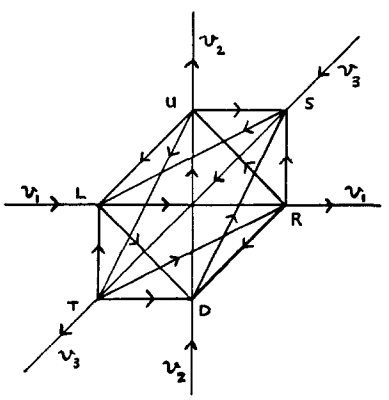


FIG. 2. The decorating hexagon to be placed at each vertex of L.

[23] A *transition cycle* is a closed circuit of bonds formed by superposing some standard configuration of bonds counted $+1$ on another configuration whose orientation parity is to be determined. The *orientation parity* of a transition cycle (which by definition contains an even number of bonds) is the parity of the number of arrows pointing in either direction around the cycle.

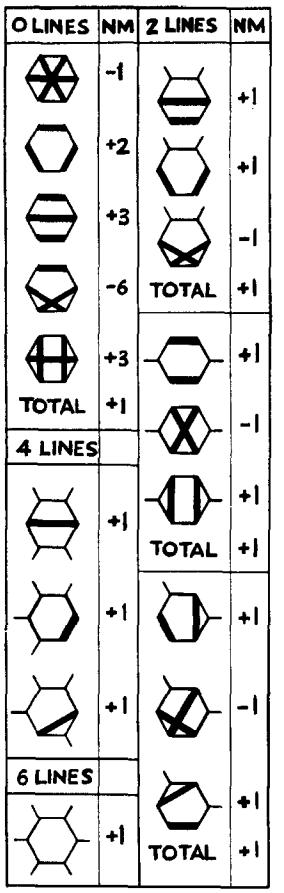


FIG. 3. Transition diagrams for 0, 2, 4, and 6 lines meeting at a vertex of L. (M is the multiplicity and N the parity.)

(in the limit of large N) and its nonvanishing elements are 6×6 matrices with the explicit values

$$A(\alpha, \beta; \alpha, \beta) = \begin{array}{c} \\ U \\ S \\ R \\ D \\ T \\ L \end{array} \begin{array}{c} \begin{array}{cccccc} U & S & R & D & T & L \end{array} \\ \begin{bmatrix} 0 & 1 & -1 & -1 & 1 & 1 \\ -1 & 0 & -1 & -1 & 1 & 1 \\ 1 & 1 & 0 & 1 & -1 & -1 \\ 1 & 1 & -1 & 0 & -1 & -1 \\ -1 & -1 & 1 & 1 & 0 & 1 \\ -1 & -1 & 1 & 1 & -1 & 0 \end{bmatrix} \end{array}, \qquad (2.8)$$

$$\begin{aligned} A(\alpha, \beta; \alpha, \beta + 1)_{UD} &= v_2, \\ A(\alpha, \beta; \alpha + 1, \beta + 1)_{ST} &= -v_3, \\ A(\alpha, \beta; \alpha + 1, \beta)_{RL} &= v_1, \\ A(\alpha, \beta; \alpha, \beta - 1)_{DU} &= -v_2, \\ A(\alpha, \beta; \alpha - 1, \beta - 1)_{TS} &= v_3, \\ A(\alpha, \beta; \alpha - 1, \beta)_{LR} &= -v_1. \end{aligned} \qquad (2.9)$$

The Pfaffian corresponding to a single decorating hexagon has the value $+1$ [from (2.8)].[24] The matrix $\mathbf{A}$ can be block diagonalized by a Fourier-type unitary transformation and the value of $|\mathbf{A}|$ can be obtained from the determinant of

$$\mathbf{A}(\phi_1, \phi_2) = A(0,0) + A(0,1)e^{i\phi_2} + A(1,1)e^{i(\phi_1+\phi_2)} + A(1,0)e^{i\phi_1} + A(0,-1)e^{-i\phi_2} + A(-1,-1)e^{-i(\phi_1+\phi_2)} + A(-1,0)e^{-i\phi_1}$$

$$= \begin{array}{c} \\ U \\ S \\ R \\ D \\ T \\ L \end{array} \begin{array}{c} \begin{array}{cccccc} U & S & R & D & T & L \end{array} \\ \begin{bmatrix} 0 & 1 & -1 & (-1+v_2e^{i\phi_2}) & 1 & 1 \\ -1 & 0 & -1 & -1 & (1-v_3e^{i(\phi_1+\phi_2)}) & 1 \\ 1 & 1 & 0 & 1 & -1 & (-1+v_1e^{i\phi_1}) \\ (1-v_2e^{-i\phi_2}) & 1 & -1 & 0 & -1 & -1 \\ -1 & (-1+v_3e^{-i(\phi_1+\phi_2)}) & 1 & 1 & 0 & 1 \\ -1 & -1 & (1-v_1e^{-i\phi_1}) & 1 & -1 & 0 \end{bmatrix} \end{array} . \quad (2.10)$$

For large N,

$$\ln |\mathbf{A}| \simeq \frac{N}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \ln |\mathbf{A}(\phi_1, \phi_2)| \, d\phi_1 \, d\phi_2, \quad (2.11)$$

where

$$\begin{aligned} |\mathbf{A}(\phi_1, \phi_2)| &= \Delta(\phi_1, \phi_2) \\ &= (1+v_1^2)(1+v_2^2)(1+v_3^2) + 8v_1v_2v_3 \\ &\quad - 2v_1(1-v_2^2)(1-v_3^2)\cos\phi_1 \\ &\quad - 2v_2(1-v_3^2)(1-v_1^2)\cos\phi_2 \\ &\quad - 2v_3(1-v_1^2)(1-v_2^2)\cos(\phi_1+\phi_2). \end{aligned} \quad (2.12)$$

From Eq. (2.7), we easily see that, since for large N $Z_N \simeq Z^N$,

$$\begin{aligned} \ln Z_N/N &\simeq \ln Z \\ &\simeq \ln 2 + \frac{1}{2(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} d\phi_1 \, d\phi_2 \\ &\quad \times \ln [\cosh 2K_1 \cosh 2K_2 \cosh 2K_3 \\ &\quad + \sinh 2K_1 \sinh 2K_2 \sinh 2K_3 - \sinh 2K_1 \cos\phi_1 \\ &\quad - \sinh 2K_2 \cos\phi_2 - \sinh 2K_3 \cos(\phi_1+\phi_2)] \end{aligned} \quad (2.13)$$

for the partition function of the triangular Ising lattice.[12] If in Eq. (2.13) any one of J_1, J_2, J_3 is put equal to zero, then Onsager's formula for the partition function of the quadratic lattice is obtained.

3. CORRELATION FUNCTIONS

The method of calculating two spin correlation functions has been presented in detail by MPW so we indicate here the extension of the MPW theory for the calculation of four and six spin correlations. Since boundary effects can be neglected for an infinite lattice the correlations depend only on the relative separations of the spins. As an example we discuss the 4-spin correlation (Fig. 4d)

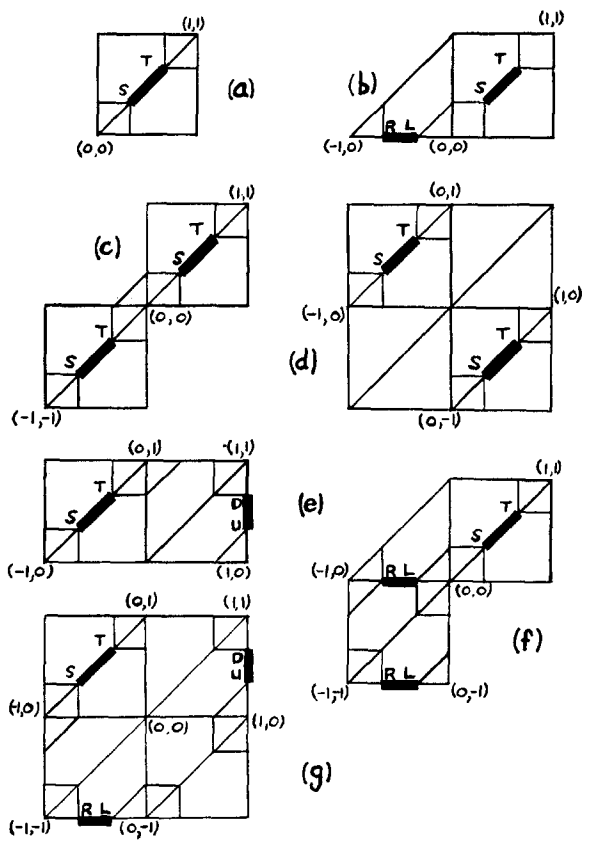


FIG. 4. Diagrams showing the perturbed dimers used in calculation of the correlations.

[24] MPW obtain a Pfaffian for a decorating square which has the value -1. This does not matter since only the *relative* signs of the transition cycles are of importance.

$$S_4 = \langle\sigma_{-1,0}\sigma_{0,1}\sigma_{0,-1}\sigma_{1,0}\rangle$$
$$= Z_N^{-1}(\cosh K_1 \cosh K_2 \cosh K_3)^N$$
$$\times \sum_{\sigma=\pm 1} \sigma_{-1,0}\sigma_{0,1}\sigma_{0,-1}\sigma_{1,0} \prod_{nn} (1 + v_1\sigma_{\alpha,\beta}\sigma_{\alpha+1,\beta})$$
$$\times (1 + v_2\sigma_{\alpha,\beta}\sigma_{\alpha,\beta+1})(1 + v_3\sigma_{\alpha,\beta}\sigma_{\alpha+1,\beta+1}). \quad (3.1)$$

Using the identities

$$\sigma_{-1,0}\sigma_{0,1}(1 + v_3\sigma_{-1,0}\sigma_{0,1}) = v_3(1 + v_3^{-1}\sigma_{-1,0}\sigma_{0,1}),$$
$$\sigma_{0,-1}\sigma_{1,0}(1 + v_3\sigma_{0,-1}\sigma_{1,0}) = v_3(1 + v_3^{-1}\sigma_{0,-1}\sigma_{1,0}), \quad (3.2)$$

which follow immediately since $\sigma_i^2 = 1$, Eq. (3.1) can be rewritten

$$Z_N v_3^{-2} S_4 = (\cosh K_1 \cosh K_2 \cosh K_3)^N$$
$$\times \sum_{\sigma=\pm 1} (1 + v_3^{-1}\sigma_{-1,0}\sigma_{0,1})(1 + v_3^{-1}\sigma_{0,-1}\sigma_{1,0})$$
$$\times \prod_{nn}{}' (1 + v_1\sigma_{\alpha,\beta}\sigma_{\alpha+1,\beta})(1 + v_2\sigma_{\alpha,\beta}\sigma_{\alpha,\beta+1})$$
$$\times (1 + v_3\sigma_{\alpha,\beta}\sigma_{\alpha+1,\beta+1}), \quad (3.3)$$

where $\prod'_{nn}$ is the product over all nearest-neighbor pairs of spins except $(\sigma_{-1,0}\sigma_{0,1})$ and $(\sigma_{0,-1}\sigma_{1,0})$. The RHS of equation (3.3) represents a perturbed partition function in which the dimers on the decorated lattice L' between sites $(-1, 0)$ and $(0, 1)$, and sites $(0, -1)$ and $(1, 0)$ are each weighted by a factor v_3^{-1} instead of v_3. This alteration of dimer weighting factors can be effected by constructing a skew-symmetric matrix $\mathbf{A} + \boldsymbol{\delta}$, where $\boldsymbol{\delta}$ will contain four nonzero elements. Squaring Eq. (3.3) and using (2.7) we obtain

$$|\mathbf{A}|\, v_3^{-4}(S_4)^2 = |\mathbf{A} + \boldsymbol{\delta}|, \quad (3.4)$$

or

$$S_4 = v_3^2\, |\mathbf{I} + \mathbf{A}^{-1}\boldsymbol{\delta}|^{\frac{1}{2}}. \quad (3.5)$$

If $\mathbf{y}$ is the 4×4 submatrix of $\boldsymbol{\delta}$ containing only those rows and columns in which nonzero elements appear, and $\mathbf{Q}$ is the 4×4 submatrix of $\mathbf{A}^{-1}$ formed from the same rows and columns, then

$$|\mathbf{I} + \mathbf{A}^{-1}\boldsymbol{\delta}|_{6N\times 6N} = |\mathbf{I} + \mathbf{Q}\mathbf{y}|_{4\times 4} \quad (3.6)$$
$$= |\mathbf{y}^{-1} + \mathbf{Q}|\, |\mathbf{y}|_{4\times 4}. \quad (3.7)$$

Now since $\mathbf{A}$, $\mathbf{A}^{-1}$, and $\boldsymbol{\delta}$ are skew symmetric, $\mathbf{y}$, $\mathbf{y}^{-1}$, $\mathbf{Q}$, and $\mathbf{y}^{-1} + \mathbf{Q}$ will be skew symmetric too, and we can express the square root in (3.5) in terms of Pfaffians. Hence

$$S_4 = \pm v_3^2 P(\mathbf{y}^{-1} + \mathbf{Q})P(\mathbf{y}), \quad (3.8)$$

where the sign is chosen to make the correlation positive at the ferromagnetic zero of temperature, $v_i = +1$. The extension of the MPW theory to other four-spin and higher (even)-order correlations is obvious.

4. THE INVERSE MATRIX

Because of the doubly cyclic structure of $\mathbf{A}$ its inverse can be obtained as

$$\mathbf{A}^{-1}(p, q; p', q') = \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \exp[(p - p')\phi_1 + (q - q')\phi_2]\mathbf{A}^{-1}(\phi_1, \phi_2)\, d\phi_1\, d\phi_2, \quad (4.1)$$

where $\mathbf{A}^{-1}(\phi_1, \phi_2)$ is the inverse matrix of $\mathbf{A}(\phi_1, \phi_2)$. Both $\mathbf{A}(\phi_1, \phi_2)$ and $\mathbf{A}^{-1}(\phi_1, \phi_2)$ are anti-Hermitian. We shall eventually require all the elements of $\mathbf{A}^{-1}(\phi_1, \phi_2)$ and since there are 36 of these, the calculation of $\mathbf{A}^{-1}(\phi_1, \phi_2)$ must be approached circumspectly. Let $\Delta = \det|\mathbf{A}(\phi_1, \phi_2)|$, and denote the cofactors of the elements in the top row of Δ by C_{UU}, C_{US}, etc. Apart from C_{UU} these cofactors have already been required for the evaluation of Δ in (2.12) and the explicit values are given in the Appendix. The remaining cofactors are most easily obtained by recognizing the transformation properties of Δ under successive rotations of the axes of the triangular lattice through 60°. Alternate permutation of the top row of Δ to the bottom and the left-hand column to the extreme right yields a succession of determinants whose relationship to the original determinant Δ provides the required group of transformations, which we list in the Appendix. $\mathbf{A}^{-1}(p, q; p', q')$ is a function of the differences $p - p'$ and $q - q'$, so we use the abbreviated notation

$$\mathbf{A}^{-1}(p, q; p', q')_{ij} = [p' - p, q' - q]_{ij}. \quad (4.2)$$

It follows that since $\mathbf{A}^{-1}(\phi_1, \phi_2)$ is anti-Hermitian $\mathbf{A}^{-1}$ is antisymmetric; or in terms of matrix elements,

$$[p' - p, q' - q]_{ij} = -[p - p', q - q']_{ji}, \qquad i \neq j. \quad (4.3)$$

We next show that

$$[k, k]_{ss} = -[k, k]_{tt} = 0. \quad (4.4)$$

From the appropriate transformation of C_{UU} we find that

$$C_{SS} = -C_{TT} = 2i[v_1(1 - v_2^2)\sin\phi_1 - v_2(1 - v_1^2)\sin\phi_2], \quad (4.5)$$

so that

$$[k, k]_{ss} = \frac{2i}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} [v_1(1 - v_2^2)\sin\phi_1 - v_2(1 - v_1^2)\sin\phi_2] \times \exp[-ik(\phi_1 + \phi_2)] \times \Delta^{-1}(\phi_1, \phi_2)\, d\phi_1\, d\phi_2, \quad (4.6)$$

where Δ^{-1} is given by (2.12). In the second term of the integral in (4.6), transform the variables as follows:

$$\begin{aligned} \phi_1 + \phi_2 &\to \omega \\ \phi_1 &\to \omega + \theta, \\ \phi_2 &\to -\theta, \end{aligned} \tag{4.7}$$

so that it becomes

$$\frac{2i}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} e^{-ik\omega}[v_2(1-v_1^2)]\sin\theta \times \Delta^{-1}(\theta,\omega)\,d\theta\,d\omega = I(v_2,v_1), \quad \text{say,} \tag{4.8}$$

where $\Delta(\theta, \omega)$ is of the form $A + B\cos\theta + C\sin\theta$, and

$$\begin{aligned} A &= (1+v_1^2)(1+v_2^2)(1+v_3^2) + 8v_1v_2v_3 \\ &\quad - 2v_3(1-v_1^2)(1-v_2^2)\cos\omega, \\ B &= -2v_2(1-v_1^2)(1-v_3^2) \\ &\quad - 2v_1(1-v_2^2)(1-v_3^2)\cos\omega, \\ C &= 2v_1(1-v_2^2)(1-v_3^2)\sin\omega. \end{aligned} \tag{4.9}$$

We note that A^2 and $(B^2 + C^2)$ are unaltered on interchange of v_1 and v_2. The integration over θ can easily be performed. It is convenient to define

$$I_n = \frac{1}{(2\pi)}\int_{-\pi}^{\pi}\frac{e^{in\theta}\cdot d\theta}{[A + B\cos\theta + C\sin\theta]}. \tag{4.10}$$

Contour integration round the unit circle gives

$$I_n = \alpha^n(A^2 - B^2 - C^2)^{-\frac{1}{2}}, \tag{4.11}$$

where

$$\alpha = [(A^2 - B^2 - C^2)^{\frac{1}{2}} - A]/(B - iC). \tag{4.12}$$

Taking the imaginary part of I_1, we obtain on substitution in (4.8)

$$I(v_2,v_1) = \frac{2i}{\pi}\int_{-\pi}^{\pi} v_1v_2(1-v_1^2)(1-v_2^2) \times \sin\omega[1 - A(A^2 - B^2 - C^2)^{-\frac{1}{2}}]/(B^2 + C^2)\,d\omega,$$

which is symmetrical in v_1 and v_2. A similar result holds for the first term in (4.6), so that the two terms cancel and $[k, k]_{ss}$ vanishes. This result will be needed later.

5. THE CORRELATION $\langle\sigma_{0,0}\sigma_{1,1}\rangle$

We first calculate the nearest-neighbor correlation $S_{01} = \langle\sigma_{0,0}\sigma_{1,1}\rangle$ which is related to the energy U_N. Since the coordination number q of the triangular lattice is six,

$$U_N = -3NJ\langle\sigma_{0,0}\sigma_{1,1}\rangle. \tag{5.1}$$

By analogy with Sec. 4 of MPW,

$$\langle\sigma_{0,0}\sigma_{1,1}\rangle = \pm v_3P(\mathbf{y}^{-1} + \mathbf{Q})P(\mathbf{y}), \tag{5.2}$$

where $\mathbf{y}$ and $\mathbf{Q}$ are given by (see Fig. 4a)

$$\mathbf{y} = \begin{matrix} \\ (0,0)S \\ (1,1)T \end{matrix}\begin{matrix} \begin{matrix}(0,0)S & (1,1)T\end{matrix} \\ \begin{bmatrix} 0 & -v_3^{-1} + v_3 \\ v_3^{-1} - v_3 & 0 \end{bmatrix} \end{matrix}, \tag{5.3}$$

$$\mathbf{Q} = \begin{matrix} \\ (0,0)S \\ (1,1)T \end{matrix}\begin{matrix} \begin{matrix}(0,0)S & (1,1)T\end{matrix} \\ \begin{bmatrix} [0,0]_{ss} & [1,1]_{st} \\ [-1,-1]_{ts} & [0,0]_{tt} \end{bmatrix} \end{matrix}. \tag{5.4}$$

From (4.4) we see that the diagonal elements of $\mathbf{Q}$ vanish and obtain

$$\mathbf{y}^{-1} + \mathbf{Q} = \begin{matrix} \\ (0,0)S \\ (1,1)T \end{matrix}\begin{matrix} \begin{matrix}(0,0)S & (1,1)T\end{matrix} \\ \begin{bmatrix} 0 & (v_3^{-1} - v_3)^{-1} + [1,1]_{st} \\ -(v_3^{-1} - v_3)^{-1} + [-1,-1]_{ts} & 0 \end{bmatrix} \end{matrix}, \tag{5.5}$$

so that on taking the Pfaffians of (5.3) and (5.5),

$$P(\mathbf{y}) = -(v_3^{-1} - v_3), \tag{5.6}$$

$$P(\mathbf{y}^{-1} + \mathbf{Q}) = (v_3^{-1} - v_3)^{-1} + [1,1]_{st}. \tag{5.7}$$

Substitution in (5.2) yields

$$\langle\sigma_{0,0}\sigma_{1,1}\rangle = \pm\{v_3 + (1 - v_3^2)[1,1]_{st}\}. \tag{5.8}$$

The matrix element in (5.8) is given by

$$[1,1]_{st} = \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} C_{TS} \times \exp[-i(\phi_1 + \phi_2)]\Delta^{-1}(\phi_1,\phi_2)\,d\phi_1\,d\phi_2, \tag{5.9}$$

where C_{TS} can be derived from C_{UD} by the appropriate transformation in the Appendix. We find that, choosing the positive sign, since the correlation is positive when $v_i = +1$,

$$\begin{aligned} \langle\sigma_{0,0}\sigma_{1,1}\rangle &= \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} v_3 - (1-v_3^2) \\ &\quad \times \{[(1-v_1^2)(1-v_2^2)]e^{-i(\phi_1+\phi_2)} - 4v_1v_2 \\ &\quad - v_3[(1+v_1^2)(1+v_2^2) + 2v_1(1-v_2^2)\cos\phi_1 \\ &\quad + 2v_2(1-v_1^2)\cos\phi_2]\}\Delta^{-1}(\phi_1,\phi_2)\,d\phi_1\,d\phi_2 \\ &= \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}[2v_3(1+v_1^2)(1+v_2^2) \end{aligned} \tag{5.10}$$

$$+ 4v_1v_2(1 + v_3^2) - v_3(1 - v_1^2)(1 - v_2^2)e^{i(\phi_1+\phi_2)}$$
$$- (1 - v_1^2)(1 - v_2^2)e^{-i(\phi_1+\phi_2)}]\Delta^{-1}(\phi_1, \phi_2)\, d\phi_1\, d\phi_2. \tag{5.11}$$

Transforming the variables as in (4.7) we get

$$\frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}(a - be^{i\omega} - ce^{-i\omega}) \times \Delta^{-1}(\theta, \omega)\, d\theta\, d\omega, \tag{5.12}$$

where

$$a = 2v_3(1 + v_1^2)(1 + v_2^2) + 4v_1v_2(1 + v_3^2), \tag{5.13}$$
$$b = v_3^2c = v_3^2(1 - v_1^2)(1 - v_2^2).$$

The θ-integration may be performed as before if we now take I_0 from (4.10), so that (5.12) becomes

$$\langle\sigma_{0,0}\sigma_{1,1}\rangle = \frac{1}{(2\pi)}\int_{-\pi}^{\pi}\frac{(a - be^{i\omega} - ce^{-i\omega})}{(A^2 - B^2 - C^2)^{\frac{1}{2}}}\, d\omega, \tag{5.14}$$

where A, B, and C are as defined in (4.9). After some lengthy algebra we find that the denominator factorizes and (5.14) can be written in the form

$$\langle\sigma_{0,0}\sigma_{1,1}\rangle = \frac{1}{(2\pi)}\int_{-\pi}^{\pi}\left(\frac{a - be^{i\omega} - ce^{-i\omega}}{a - be^{-i\omega} - ce^{i\omega}}\right)^{\frac{1}{2}} d\omega. \tag{5.15}$$

When the three interaction parameters are equal, a substantial simplification of the previous formulas occurs. A factor $(1 + v)^2$ can be removed from Δ, so that in terms of primed symbols

$$\Delta'(\phi_1, \phi_2) = (1 - 2v + 6v^2 - 2v^3 + v^4) - 2v(1 - v)^2 \times [\cos\phi_1 + \cos\phi_2 + \cos(\phi_1 + \phi_2)]$$
$$= A' + B'\cos\phi_2 + C'\sin\phi_2, \tag{5.16}$$

where

$$A' = (1 - 2v + 6v^2 - 2v^3 + v^4) - 2v(1 - v)^2\cos\phi_1,$$
$$B' = -2v(1 - v)^2(1 + \cos\phi_1), \tag{5.17}$$
$$C' = 2v(1 - v)^2\sin\phi_1.$$

$\Delta'(\phi_1, \phi_2)$ is unaltered by interchange of ϕ_1 and ϕ_2 and by the transformation (4.7) with $\omega = \phi_1$ and $\theta = \phi_2$. After removal of a factor $(1 + v)^2$ the constants in (5.14) and (5.15) become

$$a' = 2v(1 + v^2), \quad b' = v^2c' = v^2(1 - v)^2. \tag{5.18}$$

S_{01} can be expressed in terms of a complete elliptic integral of the first kind, $K(k_1)$, with modulus k_1 the same as that introduced in Eq. (1.1) but defined for the triangular lattice. In terms of the parameters for the inversion transformation[25,26] (to be denoted by $^+$),

$$k_1 = \left(\frac{\sinh 2K^+}{\sinh 2K}\right) = \left(\frac{w^{-1} - w}{z^{-1} - z}\right), \tag{5.19}$$

where

$$w^2 = \left(\frac{1 - z^2}{1 + 3z^2}\right) = \exp(-4K^+). \tag{5.20}$$

The critical point for the lattice is given by

$$z = w = 1/\sqrt{3}, \quad v = 2 - \sqrt{3}, \tag{5.21}$$

corresponding to a critical temperature $kT_c/J = 3.640957$. From Eq. (5.14) we easily check that the critical energy is $U_N(T_c) = -2NJ$. There is no corresponding antiferromagnetic solution to (5.21), but in a certain sense one may consider the absolute zero of temperature to be the critical point since there is no antiferromagnetic ordering. Putting $v = -1 (T = -0)$ in (5.14) we find, correctly, that the energy of the degenerate ground state is

$$U_N(-0) = \frac{-3N\,|J|}{\pi} \times \int_0^{\pi}\frac{(1 + 2\cos\phi)}{|1 + 2\cos\phi|}\, d\phi = -N\,|J|. \tag{5.22}$$

There is one further interesting result we can obtain from (5.14). By putting $v_3 = 0$ we regain the quadratic lattice, but $\langle\sigma_{0,0}\sigma_{1,1}\rangle$ is now the second-neighbor correlation. (5.15) becomes

$$\langle\sigma_{0,0}\sigma_{1,1}\rangle = \frac{1}{(2\pi)}\int_{-\pi}^{\pi}\left(\frac{4v_1v_2 - (1 - v_1^2)(1 - v_2^2)e^{-i\omega}}{4v_1v_2 - (1 - v_1^2)(1 - v_2^2)e^{i\omega}}\right)^{\frac{1}{2}} d\omega,$$
$$= \frac{1}{(2\pi)}\int_{-\pi}^{\pi}\left(\frac{\sinh 2K_1\sinh 2K_2 - e^{-i\omega}}{\sinh 2K_1\sinh 2K_2 - e^{i\omega}}\right)^{\frac{1}{2}} d\omega, \tag{5.23}$$

where as in (1.2) we can put $k = \sinh 2K_1\sinh 2K_2 = 1/k_1$, with

$$k_1 < 1 \quad \text{for} \quad T < T_c, \quad \text{and} \quad k_1 \geq 1 \quad \text{for} \quad T \geq T_c. \tag{5.24}$$

Now setting

$$k' = 2/(k^{\frac{1}{2}} + k^{-\frac{1}{2}}) \tag{5.25}$$

—this definition parallels the change of modulus in the Landen transformation for complete elliptic

[25] G. H. Wannier, Rev. Mod. Phys. 17, 50 (1945).
[26] R. B. Potts, thesis Oxford University, England (1951) (unpublished).

integrals[27]—we can write for (5.23)

$$\langle\sigma_{0,0}\sigma_{1,1}\rangle = \frac{2}{\pi}\int_0^{\frac{1}{2}\pi} \frac{(k+\cos 2\beta)\,d\beta}{(1+k)(1-k'^2\sin^2\beta)^{\frac{1}{2}}},$$

so that in terms of complete elliptic integrals of the first and second kinds,[28]

$$\langle\sigma_{0,0}\sigma_{1,1}\rangle = \pi^{-1}[(k^{-1}+1)E(k') + (k-1)K(k')] \tag{5.26}$$

where the modulus k' is ≤ 1 at all temperatures. The "equals" applies at the critical point for the quadratic lattice given by

$$k = \sinh 2K_1 \sinh 2K_2 = 1. \tag{5.27}$$

6. THE CORRELATION $\langle\sigma_{0,0}\sigma_{k,k}\rangle$

The previous analysis can easily be generalized for an arbitrary separation of the two spins in the diagonal direction. We quote the MPW analysis for diagonal correlations. $\mathbf{y}$ is now the $2k \times 2k$ skew-symmetric matrix

$$\mathbf{y} = (v_3^{-1} - v_3)\begin{bmatrix} \mathbf{O} & -\mathbf{I} \\ \mathbf{I} & \mathbf{O} \end{bmatrix}\begin{matrix} S \\ T \end{matrix}, \tag{6.1}$$

so that

$$\mathbf{y}^{-1} = (v_3^{-1} - v_3)^{-1}\begin{bmatrix} \mathbf{O} & \mathbf{I} \\ -\mathbf{I} & \mathbf{O} \end{bmatrix}. \tag{6.2}$$

The $2k \times 2k$ submatrix $\mathbf{Q}$ obtained from $\mathbf{A}^{-1}$ by omitting all rows and columns except those contained in $\mathbf{y}$ is

$$\mathbf{Q} = \begin{bmatrix} \mathbf{O} & \mathbf{S} \\ -\mathbf{S} & \mathbf{O} \end{bmatrix} \tag{6.3}$$

where $\mathbf{S}$ is the $k \times k$ matrix

$$\mathbf{S} = \begin{bmatrix} [1,1]_{st} & [2,2]_{st} & [3,3]_{st} & \cdots & [k,k]_{st} \\ [0,0]_{st} & [1,1]_{st} & [2,2]_{st} & \cdots & [k-1,k-1]_{st} \\ [-1,-1]_{st} & [0,0]_{st} & [1,1]_{st} & \cdots & [k-2,k-2]_{st} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ [2-k,2-k]_{st} & [3-k,3-k]_{st} & [4-k,4-k]_{st} & \cdots & [1,1]_{ts} \end{bmatrix}, \tag{6.4}$$

and we have used the fact that

$$[k,k]_{st} = -[-k,-k]_{ts},$$

and

$$[k,k]_{ss} = -[k,k]_{tt} = 0.$$

To obtain the correlation

$$\langle\sigma_{0,0}\sigma_{k,k}\rangle = \pm v_3^k P(\mathbf{y}^{-1}+\mathbf{Q})P(\mathbf{y}) \tag{6.5}$$

we require

$$P(\mathbf{y}) = \pm(v_3^{-1} - v_3)^k, \tag{6.6}$$

and

$$P(\mathbf{y}^{-1}+\mathbf{Q}) = \pm|(v_3^{-1}-v_3)^{-1}\mathbf{I} + \mathbf{S}|. \tag{6.7}$$

Substitution into (6.5) gives the result

$$\langle\sigma_{0,0}\sigma_{k,k}\rangle = |v_3\mathbf{I} + (1-v_3^2)\mathbf{S}|, \tag{6.8}$$

where the sign has been chosen so that the correlation is positive at the ferromagnetic zero. By defining

$$a_n = v_3\delta_{0n} + (1+v_3^2)[1+n, 1+n]_{st} \tag{6.9}$$

the correlation can be written as the Toeplitz determinant

$$\langle\sigma_{0,0}\sigma_{k,k}\rangle = \begin{vmatrix} a_0 & a_1 & a_2 & \cdots & a_{k-1} \\ a_{-1} & a_0 & a_1 & \cdots & a_{k-2} \\ a_{-2} & a_{-1} & a_0 & \cdots & a_{k-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{1-k} & a_{2-k} & a_{3-k} & \cdots & a_0 \end{vmatrix}. \tag{6.10}$$

a_0 has already been discussed in the previous section, so starting with the double integral representation of the matrix element in (6.9),

$$[1+n, 1+n]_{st} = \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} C_{TS} \times \exp[-i(1+n)(\phi_1+\phi_2)]\Delta^{-1}(\phi_1,\phi_2)\,d\phi_1\,d\phi_2, \tag{6.11}$$

we obtain, by an analysis similar to that leading to Eq. (5.15),

$$a_n = \frac{1}{(2\pi)}\int_{-\pi}^{\pi} e^{-in\omega}\left(\frac{a - be^{i\omega} - ce^{-i\omega}}{a - be^{-i\omega} - ce^{i\omega}}\right)^{\frac{1}{2}} d\omega, \tag{6.12}$$

[27] Bateman Manuscript Project, *Higher Transcendental Functions*, edited by A. Erdelyi (McGraw-Hill Book Company, Inc., New York, 1953), Vol. II, p. 318, Eq. (12).

[28] B. de Haan, *Nouvelles Tables d'Intégrales Définies*, (Hafner Publishing Company, New York, 1957) Table 57, Integrals 1 and 4.

where a, b, and c are as defined in (5.13). The a_n are thus seen to be the coefficients in the Fourier expansion of a function $A(\omega)$;

$$A(\omega) = \left(\frac{a - be^{i\omega} - ce^{-i\omega}}{a - be^{-i\omega} - ce^{i\omega}}\right)^{\frac{1}{2}}. \tag{6.13}$$

The reduction to the quadratic lattice can be done in two ways. If as in the discussion of $\langle\sigma_{0,0}\sigma_{1,1}\rangle$ we set $v_3=0$, we obtain the same expression for the diagonal correlations as that derived by Onsager.[29] The generating function for the Toeplitz determinant is now

$$B(\omega) = \left(\frac{\sinh 2K_1 \sinh 2K_2 - e^{-i\omega}}{\sinh 2K_1 \sinh 2K_2 - e^{i\omega}}\right)^{\frac{1}{2}} = \sum_{n=-\infty}^{\infty} b_n e^{in\omega}, \tag{6.14}$$

and the elements b_n can be expressed in terms of Legendre functions of the second kind[30],

$$b_n = \frac{1}{(2\pi)} \int_{-\pi}^{\pi} \frac{[k\cos n\omega - \cos(n+1)\omega]}{[1 - 2k\cos\omega + k^2]^{\frac{1}{2}}}\, d\omega \tag{6.15}$$

$$= \pi^{-1}[k^{\frac{1}{2}}Q_{n-\frac{1}{2}}(x) - k^{-\frac{1}{2}}Q_{n+\frac{1}{2}}(x)], \tag{6.16}$$

where

$$x = \tfrac{1}{2}(k + k^{-1}) = \cosh\eta,$$

η being defined as in MPW by $e^{-\eta} = k^{\pm 1}$ according as $T \gtrless T_c$. At the critical point $k = 1$ and $b_n = 2/[\pi(2n+1)]$.[31] Alternatively, setting $v_1 = 0$ and then replacing v_3 by v_1 we get the generating function for row correlations obtained by MPW:

$$C(\omega) = \left(\frac{(v_1 z_2 e^{i\omega} - 1)(v_1 e^{i\omega} - z_2)}{(e^{i\omega} - v_1 z_2)(z_2 e^{i\omega} - v_1)}\right)^{\frac{1}{2}} = \sum_{n=-\infty}^{\infty} c_n e^{in\omega}, \tag{6.17}$$

where

$$z_2 = (1 - v_2)/(1 + v_2) = \exp(-2K_2). \tag{6.18}$$

7. THE SPONTANEOUS MAGNETIZATION

The spontaneous magnetization is related to the long-range order by the formula

$$M^2 = R = \lim_{k\to\infty} \langle\sigma_{0,0}\sigma_{k,k}\rangle. \tag{7.1}$$

The limiting value of the correlation can be obtained by applying the following theorem on Toeplitz determinants[32]:

If $D_k(A) = \det|a_n|$ is the Toeplitz determinant of order k generated by the function

$$A(\omega) = \sum_{n=-\infty}^{\infty} a_n e^{in\omega}, \tag{7.2}$$

then

$$G(A) = \lim_{k\to\infty} |D_k|^{1/k} = \left|\exp\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln A(\omega)\, d\omega\right]\right|, \tag{7.3}$$

and further,

$$\lim_{k\to\infty} |D_k|/G^k = \exp\left[\sum_{n=1}^{\infty} n d_n d_{-n}\right], \tag{7.4}$$

where

$$d_n = \frac{1}{(2\pi)} \int_{-\pi}^{\pi} \ln A(\omega) e^{-in\omega}\, d\omega \tag{7.5}$$

are the coefficients in the Laurent expansion of $\ln A(\omega)$.

Now the generating function $A(\omega)$ defined by (6.13) for the triangular lattice can be written in the form

$$A(\omega) = e^{i\delta}, \tag{7.6}$$

where δ is a real angle. Hence

$$\ln A(\omega) = i\delta = \tfrac{1}{2}\ln\left(\frac{a - be^{i\omega} - ce^{-i\omega}}{a - be^{-i\omega} - ce^{i\omega}}\right), \tag{7.7}$$

or equivalently, factorizing the argument of the logarithm,

$$\ln A(\omega) = \tfrac{1}{2}\ln\left[\frac{(1 - c_1 e^{i\omega})(1 - c_2 e^{-i\omega})}{(1 - c_1 e^{-i\omega})(1 - c_2 e^{i\omega})}\right], \tag{7.8}$$

so that $d_0 = 0$ and $G(A) = 1$. Expanding (7.8) we obtain the Laurent coefficients

$$d_n = -(2n)^{-1}(c_1^n - c_2^n) = -d_{-n}. \tag{7.9}$$

Below the critical point,

$$\sum_{n=1}^{\infty} n d_n d_{-n} = \tfrac{1}{4}\ln\left(\frac{(1 - c_1^2)(1 - c_2^2)}{(1 - c_1 c_2)^2}\right). \tag{7.10}$$

Substituting in (7.4), the limiting value of the Toeplitz determinant for $\langle\sigma_{0,0}\sigma_{k,k}\rangle$, and hence the long-range order, is given by

$$R = \lim_{k\to\infty} \langle\sigma_{0,0}\sigma_{k,k}\rangle = \left[\frac{(1 - c_1^2)(1 - c_2^2)}{(1 - c_1 c_2)^2}\right]^{\frac{1}{2}} = (1 - k_1^2)^{\frac{1}{2}}, \tag{7.11}$$

[29] See C. Domb, Ref. 1, p. 201, Eq. (108), where the index $\frac{1}{2}$ has been omitted from the generating function.
[30] Reference 27, Vol. I, p. 156, Eq. (10).
[31] See footnote on p. 1251 of Ref. 5.
[32] U. Grenander and G. Szegö, *Toeplitz Forms and their Applications*, (University of California Press, and Cambridge University Press, Cambridge, England, 1958).

TABLE I. Row and diagonal critical-point correlations for the quadratic (Q) and triangular (T) lattices. $\omega_{ij} = \langle \sigma_{0,0}\, \sigma_{i,j} \rangle$.

k	ω_{kk}^{T}	$\omega_{kk}^{T} \times k^{\frac{1}{4}}$	ω_{kk}^{Q}	$\omega_{kk}^{Q} \times r^{\frac{1}{4}}$	ω_{k0}^{Q}	$\omega_{k0}^{Q} \times k^{\frac{1}{4}}$
1	0.66667	0.66667	0.63662	0.69424	0.70711	0.70711
2	0.56192	0.66824	0.54038	0.70079	0.59472	0.70724
3	0.50783	0.66834	0.48927	0.70219	0.53579	0.70514
4	0.47266	0.66844	0.45565	0.70270	0.49800	0.70426
5	0.44706	0.66851	0.43107	0.70294	0.47079	0.70399
6	0.42717	0.66856	0.41194	0.70308	0.44972	0.70385
7	0.39755	0.66860	0.39641	0.70316	0.43267	0.70377
8	0.38602	0.66861	0.38343	0.70321	0.41843	0.70371
...	...	...	...	...	...	...
∞	0	0.66865	0	0.70338	0	0.7035

where

$$k_1^2 = \frac{[(1 - v_1^2)(1 - v_2^2)(1 - v_3^2)]^2}{16(1 + v_1v_2v_3)(v_1 + v_2v_3)(v_2 + v_3v_1)(v_3 + v_1v_2)}, \tag{7.12}$$

and

$$M = (1 - k_1^2)^{\frac{1}{8}}. \tag{7.13}$$

Above the critical point the series (7.10) diverges and the magnetization is zero.[10,11]

8. CRITICAL-POINT CORRELATIONS

A. Ferromagnetic

The row correlations of the quadratic lattice have been studied by Onsager and Kaufman.[5] At the critical point, $v = z = \sqrt{2} - 1$, the elements of the Toeplitz determinant for the correlation $\langle \sigma_{0,0}\sigma_{k,0} \rangle$ simplify and can be evaluated explicitly. The c_n defined by (6.17) are equivalent to the Σ_{-n} of Onsager and Kaufman, and at the critical point,

$$c_n = \frac{2}{\pi} \int_0^{\frac{1}{4}\pi} [\sqrt{2} \cos 2n\alpha \sin \alpha + \sin 2n\alpha \cos \alpha]/(2 - \cos^2 \alpha)^{\frac{1}{2}}\, d\alpha, \tag{8.1}$$

which for large values of n has the asymptotic expansion

$$c_n \simeq \frac{1}{(\pi n)} \left[1 - \frac{\sqrt{2}}{2n} + \frac{1}{2n^2} - \frac{\sqrt{2}}{2n^3} + \frac{5}{4n^4} - \cdots \right]. \tag{8.2}$$

The exact critical values of c_n for $|n| \leq 3$ have been given by Onsager and Kaufman, who also give an approximate formula for the numerical value of the Toeplitz determinant for $\langle \sigma_{0,0}\sigma_{k,0} \rangle$ valid for large spin separation k:

$$\langle \sigma_{0,0}\sigma_{k,0} \rangle \simeq \prod_{s=1}^{k} \frac{\Gamma(s)\Gamma(s)}{\Gamma(s + \frac{1}{2})\Gamma(s - \frac{1}{2})}. \tag{8.3}$$

The expression (8.3) is exact for the diagonal correlations, for which

$$b_n = [\pi(n + \tfrac{1}{2})]^{-1} \tag{8.4}$$

at the critical point.[31] To first order in $1/n$, the elements of the Toeplitz determinants for row and diagonal correlations are equal, and one might reasonably assume that their mode of decay with distance is similar. It can be shown that for the diagonal correlations, for which the kth-order Toeplitz determinant is given by (8.3),

$$\langle \sigma_{0,0}\sigma_{k,k} \rangle \simeq B_0/r^{\frac{1}{4}}, \tag{8.5}$$

where r is the separation of the spins and $B_0 \simeq 0.7034$.[1,17] To test this result further for the row correlations, we have calculated $\langle \sigma_{0,0}\sigma_{k,0} \rangle$—exactly for $k \leq 4$ and to the approximation for c_n given by (8.2) for $|n| > 3$—for a few values of k. A plot of $[r^{\frac{1}{4}} \times \langle \sigma_{0,0}\sigma_{k,0} \rangle]$ against $1/k^2$ is a straight line—within the limits of the approximation—and the intercept at $1/k^2 = 0$ yields an estimate for C_0. Hence we find that for large spin separation r ($= k$ in this case),

$$\langle \sigma_{0,0}\sigma_{k,0} \rangle \simeq (C_0/r^{\frac{1}{4}})[1 + C_1/k^2], \tag{8.6}$$

where

$$C_0 \simeq 0.7035 \pm 2, \quad \text{and} \quad C_1 \simeq +0.02.$$

As a check on the reliability of the method the diagonal correlations were evaluated exactly for selected values of k, and in agreement with the exact result (8.5)[17]:

$$\langle \sigma_{0,0}\sigma_{k,k} \rangle \simeq (B_0/r^{\frac{1}{4}})[1 + B_1/k^2], \tag{8.7}$$

where $B_0 = 0.70338 \pm 2$, and $B_1 = -0.0156 \pm 5$. Thus it appears that the correlations decay with approximately radial symmetry for large spin separations. This is consistent with the formula conjectured by Onsager for the mean reciprocal range of order β.[3] Specimen numerical values are quoted in Table I.

The investigation of the triangular lattice critical-point correlations follows the same lines as above. At the ferromagnetic critical point, $v = w = 1/\sqrt{3}$ (see footnote 16), the elements a_n of the Toeplitz determinant again reduce to elementary integrals. Explicitly,

$$a_n = \frac{2}{\pi} \int_0^{\pi/2} [2 \cos 2n\alpha \sin \alpha + \sqrt{3} \sin 2n\alpha \cos \alpha]/(4 - \cos^2 \alpha)^{\frac{1}{2}}\, d\alpha, \tag{8.8}$$

which has the asymptotic expansion

TABLE II. Antiferromagnetic correlations at $T = -0$ for the triangular lattice. $\omega_{k0} = \langle \sigma_{0,0}\, \sigma_{k,0} \rangle$.

l	$\omega_{3l-2,0}$	$\omega_{3l-2,0} \times k^{\frac{1}{4}}$	$\omega_{3l-1,0}$	$\omega_{3l-1,0} \times k^{\frac{1}{4}}$	$\omega_{3l,0}$	$\omega_{3l,0} \times k^{\frac{1}{4}}$
1	−0.33333	−0.33333	−0.19285	−0.27273	+0.35852	+0.62097
2	−0.16561	−0.33123	−0.13362	−0.29878	+0.25682	+0.62909
3	−0.12320	−0.32597	−0.10802	−0.30552	+0.21027	+0.63080
4	−0.10225	−0.32334	−0.09302	−0.30852	+0.18227	+0.63141
5	−0.08925	−0.32180	−0.08291	−0.31020	+0.16311	+0.63170
6	−0.08020	−0.32080	−0.07550	−0.31128	+0.14893	+0.63186
8	−0.06813	−0.31957	−0.06518	−0.31257	+0.12901	+0.63202
10	−0.06026	−0.31886	−0.05818	−0.31332	+0.11540	+0.63209
14	−0.05029	−0.31805	−0.04906	−0.31415	+0.09754	+0.63216
17	−0.04539	−0.31770	−0.04448	−0.31451	+0.08852	+0.63218
20	−0.04168	−0.31746	−0.04098	−0.31476	+0.08162	+0.63219
...	...	...	...	...	...	...
∞	0	−0.31613	0	−0.31613	0	+0.63223

$$a_n \simeq \frac{1}{(\pi n)}\left[1 - \frac{\sqrt{3}}{\sigma n} + \frac{1}{3n^2} - \frac{\sqrt{3}}{6n^3} + \frac{1}{3n^4} - \cdots\right]. \tag{8.9}$$

The leading term is once again $(\pi n)^{-1}$, so that for large n, $a_n \simeq b_n \simeq c_n$, and we might expect that the triangular lattice correlations will decay as $1/r^{\frac{1}{2}}$ too. The exact values of a_n for $|n| \leq 3$ are

$$\begin{aligned} a_0 &= 2/3, \\ a_{\pm 1} &= (2 - 4\sqrt{3}/\pi) \\ &\pm (4\sqrt{3}/3 - 6/\pi) = \begin{cases} +0.194\,226 \\ -0.604\,857 \end{cases}, \\ a_{\pm 2} &= (22 - 40\sqrt{3}/\pi) \\ &\pm (40\sqrt{3}/3 - 72/\pi) = \begin{cases} +0.122\,543 \\ -0.228\,855 \end{cases}, \\ a_{\pm 3} &= (258 - 468\sqrt{3}/\pi) \\ &\pm (460\sqrt{3}/3 - 834/\pi) = \begin{cases} +0.088\,756 \\ -0.132\,599 \end{cases}. \end{aligned} \tag{8.10}$$

Clearly row and diagonal correlations are the same, and, as for the quadratic lattice, we have calculated $\langle \sigma_{0,0}\sigma_{k,k} \rangle = \langle \sigma_{0,0}\sigma_{k,0} \rangle$—exactly for $k \leq 4$ and to the approximation for a_n given by (8.9) for $|n| > 3$—for selected values of k. We conclude that for large spin separation r the correlations decay with approximately radial symmetry as

$$\langle \sigma_{0,0}\sigma_{k,k} \rangle \simeq (A_0/r^{\frac{1}{2}})[1 + A_1/k^2], \tag{8.11}$$

where $A_0 = 0.66865 \pm 1$, and $A_1 \simeq -0.0037 \pm 1$. Sample numerical results have been included in Table I.

Now by an argument similar to that given by Fisher for the quadratic lattice,[17] one can deduce that the parallel susceptibility should diverge as $(1 - T_c/T)^{-7/4}$.

B. Antiferromagnetic

At the antiferromagnetic zero point, the correlation $\langle \sigma_{0,0}\sigma_{k,k} \rangle$ is given by the $k \times k$ Toeplitz determinant defined by (6.13) with $v = -1$, $T = -0$. The elements a_n can be evaluated explicitly as

$$a_0 = -1/3, \quad a_n = -[2 \sin (2\pi n/3)]/(\pi n). \tag{8.12}$$

That is,

$$\begin{aligned} a_0 &= -1/3, \qquad a_{3r} = 0, \\ a_{(3r-1)} &= -a_{(3r-2)} = \sqrt{3}/(\pi n). \end{aligned} \tag{8.13}$$

Table II contains the numerical values of $\langle \sigma_{0,0}\sigma_{k,k} \rangle$ for various values of spin separation k $(= r)$. The correlations fall in groups of three; two negative and one positive, the magnitude of the positive correlation occurring at $k = 3l$, $l = 1, 2, \cdots$ being almost exactly equal to the magnitude of the sum of the two adjacent negative ones. This periodicity is related to the three sublattices into which the triangular lattice may be divided, and there is a direct correspondence between the "frozen in" state apparent from the values of $\langle \sigma_{0,0}\sigma_{k,k} \rangle$ and the ground state of leading entropy discussed by Wannier (p. 361, and Fig. 8 of Ref. 14). An analysis of the decay of the correlations with distance k on the same lines as that performed for the ferromagnetic case indicates that for a given sublattice, $i = 0, 1, 2$, $\langle \sigma_{0,0}\sigma_{k,k} \rangle \simeq \epsilon_0^{(i)}/k^{\frac{1}{2}}$ for large k, and that in detail

$$\langle \sigma_{0,0}\sigma_{k,k} \rangle \simeq \frac{\epsilon_0^{(i)}}{r^{\frac{1}{2}}}\left[1 + \frac{\epsilon_1^{(i)}}{r} + \frac{\epsilon_2^{(i)}}{r^2} + \cdots\right], \qquad i = 0, 1, 2. \tag{8.14}$$

For the sublattice containing the origin spin $\sigma_{0,0}$, the correlations are all positive:

$$\epsilon_0^{(0)} = +0.632\,226 \pm 2,$$
$$\epsilon_1^{(0)} = 0, \tag{8.15}$$
$$\epsilon_2^{(0)} = -0.187 \pm 5.$$

For the sublattice $k_1 = (3l - 2)$,

$$\epsilon_0^{(1)} = -0.316\,13 \pm 3, \quad \epsilon_1^{(1)} = +0.32 \pm 2, \tag{8.16}$$

and for the second sublattice $k_2 = (3l - 1)$,

$$\epsilon_0^{(2)} = -0.316\,13 \pm 3, \quad \epsilon_1^{(2)} = -0.32 \pm 2. \tag{8.17}$$

Thus the leading terms are such that asymptotically

$$\epsilon_0^{(1)} \simeq \epsilon_0^{(2)} \simeq -\tfrac{1}{2}\epsilon_0^{(0)}, \tag{8.18}$$

and

$$\epsilon_1^{(1)} \simeq -\epsilon_1^{(2)}. \tag{8.19}$$

The sum of the group of correlations $\langle\sigma_{0,0}, \sigma_{3l-1,0}\rangle$, $\langle\sigma_{0,0}\sigma_{3l,0}\rangle$ and $\langle\sigma_{0,0}\sigma_{3l+1,0}\rangle$ decays so rapidly with distance that Eqs. (8.18) and (8.19) are almost certainly exact. Hence the series $\sum_{k=1}^{\infty}\langle\sigma_{0,0}\sigma_{k,0}\rangle$ converges rapidly too, and has the estimated sum $-0.333\,275 \pm 1$.

9. PERPENDICULAR SUSCEPTIBILITY (i)

Fisher[18] has derived an exact formula for the perpendicular susceptibility $\chi_\perp$ of an Ising lattice in terms of the correlation functions. For a lattice with coordination number q,

$$\chi_\perp(T) = \frac{Nm^2}{kT}\Big[C_0^q(K) + C_1^q(K)\sum_{(k,l)}\langle s_k s_l\rangle + C_2^q(K)\sum_{klmn}\langle s_k s_l s_m s_n\rangle + \cdots\Big], \tag{9.1}$$

where the sums extend over all possible even spin correlations between the q nearest-neighbor spins to s_0; and the coefficients C_r^q are explicitly

$$C_r^q(K) = 2^{-q}\sum_{s_k=\pm 1}\Big[\prod_{l=1}^{2r} s_l\Big] \times \tanh\Big(K\sum_{k=1}^{q} s_k\Big)\Big/\Big(K\sum_{k=1}^{q} s_k\Big). \tag{9.2}$$

If the lattice orders both ferro- and antiferromagnetically, $\chi_\perp$ is an even function of K, rising from the finite value Nm^2/qJ at $T = 0$ to a maximum, which occurs just above the critical point, and then falling to zero as $T \to \infty$. At $T = T_c$, $\chi_\perp$ exhibits a logarithmic singularity (with vertical tangent), but remains finite. Thus $\chi_\perp$ will be a smooth continuous function of temperature for the triangular lattice with ferromagnetic interactions: J positive.

Evaluating the coefficients C_r^6 from (9.2) we get

$$\begin{aligned} C_0^6 &= \tfrac{1}{32}[\tau(6K) + 6\tau(4K) + 15\tau(2K) + 10],\\ C_1^6 &= \tfrac{1}{32}[\tau(6K) + 2\tau(4K) - \tau(2K) - 2],\\ C_2^6 &= \tfrac{1}{32}[\tau(6K) - 2\tau(4K) - \tau(2K) + 2],\\ C_3^6 &= \tfrac{1}{32}[\tau(6K) - 6\tau(4K) + 15\tau(2K) - 10], \end{aligned} \tag{9.3}$$

where

$$\tau(x) = \tanh x/x. \tag{9.4}$$

Figure 4(g) shows the nearest-neighbor spins of $s_0 = \sigma_{0,0}$. In the case of equal interactions the correlations required simplify and the sums in (9.1) become

$$\sum_{(k,l)}\langle s_k s_l\rangle = 6S_{01} + 6S_{13} + 3S_{14} = 6\langle\sigma_{0,0}\sigma_{1,1}\rangle + 6\langle\sigma_{-1,0}\sigma_{1,1}\rangle + 3\langle\sigma_{-1,-1}\sigma_{1,1}\rangle, \tag{9.5}$$

$$\sum_{(klmn)}\langle s_k s_l s_m s_n\rangle = 6S_{1234} + 6S_{1235} + 3S_{1245} = 6\langle\sigma_{1,0}\sigma_{1,1}\sigma_{0,1}\sigma_{-1,0}\rangle + 6\langle\sigma_{1,0}\sigma_{1,1}\sigma_{0,1}\sigma_{-1,-1}\rangle + 3\langle\sigma_{-1,0}\sigma_{0,1}\sigma_{0,-1}\sigma_{1,0}\rangle, \tag{9.6}$$

and the six-spin correlation

$$S_6 = S_{123456} = \langle\sigma_{-1,0}\sigma_{0,1}\sigma_{1,1}\sigma_{1,0}\sigma_{0,-1}\sigma_{-1,-1}\rangle. \tag{9.7}$$

10. CALCULATION OF CORRELATIONS

The 2-, 4-, and 6-spin correlations required for the evaluation of $\chi_\perp$ can be evaluated by the methods discussed in MPW and Secs. 3, 5, and 6 of this paper. What follows is, therefore, only an outline of the results. For each correlation, the figure (Fig. 4) shows the altered bonds in the decorated (dimer) lattice and the Pfaffian contains the corresponding matrix elements. The sign of the Pfaffian has been chosen in each case to make the correlation positive at the ferromagnetic zero; some of the correlations are negative for antiferromagnetic interactions. In general the matrix elements can be expressed in terms of elliptic integrals. At the critical points, $T = T_c$ ferromagnetic, and $T = -0$ antiferromagnetic, the integrals simplify and can be performed exactly. The values of the correlations at these points are quoted. To make the results more compact, the factor $(1 - v^2)$ has been absorbed into the matrix elements which are now denoted by rounded brackets.

The correlation S_{01} has already been discussed in Sec. 5 (see Fig. 4a), and is given by

$$S_{01} = v + (1, 1)_{st}, \tag{10.1}$$
$$S_{01}(T_c) = \tfrac{2}{3}, \tag{10.2}$$
$$S_{01}(-0) = -\tfrac{1}{3}. \tag{10.3}$$

Since the combination $[v + (1, 1)_{sT}]$ appears as an element of the Pfaffian for all the correlations, it is denoted there by S_{01} for brevity.

$$S_{13}[\text{Fig. 4(b)}] = -\begin{vmatrix} -S_{01} & (1,0)_{rs} & (2,1)_{rt} \\ (0,0)_{ls} & (1,1)_{lt} & \\ & S_{01} & \end{vmatrix}, \quad (10.4)$$

$$S_{13}(T_c) = \left[-\frac{8}{9} + \frac{8\sqrt{3}}{3\pi}\right] = +0.581\,321, \quad (10.5)$$

$$S_{13}(-0) = \left[\frac{1}{9} + \frac{2\sqrt{3}}{3\pi}\right] = +0.478\,664. \quad (10.6)$$

S_{13} is the second-nn correlation. The third-nn correlation S_{14} has already been obtained in Sec. 6 as a 2×2 determinant. For completeness the corresponding Pfaffian is

$$S_{14}(\text{Fig. 4c}) = +\begin{vmatrix} S_{01} & (1,1)_{ss} & (2,2)_{st} \\ (0,0)_{ts} & (1,1)_{tt} & \\ & S_{01} & \end{vmatrix}, \quad (10.7)$$

$$S_{14}(T_c) = \left[\frac{16}{9} - \frac{12}{\pi^2}\right] = +0.561\,924, \quad (10.8)$$

$$S_{14}(-0) = \left[\frac{1}{9} - \frac{3}{\pi^2}\right] = -0.192\,852. \quad (10.9)$$

The 4-spin correlation S_{1245} has been treated as an example in Sec. 3, and can be obtained by perturbing two dimers only,

$$S_{1245}(\text{Fig. 4d}) = +\begin{vmatrix} S_{01} & (-1,1)_{ss} & (0,2)_{st} \\ (-2,0)_{ts} & (-1,1)_{tt} & \\ & S_{01} & \end{vmatrix}, \quad (10.10)$$

$$S_{1245}(T_c) = \left[-\frac{140}{9} + \frac{160\sqrt{3}}{3\pi} - \frac{132}{\pi^2}\right] = +0.474\,256, \quad (10.11)$$

$$S_{1245}(-0) = \left[\frac{1}{9} + \frac{4\sqrt{3}}{3\pi} - \frac{6}{\pi^2}\right] = +0.238\,289. \quad (10.12)$$

The 4-spin correlation S_{1234} can also be obtained by perturbing two dimers,

$$S_{1234}(\text{Fig. 4e}) = -\begin{vmatrix} -S_{01} & (-2,0)_{us} & (-1,1)_{ut} \\ (-2,-1)_{ds} & (-1,0)_{dt} & \\ & S_{01} & \end{vmatrix}, \quad (10.13)$$

$$S_{1234}(T_c) = \left[-2 + \frac{40\sqrt{3}}{3\pi} - \frac{48}{\pi^2}\right] = +0.487\,635, \quad (10.14)$$

$$S_{1234}(-0) = \left[-\frac{1}{3} + \frac{4\sqrt{3}}{3\pi} - \frac{3}{\pi^2}\right] = +0.097\,808. \quad (10.15)$$

The 4-spin correlation S_{1235} requires the perturbation of three dimers and hence is given by a 5×5 Pfaffian. By choice of the appropriate spins, two of the dimers can be made to correspond to the correlation S_{13},

$$S_{1235}[\text{Fig. 4(f)}] = +\begin{vmatrix} -S_{01} & (0,1)_{rr} & (1,1)_{rl} & (1,1)_{rs} & (2,2)_{rt} \\ (-1,1)_{lr} & (0,1)_{ll} & (0,1)_{ls} & (1,2)_{lt} & \\ & -S_{01} & (1,1)_{sr} & (2,1)_{st} & \\ & & (0,0)_{ls} & (1,1)_{lt} & \\ & & & S_{01} & \end{vmatrix}, \quad (10.16)$$

$$S_{1235}(T_c) = \left[-\frac{28}{27} + \frac{20\sqrt{3}}{3\pi} - \frac{48}{\pi^2} + \frac{48\sqrt{3}}{\pi^3}\right] = +0.456\,414, \quad (10.17)$$

$$S_{1235}(-0) = \left[-\frac{13}{27} + \frac{2\sqrt{3}}{3\pi} - \frac{3}{\pi^2} + \frac{3\sqrt{3}}{\pi^3}\right] = -0.250\,309. \quad (10.18)$$

The 6-spin correlation S_6 also requires the perturbation of three dimers, which can be chosen so that two of them correspond to the correlation S_{1234}.

$$S_6[\text{Fig. 4(g)}] = +\begin{vmatrix} -S_{01} & (2,1)_{ru} & (2,2)_{rd} & (0,1)_{rs} & (1,2)_{rt} \\ (1,1)_{lu} & (1,2)_{ld} & (-1,1)_{ls} & (0,2)_{lt} & \\ & -S_{01} & (-2,0)_{us} & (-1,1)_{ut} & \\ & & (-2,-1)_{ds} & (-1,0)_{dt} & \\ & & & S_{01} & \end{vmatrix}, \quad (10.19)$$

$$S_6(T_c) = \left[-\frac{1463}{27} + \frac{736\sqrt{3}}{3\pi} - \frac{1008}{\pi^2} + \frac{384\sqrt{3}}{\pi^3}\right] = +0.393\,153, \quad (10.20)$$

$$S_6(-0) = \left[-\frac{5}{27} + \frac{4\sqrt{3}}{3\pi} - \frac{9}{\pi^2} + \frac{6\sqrt{3}}{\pi^3}\right] = -0.026\,803. \quad (10.21)$$

The matrix elements in the Pfaffians for the

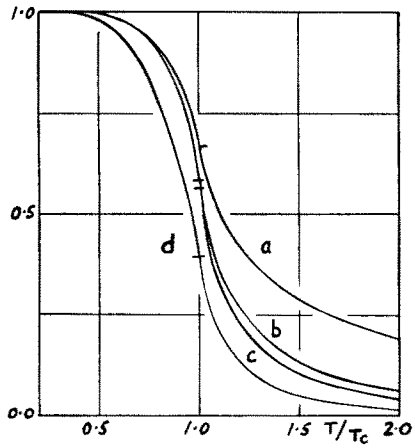


FIG. 5. Ferromagnetic correlations versus T/T_c. (a) S_{01}, (b) S_{13}, (c) S_{14}, (d) S_6. The 4-spin correlations lie between S_{01} and S_6.

correlations are not all independent. By careful use of the transformations in the Appendix the number of double integrals to be evaluated can be reduced to nineteen. In every case the first integration can be performed in terms of the integrals I_n defined previously (4.10). The remaining integrals could then be expressed in terms of elliptic integrals, but this exercise would not be very illuminating. Instead the values at the critical points $T = T_c$ and $T = -0$ have been calculated exactly and at intermediate points we have resorted to numerical integration on a digital computer. Graphs of the versus correlations are presented in Figs. 5 and 6.

As is to be expected, the correlations are all unity at the ferromagnetic zero and decrease steadily as the temperature rises. At the critical point $T = T_c$, indicated by a transverse line on the graphs (Fig. 5), the $[A + B(T - T_c) \ln |T - T_c|]$ dependence of S_{01} leads to a vertical tangent as mentioned before. The antiferromagnetic behavior of the correlations is quite different, (Fig. 6). The 6-spin correlation, S_6(d), changes sign twice, and the third-nearest-neighbor correlation, S_{14}(e), changes sign once at $T/T_c \simeq -0.7$. The remaining correlations decay steadily in magnitude as the temperature rises and have points of inflexion between $-\frac{1}{4}$ and $-\frac{1}{5}$ of T_c.

11. TRANSFORMATION OF CORRELATIONS

A useful check on the values of the correlations calculated in the previous section can be obtained by applying a result derived by Fisher to the triangular lattice.[33] From Eq. (80), p. 980 of Ref. 33 we can obtain a recurrence relation between the correlation S_{01} and the correlations between the six nearest-neighbor spins to s_0. Specifically,

$$\langle s_0 s_g \rangle = \sum_{r=1}^{6} \alpha_r \langle s_r s_g \rangle + \sum_{t=1}^{\binom{6}{3}} \beta_t \langle s_{a_t} s_{b_t} s_{c_t} s_g \rangle + \sum_{u=1}^{\binom{6}{5}} \gamma_u \langle s_{a_u} s_{b_u} s_{c_u} s_{d_u} s_{e_u} s_g \rangle, \qquad (11.1)$$

where in the products $s_{a_t} s_{b_t} s_{c_t}$, $s_{a_u} s_{b_u} s_{c_u} s_{d_u} s_{e_u}$, no spin may appear more than once, and the sums run over all possible permutations of the six nn spins. When the interaction parameters K_i are all equal the subscripts distinguishing the coefficients α, β, γ may be dropped. From the definition of $\langle s_0 s_g \rangle$,

$$\langle s_0 s_g \rangle = \sum_{s_i = \pm 1} s_0 s_g \exp\left(\sum_{(ij)} K s_i s_j\right)/Z_N, \qquad (11.2)$$

select the summation over $s_0 = \pm 1$,

$$\Sigma_0 = \sum_{s_0 = \pm 1} s_0 \exp\left(K s_0 \sum_{i=1}^{6} s_i\right), \qquad (11.3)$$

which can be expanded in the form

$$\Sigma_0 = \left[\alpha \sum_{r=1}^{6} s_r + \beta \sum_{t=1}^{\binom{6}{3}} s_{a_t} s_{b_t} s_{c_t} + \gamma \sum_{u=1}^{\binom{6}{5}} s_{a_u} s_{b_u} s_{c_u} s_{d_u} s_{e_u}\right] \sum_{s_0 = \pm 1} \exp\left(K s_0 \sum_{i=1}^{6} s_i\right). \qquad (11.4)$$

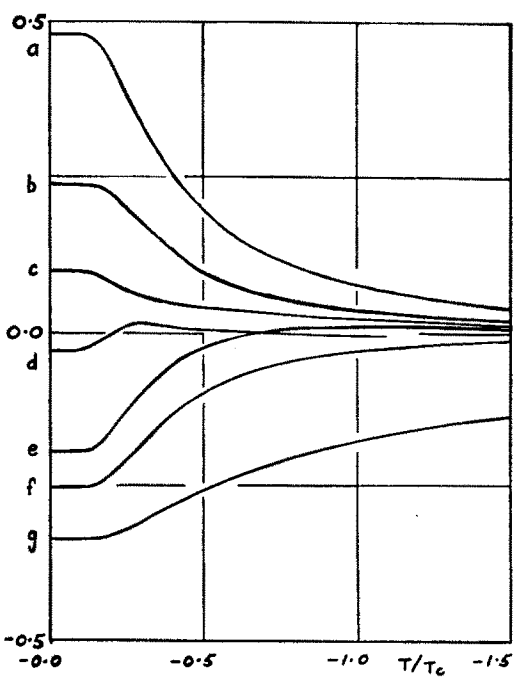


FIG. 6. Antiferromagnetic correlations versus T/T_c. Reading from top to bottom down the axis $T = -0$: (a) S_{13}, (b) S_{1245}, (c) S_{1234}, (d) S_6, (e) S_{14}, (f) S_{1235} and (g) S_{01}.

[33] M. E. Fisher, Phys. Rev. 113, 969 (1959).

Insertion of the transformation (11.4) in (11.2) yields the recurrence relation (11.1). The coefficients α, β, γ can be determined quite simply from (11.3) and (11.4) by multiplying through by $s_{r'}$, $s_{a_l}s_{b_l}s_{c_l}$, and $s_{a_u}s_{b_u}s_{c_u}s_{d_u}s_{e_u}$, in turn, and summing over the 2^6 states of the six-nn spins. For the triangular lattice this yields

$$\alpha = \tfrac{1}{32}[\tanh 6K + 4\tanh 4K + 5\tanh 2K],$$
$$\beta = \tfrac{1}{32}[\tanh 6K - 3\tanh 4K], \qquad (11.5)$$
$$\gamma = \tfrac{1}{32}[\tanh 6K - 4\tanh 4K + 5\tanh 2K].$$

Finally putting $g = 1$ in (11.1) we get the required relationship linking S_{01} and the correlations between the nn spins to s_0.

$$S_{01} = \alpha(1 + 2S_{01} + 2S_{13} + S_{14}) + \beta(4S_{01} + 4S_{13} + 2S_{14} + 4S_{1234} + 4S_{1235} + 2S_{1245}) + \gamma(2S_{1234} + 2S_{1235} + S_{1245} + S_6). \qquad (11.6)$$

The correlations calculated in the previous section satisfy this check, as may easily be verified in particular at the points $T = T_c$ and $T = -0$.

12. PERPENDICULAR SUSCEPTIBILITY (ii)

It now remains to complete the calculation of the perpendicular susceptibility of the triangular lattice by substituting the values of the correlations in Sec. 10 in the formula (9.1), together with the appropriate coefficients [Eqs. (9.3)–(9.7)]. At the ferromagnetic critical point ($kT_c/J = 3.640\,957$, $\tanh 2K = \frac{1}{2}$), writing $\chi_0 = Nm^2/qJ$ for convenience, we find that

$$\chi_\perp^T(T_c)/\chi_0 = \frac{3}{4}\left[\frac{|J|}{kT_c}\left(\frac{2795}{27} - \frac{1444\sqrt{3}}{3\pi} + \frac{2052}{\pi^2} - \frac{816\sqrt{3}}{\pi^3}\right) + \frac{1}{21}\left(-\frac{15323}{27} + \frac{8344\sqrt{3}}{3\pi} - \frac{11844}{\pi^2} + \frac{4704\sqrt{3}}{\pi^3}\right)\right]$$
$$= 1.120\,253. \qquad (12.1)$$

The corresponding value for the quadratic lattice ($q = 4$) is

$$\chi_\perp^Q(T_c)/\chi_0 = 1.136\,951, \qquad (12.2)$$

and for the honeycomb lattice ($q = 3$) is

$$\chi_\perp^H(T_c)/\chi_0 = 1.154\,701. \qquad (12.3)$$

The behavior of $\chi_\perp$ for the honeycomb, quadratic and ferromagnetic triangular lattices is very similar. The parameters at the maximum are

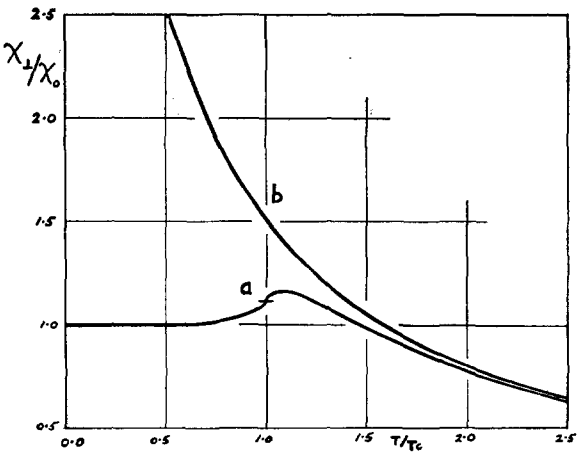


FIG. 7. $\chi_\perp/\chi_c$ vs T/T_c for (a) ferromagnetic triangular, and (b) antiferromagnetic triangular lattices.

$$\chi_\perp^T(T_{max})/\chi_0 = 1.159\,76 \text{ at } T/T_c = 1.077,$$
$$\chi_\perp^Q(T_{max})/\chi_0 = 1.183\,14 \text{ at } T/T_c = 1.088, \qquad (12.4)$$
$$\chi_\perp^H(T_{max})/\chi_0 = 1.213 \quad \text{at } T/T_c \simeq 1.1,$$

where we have quoted Fisher's results for the quadratic and honeycomb lattices. It is to be expected that the values of $\chi_\perp/\chi_0$ follow a trend related to the variation of the coordination number q between the lattices. As has been mentioned already, the antiferromagnetic triangular lattice has a singular point at the absolute zero, $T = -0$, and $\chi_\perp^T$ diverges as $1/T$. The exact form of the divergence is

$$\chi_\perp^T(T \to -0) \simeq \frac{Nm^2}{kT}\left[\frac{4}{27} + \frac{\sqrt{3}}{3\pi} - \frac{3\sqrt{3}}{4\pi^3}\right] = \frac{Nm^2}{kT} \times 0.290\,028. \qquad (12.5)$$

This result should be compared with the conclusions of Sykes and Zucker[19] for the parallel susceptibility in the antiferromagnetic case. By numerical analysis of the series expansion in $v = \tanh K$ for $\chi_\parallel^T$ they conclude that as $T \to -0$

$$\chi_\parallel^T(T \to -0) \simeq (Nm^2/kT) \times 0.138\,89 \pm 5. \qquad (12.6)$$

Both the antiferromagnetic susceptibilities are steadily increasing functions of $|K|$, $\chi_\perp^T$ having the larger amplitude of divergence. $\chi_\perp^T/\chi_0$ has been plotted against T/T_c in Fig. 7 for ferro- and antiferromagnetic interaction energies.

ACKNOWLEDGMENTS

The author would like to thank Dr. M. E. Fisher for suggesting this problem and for his interest and advice during the course of the calculations; and

Professor C. Domb and Dr. Fisher for their helpful criticisms of the first draft of this paper. The award of a postgraduate studentship for 1962–1963 by the University of London is gratefully acknowledged.

The numerical calculations reported in the paper where performed by the University of London Computer Unit Ferranti "Mercury" computer.

APPENDIX

On expanding the determinant $\Delta(\phi_1, \phi_2)$ in Eq. (2.10) by the first row, we get for the cofactors

$$C_{UU} = 2iv_1(1 - v_3^2)\sin\phi_1 + 2iv_3(1 - v_1^2)\sin(\phi_1 + \phi_2), \quad \text{(A1)}$$

$$C_{US} = 1 + v_1^2 + 2v_1v_2v_3 - [v_2v_3(1 + v_1^2) + 2v_1]e^{i\phi_1} - v_2(1 - v_1^2)e^{-i\phi_2} - v_3(1 - v_1^2)e^{i(\phi_1+\phi_2)}, \quad \text{(A2)}$$

$$C_{UR} = 1 - v_3^2 - v_1v_2(1 - v_3^2)e^{i(\phi_1-\phi_2)} - (v_1 + 2v_2v_3 + v_1v_3^2)e^{i\phi_1} - (v_2 + 2v_1v_3 + v_2v_3^2)e^{-i\phi_2}, \quad \text{(A3)}$$

$$C_{UD} = -(1 - v_1^2)(1 - v_3^2) + 4v_1v_3e^{-i\phi_2} + v_2e^{-i\phi_2}[(1 + v_1^2)(1 + v_3^2) + 2v_1(1 - v_3^2)\cos\phi_1 + 2v_3(1 - v_1^2)\cos(\phi_1 + \phi_2)], \quad \text{(A4)}$$

$$C_{UT} = -(1 - v_1^2) + v_2v_3(1 - v_1^2)e^{-i(2\phi_2+\phi_1)} + (v_2 + 2v_1v_3 + v_2v_1^2)e^{-i\phi_2} + (v_3 + 2v_1v_2 + v_3v_1^2)e^{-i(\phi_1+\phi_2)}, \quad \text{(A5)}$$

$$C_{UL} = 1 + v_3^2 + 2v_1v_2v_3 - [v_1v_2(1 + v_3^2) + 2v_3]e^{-i(\phi_1+\phi_2)} - v_1(1 - v_3^2)e^{-i\phi_1} - v_2(1 - v_3^2)e^{-i\phi_2}. \quad \text{(A6)}$$

When the interactions are all equal, a factor of $(1 + v)$ may be removed from the cofactors. This factor, together with the $(1 + v)$ from the term $(1 - v^2)$ multiplying each matrix element in the correlation Pfaffians, can be canceled with the similar factor appearing in $\Delta = (1 + v)^2\Delta'$ [Eq. (5.16)]. The matrix elements $(p, q)_{ij}$ are finite at $v = -1$, $T = -0$.

The transformations for obtaining the other cofactors are summarized in the following table. For example, to obtain the cofactor C_{TS}, apply the transformation **T** to the cofactor C_{UD}. **U** is the identity transformation:

U	S	R	D	T	L
C_{UU}	C_{SS}	C_{RR}	C_{DD}	C_{TT}	C_{LL}
C_{US}	$-C_{SR}$	C_{RD}	$-C_{DT}$	C_{TL}	$-C_{LU}$
C_{UR}	C_{SD}	C_{RT}	C_{DL}	C_{TU}	C_{LS}
C_{UD}	$-C_{ST}$	C_{RL}	$-C_{DU}$	C_{TS}	$-C_{LR}$
C_{UT}	C_{SL}	C_{RU}	C_{DS}	C_{TR}	C_{LD}
C_{UL}	$-C_{SU}$	C_{RS}	$-C_{DR}$	C_{TD}	$-C_{LT}$

(A7)

The transformations are

U	v_1 v_2 v_3	ϕ_1	ϕ_2	$(\phi_1 + \phi_2)$
S	v_2 v_3 v_1	$-\phi_2$	$(\phi_1 + \phi_2)$	ϕ_1
R	v_3 v_1 v_2	$-(\phi_1 + \phi_2)$	ϕ_1	$-\phi_2$
D	v_1 v_2 v_3	$-\phi_1$	$-\phi_2$	$-(\phi_1 + \phi_2)$
T	v_2 v_3 v_1	ϕ_2	$-(\phi_1 + \phi_2)$	$-\phi_1$
L	v_3 v_1 v_2	$(\phi_1 + \phi_2)$	$-\phi_1$	ϕ_2

(A8)

When the interactions are all equal, the integrals containing the cofactor C_{UL} can be transformed to ones containing C_{US}, and similarly for C_{UT} and C_{UR}.

Nested Hilbert Spaces in Quantum Mechanics. I*

ALEXANDER GROSSMANN†
Physics Department, New York University, New York, and Institute for Advanced Study, Princeton, New Jersey
(Received 30 April 1963)

A nested Hilbert space is a pair of Hilbert spaces H_0, H_1, each of which is in a certain sense identified with a dense subset of the other. These structures are used here to study analytic continuation into "unphysical sheets" and to discuss nonnormalizable states of quantum-mechanical systems.

I. INTRODUCTION

IN the practice of quantum mechanics the Hilbert space framework[1] is often found too restrictive. The need for more room is felt particularly in questions involving the analytic continuation into "unphysical sheets" where one has to deal with functions that increase at infinity.

The aim of this paper is to point out that structures only one step removed from Hilbert space can be used to handle such questions. They consist—roughly speaking—of a pair of Hilbert spaces, each of which may be identified with a dense subset of the other. Such identifications are possible because every element of a Hilbert space can also play the role of a linear functional.

Structures of this kind are not new. In one form or another, they are a standard tool of functional analysis and had, for example, been used by Friedrichs[2] about thirty years ago.

The present work was strongly influenced by the books of Gel'fand and collaborators on distribution theory.[3] The spaces introduced by them in Vol. IV involve sequences (rather than pairs) of Hilbert spaces. For the purposes of this paper, a pair of Hilbert spaces is sufficient, since there is no need to make continuous all sensible operations (such as differentiation).

Section II contains the starting definitions and a few examples. The relationship between operators in the two Hilbert spaces is studied in Sec. III. These relationships are used, in Secs. IV–VI, to investigate analytic continuation into "unphysical sheets". Section VII is concerned with the statistical interpretation of nonnormalizable states in quantum mechanics, and Sec. VIII with resonances and scattering.

II. DEFINITIONS AND EXAMPLES

Definition: Let H_0 and H_1 be infinitely dimensional, separable Hilbert spaces.[4] An *inclusion* of H_1 into H_0 is a linear mapping E_{01} which

(a) is defined for every $f \in H_1$ and continuous;
(b) has range dense in H_0;
(c) is one-to-one.

The inverse of E_{01} always exists, but need not be continuous. The *adjoint* of E_{01}, *denoted by* E_{10} and defined by

$$(f, E_{01}g)_0 = (E_{10}f, g)_1 \qquad (f \in H_0, g \in H_1),$$

is an inclusion of H_0 into H_1. To see this, notice that (a), (b), (c) for E_{01} imply, respectively, (a), (c), (b) for E_{10}.

Definition: A nested Hilbert space (H_0, H_1, E_{01}) is a structure that consists of two Hilbert spaces H_0, H_1, an inclusion E_{01} of H_1 into H_0, and the adjoint inclusion E_{10} of H_0 into H_1.

Note added in proof. A slightly different definition, involving at least three different spaces can be used to simplify many arguments of this paper.

The structure of nested Hilbert space determines the numbers $(f, s)_0$ (scalar products in H_0; $f \in H_0$, $s \in H_0$), the numbers $(h, k)_1$ (for all $h \in H_1$, $k \in H_1$) and the numbers $(f, E_{01}g)_0 = (E_{10}f, g)_1$ for all $f \in H_0$ and $g \in H_1$.

Example 1: Let R^n be the real n-dimensional

* The research reported in this paper has been supported in part by the Army Research Office (Durham) under Contract No. DA-ARO-(D)-31-124-G133, and by the National Science Foundation.

† Present address: Courant Institute of Mathematical Sciences, New York University, New York, New York.

[1] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, New Jersey, 1955).

[2] K. Friedrichs, Math. Ann. 109, 465 (1934). Reproduced in F. Riesz and B. Sz-Nagy, *Functional Analysis* (translation) (Frederick Ungar Publishing Company, New York, 1955), pp. 330–336.

[3] I. M. Gel'fand, G. E. Silov, and N. Ya. Vilenkin, *Generalized Functions* (State Press for Physics and Mathematics, Moscow, 1958–61), Vols. I–IV.

[4] It is assumed that the reader is familiar with the terminology of Hilbert space theory. See the books of Refs. 1 and 2, or M. H. Stone, *Linear Transformations in Hilbert Space and their Applications to Analysis* (American Mathematical Society Colloquium Publications, New York), Vol. XV.

Euclidean space, and $x \in R^n$. Let $M(x)$ be a continuous function such that $0 < M(x) \leq 1$ for every $x \in R^n$. Let $H_0 = L^{(2)}$ be the Hilbert space of functions square-integrable over R^n. Denote by $H_1 = L^{(2)}(M^{-2})$ the Hilbert space of measurable functions $f(x)$ for which

$$(f, f)_1 = \int M^{-2}(x)\, |f(x)|^2\, dx < \infty.$$

Then the correspondence E_{01}: $(E_{01}f)\ (x) = f(x)$ for every $f \in H_1$, is an inclusion of H_1 into H_0. The adjoint inclusion is:

$$(E_{10}g)(x) = M^2(x)g(x), \text{ for every } g \in L^{(2)}.$$

Example 2: Let H_0 and H_1 be arbitrary. Let $\{h_n\}$ be an orthonormal basis in H_0, and $\{e_n\}$ an orthonormal basis in H_1. Let λ_n be a sequence of positive numbers such that $\lim \lambda_n = 0$ as $n \to \infty$. Then

$$E_{01} = \sum_{n=0}^{\infty} |h_n > \lambda_n < e_n| \tag{1.1}$$

is[5] a completely continuous inclusion of H_1 into H_0. Its adjoint is

$$E_{10} = \sum_{n=0}^{\infty} |e_n > \lambda_n < h_n|. \tag{1.2}$$

Every completely continuous inclusion of H_1 into H_0 can be written in the form (1.1), with a suitable choice of the bases $\{e_n\}$, $\{h_n\}$ and of the numbers λ_n.

Example 3: Specialize Example 2 by considering $H_0 = L^{(2)}$ (over R^1). In H_0, consider the orthonormal basis of Hermite functions $h_n(x)$. Define H_1 as the Hilbert space of functions $f(x) = \sum_{n=0}^{\infty} f_n h_n(x)$ such that $\sum_{n=0}^{\infty} |f_n|^2 \lambda_n^{-2} < \infty$, with the scalar product

$$(f, g)_1 = \sum (f_n^* g_n / \lambda_n^2).$$

Let, again, as in Example 1, $(E_{01}f)(x) = f(x)$. Then $e_n(x) = \lambda_n h_n(x)$, and

$$(E_{10}g)(x) = \int K(x, y)g(y)\, dy,$$

where

$$K(x, y) = \sum_{n=0}^{\infty} h_n(x)\lambda_n^2 h_n(y). \tag{1.3}$$

If $\lambda_n = \kappa^n$ with $\kappa < 1$, then $K(x, y)$ can be explicitly computed. The Fourier transformation is a unitary operator in H_1.

Example 3a: Impose on the numbers λ_n the condition

$$\lambda_n^2 \leq \Lambda/\Gamma(n + 1) \qquad (n = 0, 1, 2 \cdots)$$

for some constant Λ. Simple estimates[6] then show that every $f \in H_1$ is entire analytic and that, for every number $b > 1$, there exists a constant $C = C(b)$, independent of f, such that

$$|f(x + iy)| \leq C\, ||f||_1 \exp(-\tfrac{1}{2}b^{-1}x^2 + \tfrac{1}{2}by^2). \tag{1.4}$$

It follows from (1.4) that, for every complex w, there exists in H_1 a vector δ_w such that,[7] for every $f \in H_1$,

$$(\delta_w, f)_1 = f(w). \tag{1.5}$$

It is convenient to consider in H_1 the family $\delta(u, w)$ of operators (of trace class) defined by

$$\delta(u, w) = |\delta_u\rangle\langle\delta_w| \qquad (u, w \text{ complex}). \tag{1.6}$$

Example 4[8]: Let again $H_0 = L^{(2)}$. Denote by S the vector space of infinitely differentiable complex-valued functions $f(x)$ such that, for $m, n = 0, 1, 2 \cdots$, the function $x^m D^n f(x)$ is bounded. Denote by p the operator $i^{-1}D$ and consider in S the sequence of positive-definite operators

$$L_0 = 1,$$
$$\cdots$$
$$L_n = xL_{n-1}x + pL_{n-1}p$$
$$\cdots$$

Let $\alpha > 0$ and $A > 0$. Define $H_1 = S(\alpha; A)$ as the Hilbert space of functions $f \in S$ for which

$$\sum_n [\Gamma(\alpha n)A^n]^{-2}(f, L_n f)_0 < \infty$$

with the scalar product

$$(f, g)_1 = \sum_{n=0}^{\infty} [\Gamma(\alpha n)A^n]^{-2}(f, L_n g)_0.$$

The Fourier transform is a unitary operator in $S(\alpha; A)$.

Every $f \in S(\alpha; A)$ decreases at $x \to \infty$ as $\exp(-a\ |x|^{1/\alpha})$ $(a > 0)$. If $\alpha < 1$, then every $f \in S(\alpha; A)$ is entire analytic of order $\leq (1 - \alpha)^{-1}$.

For $\alpha < \frac{1}{2}$, the space $S(\alpha; A)$ is trivial, i.e., consists only of the element $f = 0$.

[5] Equation (1.1) means $E_{01}\, f = \Sigma\lambda_n(e_n, f)_1\, h_n$ for every $f \in H_1$.

[6] See Appendix.

[7] Hilbert spaces of analytic functions—with elements δ_w that satisfy (1.5)—appear in many questions. See V. Bargmann Commun. Pure Appl. Math. **14**, 187 (1961) and H. Meschkowski, *Hilbertsche Räume mit Kernfunktionen* (Springer-Verlag, Berlin, 1962).

[8] The Hilbert spaces $S(\alpha, A)$ are closely related to the vector spaces "of type S" studied in Vol. II of Ref. 3. See also S. Mandelbrojt, Ann. Sci. Ec. Norm. Sup. Ser. 3 **77**, 154 (1960).

Note added in proof. The connection between Example 3 and Example 4 is discussed in a forthcoming paper entitled "Hilbert spaces of type S."

Remark: The bilinear form

$$(f, g) = (f, E_{10}h)_1 = (E_{01}f, h)_0 \qquad (f \in H_1, h \in H_0)$$

is sometimes more important than the scalar products in H_0 and H_1. This motivates the following definition:

The nested Hilbert space (H_0, H_1, E_{01}) is said to be *equivalent* to the nested Hilbert space $\bar{H}_0, \bar{H}_1, \bar{E}_{01})$ if there exists a linear mapping L from (the whole of) H_0 onto $\bar{H}_0$, and a linear mapping J from H_1 onto $\bar{H}_1$ such that

(a) L and J are bounded and have bounded inverses;
(b) the equation $E_{01} = L^* \bar{E}_{01} J$ holds.

These conditions are equivalent to the requirement that, for every $h \in H_0$ and every $f \in H_1$, one should have

$$(h, E_{01}f)_0 = (\bar{h}, \bar{E}_{01}\bar{f})_{\bar{0}},$$

with $\bar{h} = Lh$ and $\bar{f} = Jf$.

In particular, if L and J are unitary, then the two nested Hilbert spaces are isomorphic.

Remark: It is sometimes convenient to identify elements that are in correspondence through an inclusion. Since inclusions are one-to-one, this does not lead to contradictions. One has to remember, however, that it becomes then impossible to maintain, in *both* Hilbert spaces H_0 and H_1, the usual identifications between elements and continuous linear functionals. For example, assume that one has identified every $f \in H_1$ to $E_{01}f \in H_0$, and every $h \in H_0$ to $E_{10}h \in H_1$. If then one identifies every $h \in H_0$ with the linear functional $\varphi \to (h, \varphi)_0$, one is committed to identify $f \in H_1$ with the linear functional $\chi \to (f, E_{10}E_{01}\chi)_1 = (E_{01}f, E_{01}\chi)_0$ rather than with the functional $\chi \to (f, \chi)_1$.

To illustrate this, consider the case where (a) $H_0 = L^{(2)}$, (b) H_1 is a space of functions, and (c) $(E_{01}f)(x) = f(x)$ $(f \in H_1)$. Then

$$(f, E_{10}E_{01}\chi)_1 = (E_{01}f, E_{01}\chi)_0 = \int f^*(x)\chi(x)\, dx.$$

It may happen that the linear functional $\chi \to (f, \chi)_1$ can also be written as

$$(f, \chi)_1 = \int F^*(x)\chi(x)\, dx \qquad (\chi \in H_1),$$

but the function F is in general different from f. If the functions $\chi \in H_1$ decrease as $|x| \to \infty$, then F may increase as $|x| \to \infty$.

III. OPERATORS

1. Let A be any linear operator with domain and range contained in H_0. Consider, in H_1, the operators

$$e_{10}(A) = E_{10}AE_{01} \tag{2.1}$$

and

$$j_{10}(A) = (E_{01})^{-1}AE_{01}. \tag{2.2}$$

The domain of $e_{10}(A)$ is $(E_{01})^{-1}D_A$ where D_A is the domain of A.

The domain of $j_{10}(A)$ is the subset of $(E_{01})^{-1}\, D_A$ consisting of vectors which are also such that $AE_{01}f \in E_{01}H_1$.

If A is bounded and self-adjoint, then $e_{10}(A)$ is also bounded and self-adjoint, while $j_{10}(A)$ is, in general, neither bounded nor symmetric.

There exist in H_1 operators which are not of the form $e_{10}(A)$ for any operator A in H_0. Such is, for example, the operator of orthogonal projection on a one-dimensional space spanned by a vector $f \in H_1$, when f does not belong to $E_{10}H_0$.

Let $f \in H_1$ be a vector in the domain of $e_{10}(A)$. Then, for every $g \in H_1$,

$$(g, e_{10}(A)f)_1 = (g, E_{10}AE_{01}f)_1 = (E_{01}g, AE_{01}f)_0.$$

So, certain matrix elements of A—namely the ones between $D_A \cap E_{01}H_1$ and $E_{01}H_1$—can also be computed in H_1, as matrix elements of $e_{10}(A)$.

The transition from A to $e_{10}(A)$ consists essentially in chosing a suitable subset of matrix elements of A and then "putting together" these matrix elements to form an operator in H_1.

The operator $j_{10}(A)$ is, if one identifies $f \in H_1$ with $E_{01}f$, the restriction of A to $E_{01}H_1 \subseteq H_0$. Its matrix elements are different from the matrix elements of A, because of the different definition of scalar product.

The adjoint $j_{10}^*(A)$ of $j_{10}(A)$ is—if it exists and if one identifies $h \in H_0$ with $E_{10}h$—the extension of A^* from $E_{10}H_0$ to its natural domain of definition. *The eigenvalues of $j_{10}^*(A)$ include the "improper" eigenvalues of A^*, long familiar in quantum mechanics.*

2. Example: Consider the nested Hilbert space of Example 3a, Sec. II. In $H_0 = L^{(2)}$ consider the operator A of multiplication by x. Then $j_{10}(A)$ is defined for all $f \in H_1$ which are such that $xf(x)$ also belongs to H_1; it transforms f into xf. This operator has no eigenvalues.

The domain of this operator $j_{10}(A)$ contains, e.g., all finite linear combinations of the elements of the basis $\{e_n\}$; so $j_{10}(A)$ is densely defined in H_1, and the adjoint $j_{10}^*(A)$ exists.

Every complex number w^* is an eigenvalue of $j_{10}^*(A)$, and the vector $\delta_w \in H_1$, defined by (1.5) is the corresponding eigenvector. This is so because

$$(j_{10}^*(A)\delta_w, g)_1 = (\delta_w, j_{10}(A)g)_1 = wg(w) = (w^*\delta_w, g)_1 \qquad (2.3)$$

for every g in the domain of $j_{10}(A)$.

3. Definition[9]*:* Let A be an operator in H_0 such that

(a) A is densely defined (so that the adjoint A^* exists);
(b) $j_{10}(A^*)$ is also densely defined [so that its adjoint $j_{10}^*(A^*)$ also exists].

The complex number z is said to be a *generalized eigenvalue* of A if there exists in H_1 a nonzero vector f such that

$$j_{10}^*(A^*)f = zf.$$

This definition includes the usual one: Let $Ah = zh$ ($h \in H_0$, $h \neq 0$). Notice that $j_{10}^*(A^*) \supseteq E_{10}A^{**}E_{10}^{-1} \supseteq E_{10}AE_{10}^{-1}$. So $j_{10}^*(A^*)f = zf$, with $f = E_{10}h$.

4: There exists another generalization of the concept of eigenvalue of A. An isolated discrete eigenvalue of A is a pole of the resolvent $(z - A)^{-1}$; it is also,—as is seen below—a pole of $e_{10}((z - A)^{-1})$ for any inclusion E_{10}. It is natural to ask whether $e_{10}((z - A)^{-1})$ can be analytically continued beyond the domain of analyticity of $(z - A)^{-1}$ and whether the poles of such an analytic continuation have some of the properties of eigenvalues. These questions are discussed in the sections that follow, but a definition can be given already here:

Let A be a closed operator in H_0. Denote by $\rho(A)$ the resolvent set of A, i.e., the set of complex numbers z such that the operator $(z - A)^{-1}$ is bounded. Let Δ_0 be a connected component of $\rho(A)$.

Definition: The *discrete spectrum* of A, with respect to the nested Hilbert space $(H_0, H_1; E_{01})$ and with respect to Δ_0 consists of the isolated singularities of the analytic operator-valued function obtained by analytic continuation of $e_{10}((z - A)^{-1})$.

The elements of the discrete spectrum are, in general, points on a Riemann surface.

[9] For a discussion with the help of distribution theory, see Ref. 3, Vol. IV, pp. 133–153.

IV. ANALYTIC CONTINUATION

The main purpose of this section is to show that "unphysical sheets"[10] appear in the very simplest examples of analytic continuation of operator families of the form $e_{10}((z - A)^{-1})$.

1. Analytic operator families: Let Δ be an open connected set of the z plane of complex numbers. A family $G(z)$ ($z \in \Delta$) of bounded operators in a Hilbert space H is said to be holomorphic in Δ if all the functions $(f, G(z)g)$ [with $f \in H$, $g \in H$] are holomorphic in Δ. It can be shown[11] that this implies the existence of a Taylor expansion—convergent in the operator norm—around every point of Δ.

If two operator families coincide in an open subset of Δ, then they coincide in the whole of Δ; (principle of analytic continuation).

The natural boundary of an analytic operator family is defined as a line across which analytic continuation is impossible. If any one of the matrix elements $(f, G(z)g)$ has a natural boundary, then the operator family $G(z)$ has the same line as its natural boundary. However, an operator family can have a natural boundary even if every one of its matrix elements has only isolated singularities.

No new concepts are required in order to define the analyticity of an operator family on a Riemann surface. Every point on a Riemann surface has neighborhoods that correspond to open sets in the complex plane in the local uniformizing variable.[12] Analyticity on the Riemann surface is defined as analyticity in these parameters.

2. Families of the form $e_{10}(G(z))$. Let now $(H_0, H_1; E_{01})$ be a nested Hilbert space, and $G(z)$ an operator family of H_0, holomorphic in an open connected set Δ. The matrix elements of $e_{10}(G(z))$ are a subset of the matrix elements of $G(z)$. So it may be expected that $e_{10}(G(z))$ will sometimes have an analytic continuation beyond Δ.

Consider, for example, the operator family $B(k)$ in $H_0 = L^{(2)}$, defined by the kernel $B(x, y; k) = e^{ik|x-y|}$ for Im $k > 0$. This operator family cannot be

[10] For the various aspects of resonances see C. L. Dolph, *Positive Real Resolvents and Linear Passive Hilbert Systems* Ann. Acad. Sci. Fennicae Ser. **AI 336**, 9 (1963) which also contains an extensive bibliography. The author is indebted to Professor Dolph for sending him this preprint and for a very profitable discussion.

[11] See, e.g., H. G. Garnir, *Les Problemes aux Limites de la Physique Mathématique* (Birkhäuser, Basel, 1958), p. 37–50, or E. Hille and R. S. Phillips, *Functional Analysis and Semigroups* (American Mathematical Society Colloquium Publications, Providence, Rhode Island, 1957), p. 92.

[12] See, e.g., H. Behnke and F. Sommer, *Theorie der Analytischen Funktionen einer Komplexen Veränderlichen* (Springer-Verlag, Berlin, 1962), Chap. VI.

analytically continued into the lower half-plane in k. However, if H_1 consists of functions that decrease sufficiently fast, then $e_{10}(B(k))$ may be analytically continued into the lower half-plane.

Returning to the general case: Since $E_{01}H_1$ is dense in H_0 and since $G(z)$ is bounded for every $z \in \Delta$, the knowledge of the matrix elements of $e_{10}(G(z))$ is sufficient to determine $G(z)$. It is sometimes convenient to write $G(z)$ in a form which involves the analytic continuation of $e_{10}(G(z))$ beyond Δ.

If $e_{10}(G(z))$ can be analytically continued into a domain $\Delta_1 \supset \Delta$ that has several sheets, then the new sheets are sometimes called "unphysical".

In most of the examples that follow, the family $G(z)$ is of the form $G(z) = (z - A)^{-1}$.

3. Examples: (a) Let $H_0 = L^{(2)}(R^1)$. Denote by X the operator $f \to xf$ of multiplication by x. The resolvent $(z - X)^{-1}$ is defined for $\operatorname{Im} z \neq 0$ and it consists of *two* analytic operator-valued functions

$$G^{(+)}(z) = (z - X)^{-1} \qquad (\operatorname{Im} z < 0), \tag{4.1}$$

and

$$G^{(-)}(z) = (z - X)^{-1} \qquad (\operatorname{Im} z > 0). \tag{4.2}$$

The real axis is the natural boundary for each of these two operator families.

Let H_1 be the Hilbert space defined in Example 3a of Sec. II. It consists of functions $f(x)$ such that the numbers $|(h_n, f)_0|$ (where the h_n are Hermite functions) decrease sufficiently fast. Every such function f can be analytically continued into the whole plane of the complex variable $s = x + iy$. The inclusion E_{01} assigns, to every $f(s) \in H_1$, the function $f(x) \in L^{(2)}$ [the restriction of $f(s)$ to the real axis].

Proposition: There exist in H_1 two families of bounded operators, to be denoted by $G_1^{(+)}(z)$ and $G_1^{(-)}(z)$, such that

(a) Both $G_1^{(+)}(z)$ and $G_1^{(-)}(z)$ are *entire analytic* in z;

(b) For $\operatorname{Im} z < 0$, $\quad G_1^{(+)}(z) = e_{10}(G^{(+)}(z))$; (4.3)

(b′) For $\operatorname{Im} z > 0$, $\quad G_1^{(-)}(z) = e_{10}(G^{(-)}(z))$. (4.4)

Proof: Consider in H_1 the bilinear functional

$$B^{(+)}(f, g; z) = \int (z - s)^{-1} f^*(s^*) g(s)\, dx \qquad (f, g \in H_1), \tag{4.5}$$

where $s = x + iy$ and $y > \operatorname{Im} z$ (i.e., the path of integration is a parallel to the real axis passing above z). By (1.4),

$$\int |(z - s)^{-1} f^*(s^*) g(s)|\, dx \leq (y - \operatorname{Im} z)^{-1} \int |f^*(s^*) g(s)|\, dx$$

$$(y - \operatorname{Im} z)^{-1}\, ||f||_1\, ||g||_1\, C \int \exp(-b^{-1}x^2 + by^2)\, dx \leq \text{const}\, ||f||_1\, ||g||_1.$$

So the integral (4.5) is absolutely convergent and defines, for fixed z, a continuous bilinear functional of f and g. Consequently there exists an operator $G_1^{(+)}(z)$ such that

$$(f, G_1^{(+)}(z) g)_1 = B^{(+)}(f, g; z).$$

Since the functions f and g are entire analytic, the integral (4.5) is independent of the choice of y (i.e., of the path of integration), as long as $y > \operatorname{Im} z$. It follows that $G_1^{(+)}(z)$ is entire analytic and that, for $\operatorname{Im} z < 0$, the equality (4.3) holds.

Similarly, one defines the operator family $G_1^{(-)}(z)$ by

$$(f, G^{(-)}(z) g)_1 = \int f^*(s^*)(z - s)^{-1} g(s)\, dx \tag{4.6}$$

$(f \in H_1,\ g \in H_1,\ s = x + iy,\ y < \operatorname{Im} z)$,

and verifies that (4.4) holds. This proves the proposition.

(b) The operator families $G_1^{(+)}(z)$ and $G_1^{(-)}(z)$ are related by

$$G_1^{(-)}(z) - G_1^{(+)}(z) = -2i\pi\, \delta(z^*, z), \tag{4.7}$$

and

$$[G_1^{(-)}(z)]^* = G_1^{(+)}(z^*). \tag{4.8}$$

The operators $\delta(z^*, z)$ have been defined in (1.6) and in the Appendix. The equality (4.7) follows from the Cauchy formula, while (4.8) is obtained by a substitution $s \to s^*$ in the integrals.

(c) The Fourier transform of $G_1^{(-)}(z)$ is $\tilde{G}_1^{(-)}(z) = FG_1^{(-)}(z)F^{-1}$, where F is the unitary operator of Fourier transformation in H_1. A simple calculation shows that, for every $f \in H_1$, $g \in H_1$ and for every complex z,

$$(f, \tilde{G}_1^{(-)}(z) g)_1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f^*(p') K^{(-)}(p', p''; z) g(p'')\, dp'\, dp'',$$

where

$$K^{(-)}(p', p''; z) = i\theta(p' - p'') \exp(z\,|p' - p''|).$$

(d) Consider now, in $L^{(2)} = H_0$ the operator

X^2 of multiplication by x^2. If k is a number such that

$$\operatorname{Im} k > 0 \quad \text{and} \quad k^2 = z,$$

then the operator identity

$$(z - X^2)^{-1} = (2k)^{-1}[G^{(-)}(k) - G^{(+)}(-k)]$$

holds. Consequently,

$$e_{10}((z - X^2)^{-1}) = (2k)^{-1}[G_1^{(-)}(k) - G_1^{(+)}(-k)]. \quad (4.9)$$

It has been shown above that $G_1^{(-)}(k)$ and $G_1^{(+)}(-k)$ can be analytically continued, from the half-plane $\operatorname{Im} k > 0$, into the whole plane. This means that $e_{10}((z - X^2)^{-1} = G_1(z)$ can be analytically continued, from the cut z plane, into the finite points of the Riemann surface of $z^{\frac{1}{2}}$.

If ζ' and ζ'' are two points on the Riemann surface lying above the same number z, then

$$G_1(\zeta') - G_1(\zeta'') = i\pi k^{-1}[\delta(k^*, k) - \delta(-k^*, -k)],$$

where $k^2 = z$ and the sign of k is irrelevant.

(e) The preceding results can be extended to an arbitrary polynomial and to several independent variables. The one-dimensional paths of integration are then replaced by so-called Hörmander ladders.

(f) Let $H_0 = L^{(2)}$ (over R^n). Consider in H_0 a family of integral operators

$$G(z) : f(x) \to \int G(x, x'; z)f(x')\,dx'$$

$$(f \in L^{(2)},\ z \in \Delta_0)$$

holomorphic in some open set Δ_0. Let $M(x)$ be a continuous function such that $0 < M(x) \leq 1$ and that, *still in* $L^{(2)}$; the family

$$B(z) : f(x) \to \int B(x, x'; z)f(x')\,dx' \quad (f \in L^{(2)}, z \in \Delta_1)$$

$$B(x, x'; s) = M(x)G(x, x'; z)M(x')$$

is holomorphic in a set $\Delta_1 \supset \Delta_0$.

Then, in the Hilbert space $L^{(2)}(M^{-2})$ of Example 1, Sec. II, the operator family $e_{10}(G(z))$ can be analytically continued into Δ_1.

This assertion follows easily from $(E_{01}f)(x) = f(x)$, $(E_{10}g)(x) = M^2(x)g(x)$, and the fact that $f \to M(x)f$ is a unitary correspondence between $L^{(2)}(M^{-2})$ and $L^{(2)}$.

For example, if $G(x, x'; z)$ is the familiar kernel, (in R^3),

$$G^{(0)}(x, x'; z) = -(4\pi)^{-1}\,|x - x'|^{-1}$$

$$\times \exp(z^{\frac{1}{2}}\,|x - x'|) \qquad (\operatorname{Im} z^{\frac{1}{2}} > 0), \qquad (4.10)$$

and if $M(x) = \exp(-x^2)$, then $e_{10}(G(z))$ can be analytically continued to a family entire analytic on the Riemann surface of $z^{\frac{1}{2}}$. This example is closely related to the Example (d) above.

Remark: The operators of Example (a) are quite familiar in the form $1/(x - z \pm i\epsilon)$. The point here is that—due to the regularity properties of elements of H_1 and the definition of scalar product in H_1—these operators are defined for all z and depend analytically on z.

V. PRESERVATION OF ISOLATED SINGULARITIES

Let there be given, in the Hilbert space H_0, a family $A_0(z)$ of operators holomorphic in an open connected set Δ_0.

Define $A_1(z)$, $(z \in \Delta_1)$ as the family of operators in H_1 obtained by continuing analytically, as far as possible, the family $e_{10}(A_0(z)) = E_{10}A_0(z)E_{01}$, $(z \in \Delta_0)$. Clearly $\Delta_1 \supseteq \Delta_0$, since the operators E_{01} and E_{10} are bounded. The examples of the preceding section show that Δ_1 can be considerably larger than Δ_0.

It is shown now that, no matter what E_{01} and H_1 are, this enlargement of the domain of analyticity cannot occur at the expense of isolated singularities, but results from a receding of natural boundary lines.

An easy consequence is that the position of isolated singularities of $A_1(z)$ is to a certain degree independent of the choice of H_1 and of E_{01}.

1. A point z_0 is said to be an isolated singularity of an operator family $A(z)$ if there exists an open neighborhood Z of z_0 such that $A(z)$ is holomorphic in $Z\backslash\{z_0\}$. An isolated singularity z_0 is said to be removable, if $A(z_0)$ can be defined so that $A(z)$ becomes holomorphic in all of Z.

It can be shown that an isolated singularity at z_0 is removable if and only if the bound norm $||A(z)||$ remains bounded in some neighborhood of z_0.

An isolated singularity which is not removable is said to be a pole of $A(z)$ if, for some positive integer n, the singularity at z_0 of the family $(z - z_0)^n A(z)$ is removable. If no such n exists, then z_0 is said to be an essential singularity of $A(z)$.

2. Theorem: Let A_0 be a family of operators in H_0, having an isolated singularity at a point z_0. Let H_1 be any inclusion of some Hilbert space H_1 into H_0. Then the family $e_{10}(A_0(z))$ has, at z_0, an isolated singularity of the same kind as $A_0(z)$.

Proof: It is sufficient to show that the singularity of $e_{10}(A_0(z))$ at z_0 cannot be "less bad" than the singularity of $A_0(z)$. This follows immediately from

Lemma: If z_0 is a removable singularity of

$e_{10}(A_0(z))$, then z_0 is also a removable singularity of $A_0(z)$.

In order to prove the lemma, consider a closed contour $\mathcal{C}$ around z_0, such that z_0 is the only singularity of $e_{10}(A_0(z))$ in the interior of $\mathcal{C}$. We show that $||A_0(z)||$ is bounded in the interior of $\mathcal{C}$.

Since the singularity of $e_{10}(A_0(z))$ was assumed removable, the function $(f, e_{10}(A_0(z)(g)_1$, suitably completed at z_0, is holomorphic in the interior of $\mathcal{C}$ for every $f \in H_1$, $g \in H_1$. The absolute value of this function cannot have a maximum in the interior of $\mathcal{C}$. So, for every z in the interior of $\mathcal{C}$,

$$|(f, e_{10}(A_0(z))g)_1| = |(E_{01}f, A_0(z)E_{01}g)_0| \leq C\,||E_{01}f||_0\,||E_{01}g||_0,$$

where $C = \sup ||A_0(w)||$ $(w \in \mathcal{C})$. The number C is finite, since $A_0(z)$ is holomorphic in some neighborhood of every point of $\mathcal{C}$. Since $E_{01}H_1$ is dense in H_0, it follows that $||A_0(z)||$ is bounded. This proves the lemma and the theorem.

3. Let now $\bar{H}_1$ be a Hilbert space different from H_1, and $\bar{E}_{01}$ an inclusion of $\bar{H}_1$ into H_0. The above theorem makes it possible to compare—under suitable assumptions—certain analyticity properties of $\bar{A}_1(z)$ [the analytic continuation of $\bar{E}_{10}A_0(z)\bar{E}_{01}$] with analyticity properties of $A_1(z)$. Namely:

If there exists a Hilbert space H_2, an inclusion E_{12} of H_2 into H_1, and an inclusion $\bar{E}_{12}$ of $\bar{H}_2$ into $\bar{H}_1$ such that

$$E_{01}E_{12} = \bar{E}_{01}\bar{E}_{12}, \tag{5.1}$$

then, in the region where both A_1 and $\bar{A}_1$ are holomorphic except for isolated singularities, the isolated singularities of $\bar{A}_1$ are the same as the ones of A_1.

In order to see this, notice that $E_{02} = E_{01}E_{12} = \bar{E}_{01}\bar{E}_{12}$ is an inclusion of H_2 into H_0. So the domain of holomorphy of $A_2(z)$ [the analytic continuation of $E_{20}A_0(z)E_{02}$] contains both the domain of holomorphy of $A_1(z)$ and that of $\bar{A}_1(z)$. By the preceding theorem, the isolated singularities of any one of these operators appear as isolated singularities of A_2.

A simple example of (5.1) occurs if E_{01} and $\bar{E}_{01}$ are of the form

$$E_{01} = \sum |h_n > \lambda_n < e_n|, \qquad \bar{E}_{01} = \sum |h_n > \bar{\lambda}_n < \bar{e}_n|, \tag{5.2}$$

where $\{e_n\}$ and $\{\bar{e}_n\}$ are arbitrary orthonormal bases in H_1 and $\bar{H}_1$, respectively, while $\{h_n\}$ is an orthonormal basis in H_0; *the same* for both operators (5.2).

4. It has been shown that the discrete spectrum defined at the end of Sec. III contains all the isolated eigenvalues of A. Indeed, such eigenvalues are poles of $(z - A)^{-1}$. By the theorem on preservation of isolated singularities, they are also poles of $e_{10}((z - A)^{-1})$.

The discrete spectrum with respect to a nested Hilbert space consists, in general, not of numbers, but of points on a Riemann surface. This is so because the family $G_1(z)$ is, in general, defined on a Riemann surface, as shown by the examples of Sec. IV.

If A is a Hamiltonian then the points of the discrete spectrum that belong to $\rho(A)$ (i.e., lie in the "physical sheet") are called binding energies, while the remaining points are called virtual binding energies and resonances.

VI. INTERACTIONS

In the perturbation theory of Hilbert space operators, it has long been known that a discrete eigenvalue can wander into the continuous spectrum and then disappear.

This behavior can also be described without the help of perturbation theory, as is seen below. The main idea is this: Consider an operator of the form $A = A^{(0)} + V$ in the Hilbert space H_0. Assume that, in a nested Hilbert space (H_0, H_1, E_{01}) the operator family $e_{10}((z - A^{(0)})^{-1})$—corresponding to the "noninteracting" $A^{(0)}$—can be analytically continued into some Riemann surface (as was illustrated by the examples of Sec. IV). Impose on the interaction V restrictions which ensure that the equation

$$(z - A)^{-1} = (z - A^{(0)})^{-1} + (z - A)^{-1}V(z - A^{(0)})^{-1} \tag{6.1}$$

—multiplied by E_{10} from the left and by E_{01} from the right—becomes, in H_1, an equation of the form

$$G_1(z) = G_1^{(0)}(z) + G_1(z)K(z), \tag{6.2}$$

where

$$G_1(z) = e_{10}((z - A)^{-1}), \tag{6.3}$$

$$G_1^{(0)}(z) = e_{10}((z - A^{(0)})^{-1}). \tag{6.4}$$

If (6.2) can be analytically continued beyond the domain of validity of (6.1), then it provides a means for studying the "unphysical" singularities of $G_1(z)$ in their dependence on V.

A related problem is the study of the operator family

$$T(z) = V + V(z - A)^{-1}V \tag{6.5}$$

in the Hilbert space H_0 (see Sec. VIII).

Another question discussed in this section is the

relation between the "improper" eigenvectors of $A^{(0)}$ and of A.

1. Consider in H_0 an operator of the form

$$A = A^{(0)} + V. \tag{6.6}$$

Assume that

(a) $A^{(0)}$ and A are closed;
(b) The domain of A is the same as the domain of $A^{(0)}$.

Then (6.1) holds[13] for every $z \in \rho(A) \cap \rho(A^{(0)})$.

Denote by Δ_0 a connected component of $\rho(A) \cap \rho(A^{(0)})$. Assume further

(c) The family $e_{10}((z - A^{(0)})^{-1})$ can be analytically continued from Δ_0 into a domain $\Delta_1 \supset \Delta_0$ (which may have several sheets).

The problem is to study, in Δ_1, the analytic continuation of $e_{10}((z - A)^{-1})$ and to examine the isolated singularities of that analytic continuation.

The main assumption is:

(d_1) For every $z \in \Delta_0$, the operator

$$K(z) = j_{10}(V(z - A^{(0)})^{-1}) \tag{6.7}$$

is bounded.

(d_2) The operator family $K(z)$ can be analytically continued to a family holomorphic in Δ_1.

If the interaction V is of the form

$$V = e_{01}(w) \tag{6.8}$$

where W is bounded, then (d_1) and (d_2) follow from (c), with

$$K(z) = We_{10}((z - A^{(0)})^{-1}). \tag{6.9}$$

Theorem: If the conditions (a), (b), (c), (d) are satisfied, then

$$G_1(\zeta) = G_1^{(0)}(\zeta)[1 - K(\zeta)]^{-1} \tag{6.10}$$

for every point $\zeta \in \Delta_1$ which can be reached from Δ_0 by a path along which $1 - K(\zeta)$ has a bounded inverse.

Proof: For $\zeta \in \Delta_0$, Eq. (6.2) is a consequence of (6.1) and of (d_1). The solution of (6.2) is (6.10). By (c) and (d_2), (6.10) can be analytically continued from Δ_0 into Δ_1, which proves the assertion.

The singularities of $G(\zeta)$ can occur, by (6.10) only at points where either $G_1^{(0)}(\zeta)$ or $[1 - K(\zeta)]^{-1}$ is singular.

2. Example: Let $(z - A^{(0)})^{-1}$ be the integral operator in $L^{(2)}(R^3)$ defined by (4.10). Let $M(x)$ be a continuous function over R^3 such that $0 < M(x) \leq 1$ and that, as $|x| \to \infty$, $M(x) = O(e^{-a|x|})$ for every $a > 0$. Let $V(x)$ be a function of the form

$$V(x) = (M(x))^2 W(x),$$

where $W(x)$ is measurable and bounded. Consider the nested Hilbert space of Example 1, Sec. II. Then, by Example (f) of Sec. IV, the operator $e_{10}((z - A^{(0)})^{-1})$ can be analytically continued into Δ_1—the Riemann surface of $z^{\frac{1}{2}}$. Furthermore, K is the integral operator defined in $H_1 = L^{(2)}(M^{-2})$ by the kernel

$$K(x, x'; \zeta) = W(x)M^2(x)G^{(0)}(x, x'; \zeta)$$
$$= V(x)G^{(0)}(x, x'; \zeta).$$

Since K is completely continuous, the family $1 - K(\zeta)$ has, in Δ_1, singularities if and only if the operator $K(\zeta)$ has an eigenvalue equal to one. Because of the unitary correspondence $f \to Mf$ between $L^{(2)}(M^{-2})$ and $L^{(2)}$, this happens whenever, in $H_0 = L^{(2)}$, the kernel

$$Q(x, x'; \zeta) = M^{-1}(x)K(x, x'; \zeta)M(x)$$
$$= M(x)W(x)G^{(0)}(x, x'; \zeta)M(x')$$

has an eigenvalue equal to one.

3. Remark: If V is replaced by wV (where w is a complex number), then K is replaced by wK. The expression the rhs of (6.10) can be studied by standard methods. For example, if $||K|| < 1$, then $(1 - K)^{-1}$ exists. If K is completely continuous, then the Fredholm alternative holds. If K is of trace class, then the singularities of $[1 - wK(\zeta)]^{-1}$ lie on the manifold—in the ζ–w space—obtained by setting to zero the Fredholm determinant.

4. We consider now the relation between the generalized eigenvalues of $A^{(0)}$ and of $A^{(w)} = A^{(0)} + wV$ (w complex).

Assume

(a) For every w, the operator $j_{10}(A^{(w)})$ is densely defined;
(b) The number z satisfies $z \in \rho(A^{(0)})$ and $z^* \in \rho(A^{(0)*})$;
(c) The operator $E_{01}^{-1}V^*G_0^{(0)*}(z)E_{01}$ is bounded. [Here $G_0^{(0)}(z) = (z - A^{(0)})^{-1}$.]

Define $L(z)$ as the closure of $E_{10}G_0^{(0)}(z)VE_{10}^{-1}$ and $\Omega(z, w)$ as

$$\Omega(z, w) = [1 - wL(z)]^{-1}.$$

A simple calculation shows then that

$$j_{10}^*(z^* - A^{(0)*}) \supseteq j_{10}^*(z^* - A^*)\Omega(z, w).$$

[13] See, e.g., Hille-Phillips, Ref. 11, p. 197.

Consequently: If z is a generalized eigenvalue of $A^{(0)}$, [with $j_{10}^*(z^* - A^{(0)*})f = 0$], and if $\Omega(z, w)f$ is in the domain of $j_{10}^*(A^{(w)*})$, then $j_{10}^*(z^* - A^{(w)*}) \cdot \Omega(z, w)f = 0$. The operators $\Omega(z, w)$ establish a correspondence between generalized eigenvectors of $A^{(0)}$ and of $A^{(w)}$.

VII. STATES IN QUANTUM MECHANICS

In the standard mathematical formulation[1] of quantum mechanics, the states are unit vectors in a Hilbert space H. Every self-adjoint operator in H corresponds to an observable and every projection operator in H to a question.[14]

One might ask whether the concept of state may be modified so as to accomodate wavefunctions that are not square-integrable (such as eigenfunctions of the continuous spectrum). A way of doing this is described in this section.

It turns out, not surprisingly, that the generalized states allow only the computation of relative probabilities,[1] and that they allow only a subset of questions.

Physically, the last condition implies that every allowed complete measurement leaves the system in a state, the wavefunction of which is not only square-integrable, but also "well behaved". Such a requirement seems reasonable.

1. Definition: Let (H_0, H_1, E_{01}) be a nested Hilbert space.

An *admissible question* is defined as an orthogonal projection operator in H_0 which is of the form

$$P = e_{01}(\Pi) = E_{01}\Pi E_{10}$$

where Π is a bounded operator in H_1.

A *state* is defined as an arbitrary nonzero vector in H_1. Nonzero multiples of a vector $f \in H_1$ define one and the same state.

A *normalizable state* is a nonzero element of the subset $E_{10}H_0 \subset H_1$. To every normalizable state $f = E_{10}h$ $(h \in H_0)$, there corresponds the unit vector $h = ||h||_0^{-1}h$ in H_0.

Notice that an admissible question is an operator, not in the space H_1 of states, but in the space H_0.

The *wavefunction* of a state $f \in H_1$ is defined as the linear functional which, to every $\varphi \in H_1$ assigns the number $(f, \varphi)_1$.

If the Fourier transformation F is a unitary operator in H_1, then it is convenient to consider also the functionals $\varphi \to (f, F^j\varphi)_1$ $(j = 1, 2, 3)$. They define the wavefunction in the p, $-x$, and $-p$ representations.

A density operator ρ in H_1 is defined by $\rho \neq 0$ and by the condition $0 \leq \text{tr } (\Pi\rho) < \infty$ whenever $e_{01}(\Pi)$ is an admissible question. If E_{01} is of Hilbert–Schmidt class, then every positive-definite, bounded, self-adjoint operator is a density operator.

If P' and P'' are admissible questions that commute (i.e., if $P'P'' = P''P'$), then $P' + P''$ and $P'P''$ are also admissible questions.

The unit operator is, in general, *not* an admissible question, since there need not exist any bounded operator Π in H_1 such that $1 = e_{01}(\Pi)$.

If P is an admissible question, then $1 - P$ is, in general, not an admissible question.

2. The statistical interpretation is given by the rules:

(a) If a system is in the state $f \in H_1$, then the *relative* probability of positive answer to the questions $P' = e_{01}(\Pi')$ and $P'' = e_{01}(\Pi'')$ is the ratio

$$(f, \Pi'f)_1/(f, \Pi''f)_1.$$

(b) If the range of P is one-dimensional, and if the positive answer to the question P is observed, then the system is left in the normalizable state $E_{10}PH_0$.

An admissible question P of one-dimensional range is an operator of the form $Pg = (h, g)_0h$, where $h \in E_{01}H_1$ and $||h||_0 = 1$. By (b), the complete measurement P leaves the system in the state $E_{10}h = E_{10}E_{01}f$, where $f \in H_1$. The wavefunction of this state is $\varphi \to (E_{10}E_{01}f, \varphi)_1 = (E_{01}f, E_{01}\varphi)_0$.

Consider a system in the state of $f \in H_1$. Let $\varphi' \in H_1$ and $\varphi'' \in H_1$ be such that $||E_{01}\varphi'||_0 = ||E_{01}\varphi''||_0 = 1$. Then the relative probability of finding the system in the normalizable states $E_{10}E_{01}\varphi'$ and $E_{10}E_{01}\varphi''$ is the ratio $|(f, \varphi')_1|^2/(f, \varphi'')_1|^2$. In the special case that $f = E_{10}h$ $(||h||_0 = 1)$ is a normalized state, the scalar product $(f, \varphi')_1$ becomes $(f, \varphi')_1 = (E_{10}h, \varphi')_1 = (h, E_{01}\varphi')_0$ in agreement with the usual interpretation.

3. Examples: Assume that $H_0 = L^{(2)}$, and that $(E_{01}\varphi)(x) = \varphi(x)$ for every $\varphi \in H_1$.

The wavefunction of a normalizable state $f = E_{10}\psi$, $(\psi \in L^{(2)})$ is the functional

$$(f, \varphi)_1 = (E_{10}\psi, \varphi)_1 = (\psi, E_{01}\varphi)_0 = \int \psi^*(x)\varphi(x)\,dx$$

for every $\varphi \in H_1$. This functional can be identified with the function $\psi(x)$. So the definition of wavefunction, given above, coincides, for normalizable states, with the usual definition.

A function $\psi(x)$ may be the wavefunction of some

[14] G. W. Mackey, Lecture Notes on the Mathematical Foundations of Quantum Mechanics (Harvard University, Cambridge, Massachusetts, 1960).

state $f \in H_1$ even if $\psi(x)$ is not square-integrable. If H_1 is the Hilbert space of Example 3a, Sec. II, then every function $\psi(x)$ such that $\exp(-\frac{1}{2}x^2)\psi(x) \in L^{(2)}$ is the wavefunction of some state $f \in H_1$.

For arbitrary complex w, the wavefunction of the state $\delta_w \in H_1$ [see (1.5)] is the functional $\varphi \to \varphi(w)$. In the p representation, the wavefunction of the same state is $\exp(-iw^*p)$.

Consider the same nested Hilbert space as above. Let I be a finite nonempty interval of the x axis. Let χ_I be the characteristic function of I, normalized to unity ($||\chi_I||_0 = 1$). The projection operator defined by the kernel $\chi_I(x)\,\chi_I(x')$ is *not* an admissible question, since χ_I does not belong to H_1.

The approximations of χ_I by partial sums

$$\chi_{IN}(x) = \sum_{n=0}^{N} c_n h_n(x)$$

[$N = 0, 1, 2, \cdots$, $c_n = (h_n, \chi_I)_0$; $h_n(x)$ Hermite functions] do belong to H_1. Define

$$q_{IN}(x) = \left(\int |\chi_{IN}(x')|^2\, dx'\right)^{-\frac{1}{2}} \chi_{IN}(x).$$

The projection operator in H_0, defined by the kernel $q_{IN}(x)q_{IN}(x')$, is an admissible question. If the "yes" answer to this question is observed, the system is left in a state with wavefunction $q_{IN}(x)$ which tends, as $N \to \infty$, to a wave packet concentrated in I.

It is an elementary, but instructive, exercise to study the relative probabilities of positive answers to these questions, for a system in the state δ_w (w arbitrary complex), both in the x and the p representation. The sums can be evaluated with the help of the Christoffel summation formula.

VIII. VARIOUS APPLICATIONS

This section contains the outline of two questions of nonrelativistic quantum mechanics where the language of nested Hilbert spaces seems convenient. No attempt will be made here to construct a general theory on the basis of the definitions of the preceding section.

1. Resonances: Let A be an operator in a Hilbert space H_0. Assume that A is self-adjoint and bounded from below. If A is the Hamiltonian of a system, then

$$U(t) = \exp(-iAt)$$

is the time development operator. Since $U(t)$ is bounded, it is completely determined by its matrix elements between elements of any dense subset of H_0. Consequently, if H_1 is any Hilbert space and E_{01} any inclusion of H_1 into H_0, then the operator $U(t)$ is completely determined by the operator $e_{10}(U(t))$.

It is often possible to write $U(t)$ as a contour integral

$$U(t) = (2i\pi)^{-1} \oint e^{-izt}(z - A)^{-1}\, dz \qquad (8.1)$$

taken around the spectrum of A. The conditions for the validity of (8.1) will not be discussed here. It follows from (8.1) that

$$e_{10}(U(t)) = (2i\pi)^{-1} \oint e^{-izt} G_1(z)\, dz, \qquad (8.2)$$

where $G_1(z) = e_{10}((z - A)^{-1})$ is the operator family discussed in Secs. III–VI; it is analytic in a domain which will, in general, have several sheets.

If the path of integration in (8.2) can be shrunk to a point while crossing only isolated singularities, $\zeta_1, \zeta_2, \zeta_i \cdots$, then

$$e_{10}(U(t)) = \sum_i e^{-iz_i t} R_i, \qquad (8.3)$$

where R_i is the residue of G_1 at ζ_i, and z_i is the ground point of ζ_i. Equation (8.3) is the decomposition of $e_{10}(U(t))$ into contributions from individual points of the discrete spectrum with respect to the given nested Hilbert space.

2. The operator $T(z)$: If $A = A^{(0)} + V$ is again the Hamiltonian, then the operator family

$$T(z) = V + V(z - A)^{-1}V \qquad (8.4)$$

is related to the scattering amplitude. The fact that $T(z)$ is "better" than $(z - A)^{-1}$ —provided V is suitably restricted—has been often exploited.[15] The reason for this fact is the obvious proposition below:

Proposition: If V is of the form $e_{01}(W)$ where W is a bounded operator in H_1, then

$$T(z) - V = e_{01}(WG_1(z)W) = E_{01}WG_1(z)WE_{10};$$

here $G_1(z) = e_{10}((z - A)^{-1}) = E_{10}(z - A)^{-1}E_{01}$.

This means not only that $T(z)$ can be analytically continued as far as $G_1(z)$, but also that the matrix elements of $T(z)$ can be extended to a space "larger" than H_0, with the help of the procedure used, e.g., in Sec. VII.

A formulation of scattering theory in the language of nested Hilbert spaces will be given in another paper of this series.

ACKNOWLEDGMENTS

Part of this work was done in the summer of 1962 while the author was at New York University. He

[15] See, e.g., A. Grossmann and T. T. Wu, J. Math. Phys. **2**, 710; **2**, 714; and **3**, 684 (1962).

would like to thank Professor B. Zumino for his hospitality. He is indebted to Professor J. R. Oppenheimer for the opportunity to spend the year 1962–63 at the Institute for Advanced Study, Princeton, New Jersey.

APPENDIX

This appendix contains details about the nested Hilbert space of Example 3a, Sec. II.

The positive numbers λ_n were assumed to satisfy

$$\lambda_n^2 \leq (n!)^{-1}$$

for $n = 0, 1, 2, \cdots$. The (orthonormal) Hermite functions were denoted by $h_n(x)$.

Proposition: For every complex z, the series with positive terms

$$r(z) = \sum_{n=0}^{\infty} \lambda_n^2 \, |h_n(z)|^2 \tag{A1}$$

is convergent. For every $b > 1$ there exists a constant $C = C(b)$ such that, for every $z = x + iy$,

$$r(z) \leq C \exp(-b^{-1}x^2 + by^2). \tag{A2}$$

Proof: The Hermite polynomial $H_n(z)$ can be written as a contour integral around the origin:

$$H_n(z) = (2i\pi)^{-1}\Gamma(n+1) \oint u^{-n-1} \exp(2zu - u^2)\, du.$$

The assertion follows by a straightforward estimate of $h_n(z) = \pi^{-\frac{1}{4}}(n!)^{-\frac{1}{2}} \exp(-\frac{1}{2}z^2)H_n(z)$.

An immediate consequence is

Proposition: Let f_n be a sequence of complex numbers such that

$$\sum_{n=0}^{\infty} \lambda_n^{-2} \, |f_n|^2 < \infty. \tag{A3}$$

Then the series $\sum_{n=0}^{\infty} f_n h_n(z)$ is absolutely convergent for every z. The convergence is uniform in every compact of the z plane.

Proof: If n and q are any positive integers, then

$$\sum_{m=n}^{n+q} |f_m| \, |h_m(z)| = \sum_{m=n}^{n+q} |f_m| \, \lambda_m^{-1}\lambda_m \, |h_m(z)| \leq \left(\sum_{m=n}^{n+q} |f_m|^2 \, \lambda_m^{-2}\right)^{\frac{1}{2}} (r(z))^{\frac{1}{2}},$$

where $r(z)$ is defined by (A1). The first factor on the rhs of this inequality can be made arbitrarily small by the choice of sufficiently large n. The second factor is bounded in every compact of the z plane.

Definition of H_1 and inclusion of H_1 into $L^{(2)}$: Consider the set H_1 of all the functions

$$f(x) = \sum_{n=0}^{\infty} f_n h_n(x),$$

such that the coefficients f_n satisfy (A3). For $f \in H_1$, $g \in H_1$,

$$g(x) = \sum_{n=0}^{\infty} g_n h_n(x)$$

define a scalar product

$$(f, g)_1 = \sum_{n=0}^{\infty} f_n^* \lambda_n^{-2} g_n.$$

In particular,

$$(h_m, h_n)_1 = \lambda_n^{-2}\delta_{mn}.$$

It is easy to verify that H_1 is complete with respect to the scalar product, and that $(f, f)_1 = 0$ implies $f = 0$. So H_1 is a Hilbert space.

It is convenient to consider in H_1 the orthonormal basis e_n, the elements of which are defined by

$$e_n(x) = \lambda_n h_n(x).$$

For any $f \in H_1$, define $E_{01}f \in L^{(2)}$ by

$$(E_{01}f)(x) = f(x).$$

In order to verify that E_{01} is continuous, notice that

$$||E_{01}f||_0^2 = \sum_{n=0}^{\infty} |f_n|^2 < (\max \lambda_n^2) \sum_{n=0}^{\infty} \lambda_n^{-2} \, |f_n|^2 = (\max \lambda_n^2) \, ||f||_1^2.$$

The range of E_{01} is dense in $L^{(2)}$, since it contains all finite linear combinations of Hermite functions. Finally, $E_{01}f = 0$ implies $f_n = 0$ for all n, and $f = 0$.

So E_{01} is an inclusion in the sense of Sec. II.

The inclusion E_{01} can also be written as

$$E_{01} = \sum_{n=0}^{\infty} |h_n >_0 \lambda_n <_1 e_n|,$$

which means that, for every $f \in H_1$,

$$(E_{01}f)(x) = \sum_{n=0}^{\infty} h_n(x)\lambda_n(e_n, f)_1 = \sum_{n=0}^{\infty} h_n(x)\lambda_n^2(h_n, f)_1 = \sum_{n=0}^{\infty} h_n(x) f_n = \sum_{n=0}^{\infty} h_n(x)(f, h_n)_0.$$

This is just the expansion of f into a Hermite series.

The adjoint of E_{01} is, with $e_n(x) \equiv \lambda_n h_n(x)$,

$$E_{10} = \sum_{n=0}^{\infty} |e_n >_1 \lambda_n <_0 h_n|.$$

To every $g \in H_0 = L^{(2)}$, it associates $E_{10}g \in H_1$,

defined by

$$(E_{10}g)(x) = \sum_{n=0}^{\infty} e_n(x)\lambda_n g_n = \sum_{n=0}^{\infty} h_n(x)\lambda_n^2 g_n$$

$$= \sum_{n=0}^{\infty} h_n(x)\lambda_n^2 \int h_n(y)g(y)\,dy = \int K(x, y)g(y)\,dy,$$

where

$$K(x, y) = \sum_{n=0}^{\infty} h_n(x)\lambda_n^2 h_n(y).$$

Basic properties of elements of H_1: Because of the assumptions on λ_n, the elements of H_1 are functions with many regularity properties.

Theorem: Every $f \in H_1$ is entire analytic. If b is any real number such that $b > 1$, then there exists a constant $C = C(b)$ such that, for every $z = x + iy$,

$$|f(x + iy)| \leq C\,||f||_1 \exp(-\tfrac{1}{2}b^{-1}x^2 + \tfrac{1}{2}by^2). \quad \text{(A4)}$$

Proof: The function $f(z)$ is the sum of a series of entire functions which converges uniformly in every compact; consequently $f(z)$ is entire. Furthermore,

$$f(z)| \leq \sum^{\infty} |f_n|\,|h_n(z)| \leq ||f||_1\,(r(z))^{\frac{1}{2}}$$

$$\leq C\,||f||_1 \exp(-\tfrac{1}{2}x^2 + ay^2).$$

Q.E.D.

The above theorem says that every $f \in H_1$ decreases fast on parallels to the real axis, and that in the whole complex plane it increases at a rate not faster than $\exp(a|z|^2)$ $(a > \frac{1}{2})$.

An easy consequence of the definition of H_1 is

Proposition: If $f \in H_1$ then its Fourier transform Ff, also belongs to H_1 and $||Ff||_1 = ||f||_1$.

Proof: Notice that

$$Fe_n = i^{-n}e_n.$$

Write

$$f = \sum_{n=0}^{\infty} \varphi_n e_n,$$

where

$$\varphi_n = \lambda_n^{-1} f_n \quad \text{and} \quad \sum_{n=0}^{\infty} |\varphi_n|^2 < \infty.$$

Define

$$f^{(j)} = \sum_{k=0}^{\infty} \varphi_{4k+j} e_{4k+j} \qquad (j = 0, 1, 2, 3).$$

Then

$$f = \sum_{j=0}^{3} f^{(j)}, \qquad Ff^{(j)} = i^{-j}f^{(j)},$$

and, for $j_1 \neq j_2$ $(j_1, j_2 = 0, 1, 2, 3)$,

$$(f^{(j_1)}, f^{(j_2)})_1 = 0.$$

So

$$||Ff||_1^2 = \sum_{j=0}^{3} ||Ff^{(j)}||_1^2 = \sum_{j=0}^{3} ||f^{(j)}||^2 = ||f||_1^2.$$

Q.E.D.

Since clearly $FH_1 = F^{-1}H_1 = H_1$, the Fourier transformation is a unitary operator in H_1. If A_1 is any bounded operator in H_1, its Fourier transform will be defined as

$$\tilde{A}_1 = FA_1F^{-1} = FA_1F^*,$$

so that $(Fg, \tilde{A}_1Ff)_1 = (g, A_1f)_1$, for all $f, g \in H_1$.

The elements δ_w: There exists a canonical one-to-one correspondence between elements of a Hilbert space and continuous linear functionals in that space. We now study the elements of H_1 that define functionals which are "δ functions with complex argument".

An immediate consequence of (A4) is

Proposition: Let w be any complex number. Then the correspondence which to every $f \in H_1$ associates the number $f(w)$, is a continuous linear functional in H_1.

Consequently there exists in H_1 a unique vector (to be denoted by δ_w), such that, for every $f \in H_1$,

$$(\delta_w, f)_1 = f(w). \quad \text{(A5)}$$

A family of vectors $g_w \in H_1$ (H_1 a Hilbert space) is said to be holomorphic in an open set Δ of the w plane, if, for every $f \in H_1$, the complex-valued function $(f, g_w)_1$ is holomorphic in Δ.

Then, because of the analyticity of every $f \in H_1$,

Proposition: The family δ_w of elements of H_1 is entire analytic in w^* (the complex conjugate of w).

Next, the function $\delta_w(z)$ will be expanded in the orthonormal basis $\{e_n\} = \{\lambda_n h_n\}$ of H_1. Write

$$\delta_w = \sum_{m=0}^{\infty} d_m e_m;$$

then

$$(\delta_w, e_n)_1 = \sum_{m=0}^{\infty} d_m^*(e_m, e_n)_1 = d_n^* = e_n(w),$$

so that

$$\delta_w(z) = \sum_{m=0}^{\infty} e_m^*(w)e_m(z) = \sum_{m=0}^{\infty} \lambda_m^2 h_m^*(w)h_m(z).$$

Since $\delta_w \in H_1$, this series is strongly convergent in H_1, and consequently uniformly convergent in compacts of the z plane [see (A4)].

For any two complex w and u, the scalar product, in H_1, of δ_u and δ_w is

$$(\delta_u, \delta_w)_1 = \sum_{m=0}^{\infty} e_m(u)e_m^*(w) = \sum_{m=0}^{\infty} \lambda_m^2 h_m(u)h_m^*(w).$$

The Fourier transform of δ_w is

$$(F\delta_w)(z) = \sum_{m=0}^{\infty} e_m^*(w)i^{-m}e_m(z). \tag{A6}$$

Since the Fourier transform is a unitary operator in H_1, we have for every $f \in H_1$

$$(F\delta_w, f)_1 = (\delta_w, F^{-1}f)_1 = (2\pi)^{-\frac{1}{2}} \int e^{iwx} f(x)\, dx.$$

This can also be directly verified, with the help of (A6).

The operators $\delta(u, w)$: It is often convenient to consider the family of bounded operators in H_1 defined, in the Dirac notation, by

$$\delta(u, w) = |\delta_u\rangle\langle\delta_w|.$$

That is: for every $f \in H_1$,

$$\delta(u, w)f = \delta_u(\delta_w, f)_1,$$

and

$$(g, \delta(u, w)f)_1 = g^*(u)f(w).$$

The family $\delta(u, w)$ is entire analytic in u^* and in w. In particular, the family $\delta(w^*, w)$ is entire analytic in w.

For every u, and w, the equality

$$\delta^*(u, w) = \delta(w, u)$$

holds. In particular, $\delta(w, w)$ is self-adjoint for every complex w.

For every u and w, $\delta(u, w)$ is of trace class, and $\operatorname{tr}(\delta(u, w)) \leq [r(u)r(w)]^{\frac{1}{2}}$ where $r(z)$ is defined by (A1). Indeed,

$$(e_m, \delta(u, w)e_n)_1 = e_m^*(u)e_n(w),$$

and

$$\sum_{m=0}^{\infty} |e_m(u)e_m(w)| \leq \left[\sum_{m=0}^{\infty} |e_m(u)|^2\right]^{\frac{1}{2}} \left[\sum_{n=0}^{\infty} |e_n(w)|^2\right]^{\frac{1}{2}}$$
$$= [r(u)r(w)]^{\frac{1}{2}} < \infty.$$

The Fourier transform of $\delta(u, w)$ is

$$\tilde{\delta}(u, w) = F\delta(u, w)F^{-1}$$

so that, for all $f, g \in H_1$,

$$(f, \tilde{\delta}(u, w)g)_1$$
$$= (F^{-1}f, \delta(u, w)F^{-1}g)_1 = (2\pi)^{-1} \iint f^*(x')$$
$$\times \exp(-iw^*x' + iux'')g(x'')\, dx'\, dx''.$$

For every complex w, the self-adjoint operator $\delta(w, w)$ is positive-semidefinite, since $(f, \delta(w, w)f)_1 = |f(w)|^2 \geq 0$ for every $f \in H_1$.

Finite and Disconnected Subgroups of SU_3 and their Application to the Elementary-Particle Spectrum*

W. M. FAIRBAIRN†, T. FULTON, AND W. H. KLINK
The Johns Hopkins University, Baltimore, Maryland
(Received 15 November 1963)

An attempt is made to fit the symmetries of the currently observed elementary-particle spectrum into the structure of finite or disconnected subgroups of SU_3. Surprisingly, the detailed properties of these subgroups have not been elucidated previously. As a first step, therefore, character tables and other relevant properties are derived for these groups. Next, the classification of elementary particles is made on the basis of the representations of the groups discussed. The techniques previously employed by Case, Karplus, and Yang for the application of finite subgroups of SU_2 to isotopic spin are extended to the subgroups of SU_3. The structure of SU_3 is utilized to suggest how charge and hypercharge operators are to be assigned in the subgroups. The results obtained are similar to those of SU_2 and isotopic spin. There is an upper limit, for any given group, to the dimension of the irreducible representation. For some of the groups considered, these upper limits are eight and even ten. There exist finite groups which can accommodate the eight baryons in one of the irreducible representations. However, when one looks at scattering problems, use of the finite groups, as expected, gives charge or hypercharge conservation only modulo an integer determined by the group. Charge independence is also lost. In a representative group analyzed in detail, the imposition of exact charge conservation leads automatically to the exact conservation of hypercharge and to the full SU_3 symmetry. Exact charge and hypercharge conservation can be maintained for the disconnected groups, but the maximum dimension of the irreducible representations is six, and only charge symmetry, not charge independence, is satisfied. A short discussion of the representations of the group SU_3/C is included in the appendix.

I. INTRODUCTION

OF all the groups proposed thus far to account for the elementary-particle spectrum,[1] the three-dimensional unitary unimodular group SU_3, and, in particular, the version founded on the eight-dimensional representation[2,3] seems to have been the most successful. The groups considered in attempts to account for the elementary-particle multiplicity have been principally compact, semisimple Lie groups of rank two. Two characteristics of these groups which are of significance in elementary-particle physics are that there are no limits to the dimensions of the irreducible representations and that exact conservation laws are associated with the physical quantities (such as Q and Y) related to operators of the group.

Probably one of the principal reasons that only Lie groups (or trivial extensions of them[4]) have been considered up to now is that their mathematical properties are well known. One may well ask whether there may be other groups, characterized by different symmetries, which can accommodate the elementary particles.[5] In view of the success of the models based on SU_3, we have restricted ourselves to its subgroups, both finite and infinite. One obvious infinite subgroup of SU_3 is SU_2, which, however, is not acceptable in the present context, since it has only one independent quantum number associated with it (i.e., it is a rank-one group). Our interest is in subgroups in which there exist operators corresponding to two quantum numbers (e.g., Q, Y), that is, groups which have a structure similar to that of a rank-two group.

The finite subgroups of SU_3 with which we are concerned in this paper are listed by Miller, Blichfeld, and Dickson[6]; the disconnected subgroups are given by the same authors,[6] and by the preprint version of Speiser and Tarski's paper.[5] There has been no concern with these groups comparable to that with the crystal groups. All the detailed discussions of them of which we are aware[6-8] are not very recent and often give only the order and generators of the group. They are not concerned with group-theoretical properties such as classes, character tables, and irreducible representations. Section II of this paper is devoted to obtaining these

* Supported in part by the National Science Foundation.
† Visiting Professor, 1962–63. Permanent address: The College of Science and Technology, Manchester, England.
[1] R. E. Behrends, J. Dreitlein, C. Fronsdal, and B. W. Lee, Rev. Mod. Phys. **34**, 1 (1962).
[2] M. Gell-Mann, "The Eightfold Way," CTSL-20 (1961); Phys. Rev. **125**, 1067 (1962).
[3] Y. Ne'eman, Nucl. Phys. **26**, 222 (1961).
[4] See for example, T. D. Lee and C. N. Yang, Phys. Rev. **122**, 1954 (1961).
[5] D. R. Speiser and J. Tarski, J. Math. Phys. **4**, 588 (1963) (preprint version).
[6] G. A. Miller, H. F. Dickson, and L. E. Blichfeldt, *Theory and Applications of Finite Groups* (G. E. Stechert and Company, New York, 1938), Chap. XII.
[7] C. Jordan, J. Reine Angew. Math. **84**, 93 (1878).
[8] C. Jordan, Atti Reale Accad. Napoli, **8**, No. 11 (1879).

properties, which are missing from the older discussions of the subgroups of SU_3.

The finite subgroups of SU_3 are of two types. The first type are analogs of the crystal groups and their double groups.[9] They are groups of order 36φ, 72φ, 216φ, 60, 168, and 360φ, where $\varphi = 1$ or 3, depending on whether the groups are analogs of the crystal groups or their double groups, respectively. The second type of finite groups are the analogs of the dihedral groups, and are of order $3n^2$ and $6n^2$, where n is an integer.

The infinite subgroups of SU_3, other than SU_2 or direct products of Abelian phase groups, are the analogs of the infinite dihedral group. These groups can be constructed by letting n approach infinity in the "dihedral-like" groups. The eight continuous parameters of SU_3 are then reduced to two continuous parameters. The infinite "dihedral-like" groups also contain a number of discrete elements other than simple reflections. Therefore, unlike Lie groups physicists generally deal with, which are connected groups, these groups are disconnected.

In Sec. III we turn to the application of the groups discussed in Sec. II to the spectrum of elementary particles. Only those groups will be considered suitable for such application for which it is possible to assign operators which satisfy the following minimum physical requirements:

(1) It should be possible to assign two independent quantum numbers (e.g., T_3, Y) within a multiplet. That is, the group should have a structure similar to that of a rank-two group.

(2) Charge should be exactly conserved in all reactions. However, we are willing to relax exact hypercharge conservation and allow hypercharge conservation modulo n ($n \geq 3$). The nonconservation of hypercharge, coupled with the conservation of charge, implies nonconservation of T_3, due to the Gell-Mann–Nishijima relation, and therefore lack of charge independence. We would like to allow for such a violation of charge independence only in situations where insufficient direct evidence for charge independence exists. We would therefore minimally require that

(3) the isotopic submultiplet structure and charge independence should exist for $T \leqq \frac{3}{2}$.

Beyond these minimum conditions it is convenient in view of the success of the "eightfold way" to require also that the groups considered should possess eight-dimensional irreducible representations to which the baryon and meson octets could be assigned. Finally, for mathematical convenience we would like to demand that all of the infinite subgroups be compact, so that their irreducible representations be finite-dimensional and equivalent to unitary ones.

It turns out that none of the groups considered in Sec. II satisfy all of the requirements listed above. In the case of the finite subgroups, although eight- and ten-dimensional representations exist and operators corresponding to charge and hypercharge can be defined, both of the conditions (2) and (3) listed above are violated. That is, if we consider scattering processes, the isotopic submultiplet structure is lost, even for low isotopic spins, and charge is conserved only modulo an integer determined by the group. Case, Karplus, and Yang[10] encounter the same difficulty with charge conservation, in their application of finite subgroups of SU_2 to isotopic spin. We can of course follow Ref. 10 and make charge conservation a separately imposed postulate. This postulate, however, represents an additional symmetry condition, and therefore serves to generate a new group. Charge conservation, when imposed as an additional condition on the subgroups of SU_2, generates the full SU_2 group.[10] Since two quantum numbers are defined for the subgroups of SU_3, one may be led to expect that the imposition of charge conservation is less restrictive than for subgroups of SU_2 and does not lead back to the full SU_3 group. For example, one may hope that, although charge may be made to be conserved exactly in all reactions, hypercharge could still be conserved only modulo n. We have found no groups realizing this possibility; perhaps there are group-theoretical reasons why they cannot exist.

Section III A is devoted to a review and expansion of the treatment of crystal groups as isotopic spin groups in Ref. 10. We discuss in detail how the imposition of exact charge conservation in particle reactions generates the full SU_2 symmetry. In Sec. III B a similar analysis is carried out for the analogs in SU_3 of the crystal groups. In particular, we consider extensively the group of order 216. The postulation of exact charge conservation leads to the exact conservation of hypercharge and to the full SU_3 symmetry.

The physical applications of the analogs of the dihedral groups are unsuccessful. An immediate difficulty is that the maximum dimension of the irreducible representations for both infinite and finite

[9] M. Hamermesh, *Group Theory* (Addison Wesley Publishing Company, Reading, Massachusetts, 1962), Chap. 9, Sec. 7.

[10] K. M. Case, R. Karplus, and C. N. Yang, Phys. Rev. 101, 874 (1956).

groups is six. Thus it is not possible to construct a particle scheme like that of the "eightfold way"; only models like Sakata's[11] are possible. If we pass over this difficulty, we still run into the same problem with charge conservation for the finite groups as we did previously. Exact charge and hypercharge conservation are automatically satisfied in the infinite groups. However, the models possess only charge symmetry, rather than the desired minimal charge independence. These points are illustrated by a specific example considered in Sec. III C.

Brief reference has occasionally been made[12,13] to the special role which SU_3/C plays in the "eightfold way" scheme. In the Appendix we discuss this group in somewhat more detail and derive an expression for the irreducible representations which is identical with the empirical rule given by Sakurai.[14] The discussion in the Appendix is germane to our previous considerations, since it leads us to consider only finite groups of order $n\varphi$ with $\varphi = 1$ (i.e., where the center has been factored out).

II. ANALYSIS OF THE SUBGROUPS

We propose to restrict ourselves in what follows to subgroups of SU_3 other than SU_2, its subgroups, and direct products of Abelian phase groups. The properties of these excluded groups are well known. They either do not possess a structure similar to that of a rank-two group demanded by our subsequent physical applications, as mentioned in the introduction, or else are trivial groups. The remaining finite subgroups of SU_3 are, on the one hand, the crystal-like groups and their double groups. We will denote these groups by the symbol $\Sigma(n\varphi)$, where n is the order of the group. The groups are $\Sigma(60)$, $\Sigma(168)$, $\Sigma(360\varphi)$, $\Sigma(36\varphi)$, $\Sigma(72\varphi)$ and $\Sigma(216\varphi)$, with $\varphi = 1$ or 3. On the other hand, there are the analogs of the dihedral groups. These groups will be denoted by the symbol $\Delta(3n^2)$ and $\Delta(6n^2)$, where n is an integer, and once again, the number in parentheses is the order of the group. The disconnected groups are $\Delta(3\infty^2)$ and $\Delta(6\infty^2)$, and can be obtained from the corresponding Δ groups of finite order by letting n approach infinity.

The symbol φ is introduced in labeling the order of the groups to distinguish subgroups of $SU_3(\varphi = 3)$ from subgroups of $FL_3(\varphi = 1)$—the fractional linear group of SU_3, isomorphic to SU_3/C. The groups $\Sigma(60)$ and $\Sigma(168)$ are subgroups of both. In the Δ groups the φ labeling is not a convenient one, although here too, one can consider subgroups of SU_3, FL_3, and of both. These properties will be elucidated in detail in Sec. II C.

We will investigate only those Σ groups with $\varphi = 1$, since, according to the discussion in the appendix, the group SU_3/C is more relevant to particle-symmetry applications than the full SU_3 group.

The existence of the two groups SU_3 and SU_3/C is the analog of the corresponding existence of SU_2 and $SU_2/C \cong O_3$. The parameter analogous to φ in this latter case takes on the values 1 and 2, corresponding, respectively, to the two square roots of unity (I and $-I$). When a finite group has its center factored out (e.g., $\varphi = 1$ for Σ groups), all elements satisfy

$$(C_n)^n = I, \tag{II 1}$$

where I is the identity element. However, if the center is not divided out, then

$$(C_n)^n = \{C\}, \tag{II 2}$$

where $\{C\}$ is the center of the group. In particular, if $\varphi = 3$ in the Σ groups,

$$(C_n)^n = \{C\} = \{I, \omega I, \omega^2 I\}, \qquad \omega = e^{2\pi i/3}. \tag{II 3}$$

For example, the subgroup of SU_2 corresponding to tetrahedral symmetry is the double tetrahedral group[9] T', while the corresponding subgroup of O_3 is tetrahedral group T.

We now examine the various Σ and Δ groups in detail.

A. The Groups $\Sigma(60)$, $\Sigma(168)$ and $\Sigma(360)$

Any finite group can be considered as a subgroup of a permutation group. From this point of view $\Sigma(60)$ is the alternating group on five letters[6] A_5, a subgroup of the permutation group on five letters S_5. Its generators are listed in Table I, together with the generators of all the other groups we will consider subsequently. Using the criteria and relations given by Littlewood,[15] and the character table of S_5 given by him,[16] one can construct Table II. This table is identical to the character table of the icosahedral group (I), and thus $\Sigma(60)$ is isomorphic to I and its character table is well known.[10] We list it only for the sake of completeness. The first row of Table II lists the permutation type of each class.

[11] S. Sakata, Progr. Theoret. Phys. (Kyoto) **16**, 686 (1956).
[12] L. Michel, "Invariance in Quantum Mechanics and Group Extensions," Lectures, The Istanbul Summer School, 1962 (to be published).
[13] A. Salam, *Proceedings of the International Conference on Nucleon Structure, 1963*, Stanford University Press, Palo Alto, California, 1963.
[14] J. J. Sakurai, Phys. Rev. Letters **10**, 446 (1963).
[15] D. E. Littlewood, *The Theory of Group Characters* (The Clarendon Press, Oxford, England, 1940), Chap. 9.
[16] Reference 15, Appendix.

TABLE I. Generators for the subgroups of SU_3.

Group	Elements			
$\Sigma(60)$	$A(0, \pi)$	$E(0, 0)$	W	
$\Sigma(168)$	$A(\frac{2}{7}\pi, \frac{4}{7}\pi)$	$E(0, 0)$	Z	
$\Sigma(360)$	$A(0, \pi)$	$E(0, 0)$	W	$B(\pi, \frac{5}{3}\pi)$
$\Sigma(36)$	$A(0, \frac{2}{3}\pi)$	$E(0, 0)$	V	
$\Sigma(72)$	$A(0, \frac{2}{3}\pi)$	$E(0, 0)$	V	$A(\frac{4}{9}\pi, \frac{4}{9}\pi)\, V A^{-1}(\frac{4}{9}\pi, \frac{4}{9}\pi)$
$\Sigma(216)$	$A(0, \frac{2}{3}\pi)$	$E(0, 0)$	V	$A(\frac{4}{9}\pi, \frac{4}{9}\pi)$
$\Delta(3n^2)$	$A\left(\frac{2\pi}{n} j, \frac{2\pi}{n} k\right)$	$E(0, 0)$		
$\Delta(6n^2)$	$A\left(\frac{2\pi}{n} j, \frac{2\pi}{n} k\right)$	$E(0, 0)$	$B\left(\frac{2\pi}{n} j, \frac{2\pi}{n} k\right)$	j, k integers
$\Delta(3\infty^2)$	$A(\alpha, \beta)$	$E(0, 0)$		
$\Delta(6\infty^2)$	$A(\alpha, \beta)$	$E(0, 0)$	$B(\alpha, \beta)$	

$$A(\alpha, \beta) = \begin{bmatrix} e^{i\alpha} & 0 & 0 \\ 0 & e^{i\beta} & 0 \\ 0 & 0 & e^{-i(\alpha+\beta)} \end{bmatrix}; \quad B(\alpha, \beta) = \begin{bmatrix} e^{i\alpha} & 0 & 0 \\ 0 & 0 & e^{i\beta} \\ 0 & e^{i(\pi-\alpha-\beta)} & 0 \end{bmatrix}; \quad E(\alpha, \beta) = \begin{bmatrix} 0 & e^{i\alpha} & 0 \\ 0 & 0 & e^{i\beta} \\ e^{-i(\alpha+\beta)} & 0 & 0 \end{bmatrix}$$

$$V = \frac{1}{3^{\frac{1}{2}}i} \begin{bmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega \end{bmatrix}; \quad A(\tfrac{4}{9}\pi, \tfrac{4}{9}\pi)\, V A^{-1}(\tfrac{4}{9}\pi, \tfrac{4}{9}\pi) = \frac{1}{3^{\frac{1}{2}}i} \begin{bmatrix} 1 & 1 & \omega^2 \\ 1 & \omega & \omega \\ \omega & 1 & \omega \end{bmatrix}$$

$$W = \tfrac{1}{2} \begin{bmatrix} -1 & \mu_2 & \mu_1 \\ \mu_2 & \mu_1 & -1 \\ \mu_1 & -1 & \mu_2 \end{bmatrix} \qquad Z = \frac{1}{7^{\frac{1}{2}}i} \begin{bmatrix} a & b & c \\ b & c & a \\ c & a & b \end{bmatrix}$$

$$\mu_1 = \tfrac{1}{2}(-1 + 5^{\frac{1}{2}}) \qquad a = \xi^4 - \xi^3$$

$$\mu_2 = \tfrac{1}{2}(-1 - 5^{\frac{1}{2}}) \qquad b = \xi^2 - \xi^5$$

$$\omega = e^{\frac{2}{3}\pi i} \qquad c = \xi - \xi^6$$

$$\xi^7 = 1$$

Although the order of the element in the class can immediately be obtained from this, we list it explicitly in the second row. The number of elements which commute with any given element is given by the ratio of the order of the group to the order of the class to which the element under consideration belongs. These numbers are listed in the fourth row. The irreducible representations of $\Sigma(n)$ are denoted by $\Sigma_d(n)$, where d is the dimension of the representation. When there is no danger of ambiguity, the (n) will be suppressed. In subsequent character tables, the notation $\bar{\Sigma}_d$ and $\bar{\Sigma}'_d$ will be used to denote complex-conjugate representations.

From the numbers listed in the fourth row of Table II, it is possible, on the basis of a self-consistency argument, to specify which elements commute. For $\Sigma(60)$ a given element C_3 or C_5 commutes with elements in its own period. A given element C_2 commutes with its period and its period multiplied by another C_2.

TABLE II. Character table for the group $\Sigma(60)$.

Permutation type	1^5	$1^2\,3$	12^2	5	5
Element type	E	(C_3, C_3^2)	C_2	(C_5, C_5^4)	(C_5^2, C_5^3)
Order of class	1	20	15	12	12
Number of commuting elements	60	3	4	5	5
Σ_1	1	1	1	1	1
Σ_3	3	0	-1	$\frac{1}{2}(1 + 5^{\frac{1}{2}})$	$\frac{1}{2}(1 - 5^{\frac{1}{2}})$
Σ'_3	3	0	-1	$\frac{1}{2}(1 - 5^{\frac{1}{2}})$	$\frac{1}{2}(1 + 5^{\frac{1}{2}})$
Σ_4	4	1	0	-1	-1
Σ_5	5	-1	1	0	0

TABLE III. Character table of the group $\Sigma(168)$.

Permutation type	1^7	$1^3\,2^2$	124	13^2	7	7
Element type	E	C_2	(C_4, C_4^3)	(C_3, C_3^2)	(C_7, C_7^2, C_7^4)	(C_7^3, C_7^5, C_7^6)
Order of class	1	21	42	56	24	24
Number of commuting elements	168	8	4	3	7	7
Σ_1	1	1	1	1	1	1
Σ_3	3	−1	1	0	$\frac{1}{2}(-1+i7^{\frac{1}{2}})$	$\frac{1}{2}(-1-i7^{\frac{1}{2}})$
Σ_3^*	3	−1	1	0	$\frac{1}{2}(-1-i7^{\frac{1}{2}})$	$\frac{1}{2}(-1+i7^{\frac{1}{2}})$
Σ_6	6	2	0	0	−1	−1
Σ_7	7	−1	−1	1	0	0
Σ_8	8	0	0	−1	1	1

TABLE IV. Character table for the group $\Sigma(360)$.

Permutation type	1^6	$1^3\,3$	$1^2\,2^2$	15	15	24	3^2
Element type	E	(C_3, C_3^2)	C_2	(C_5, C_5^4)	(C_5^2, C_5^3)	(C_4, C_4^3)	(C'_3, C'^2_3)
Order of class	1	40	45	72	72	90	40
Number of commuting elements	360	9	8	5	5	4	9
Σ_1	1	1	1	1	1	1	1
Σ_5	5	2	1	0	0	−1	−1
Σ'_5	5	−1	1	0	0	−1	2
Σ_8	8	−1	0	$\frac{1}{2}(1+5^{\frac{1}{2}})$	$\frac{1}{2}(1-5^{\frac{1}{2}})$	0	−1
Σ'_8	8	−1	0	$\frac{1}{2}(1-5^{\frac{1}{2}})$	$\frac{1}{2}(1+5^{\frac{1}{2}})$	0	−1
Σ_9	9	0	1	−1	−1	1	0
Σ_{10}	10	1	−2	0	0	0	1

$\Sigma(168)$ is a subgroup[6] of S_7. Its character table is given in Littlewood,[16] and is reproduced in Table III. The commuting elements for this group are as follows: elements C_3, C_4, and C_7 commute with elements in their own periods; C_2 commutes with the period of a corresponding C_4 element ($C_4^2 = C_2$), and with the period of this C_4 element multiplied by another C_2.

$\Sigma(360)$ is the alternating group on six letters[6] A_6, a subgroup of S_6. We have constructed Table IV from Littlewood's character table[16] for S_6. The commuting elements for this group are C_4 and C_5, both with their own periods, C_3 with its own period and its period multiplied by another C_3, and C_3^2 where C_3 is an element of the class $1^3 3$. Similarly, C'_3 commutes with its period and with another C'_3 and its period. C_2 has the same commutation properties as the C_2 elements of $\Sigma(168)$.

B. The Hessian Group, $\Sigma(216)$, and Its Subgroups $\Sigma(36)$ and $\Sigma(72)$

These groups all have the common characteristic that they contain invariant subgroups[6] of order 9 and 18. The group of order 18 is a dihedral group, and the one of order 9 is an Abelian group. Since these groups are subgroups of SU_2, we will not discuss them any further. One can also see by looking at the table of generators that $\Sigma(36)$ is a subgroup of $\Sigma(72)$ and of $\Sigma(216)$. To obtain character tables, we use the fact that the groups under consideration must be subgroups of some permutation group. Since the order of the permutation group on n letters is $n!$, and $216 = 2 \times 3 \times 4 \times 9$, we see that the lowest-order permutation group, which could have $\Sigma(216)$ as a subgroup, is S_9. Using Littlewood's techniques for finding compound characters,[15] we attempt to find a group of order 216 which is a subgroup of S_9 (the character table for S_9 is given by Littlewood[16]). Such a group of order 216 exists, since we can exhibit its character table explicitly. Furthermore, we can construct the character table for a subgroup of this group which is of order 72. This group has in turn a subgroup of order 36. Subsequent analysis of the three character tables obtained shows that the corresponding groups all have invariant subgroups of orders 9 and 18. Since the structure of both the invariant and noninvariant subgroups is the same as that required for $\Sigma(216)$, $\Sigma(72)$, and $\Sigma(36)$, we can identify the three subgroups of S_9 that we have obtained with the Σ groups of the corresponding order.

Table V gives the character table for the group

TABLE V. Character table of the group $\Sigma(36)$.

Permutation type	1^9	14^2	14^2	12^4	3^3	3^3
Element type	E	C_4	C_4^3	C_2	C_3	C_3^2
Order of class	1	9	9	9	4	4
Number of commuting elements	36	4	4	4	9	9
Σ_1	1	1	1	1	1	1
Σ'_1	1	−1	−1	1	1	1
$\bar{\Sigma}_1$	1	i	$-i$	−1	1	1
$\bar{\Sigma}_1^*$	1	$-i$	i	−1	1	1
Σ_4	4	0	0	0	−2	1
Σ'_4	4	0	0	0	1	−2

TABLE VI. Character table of the group $\Sigma(72)$.

Permutation type	1^9	14^2	14^2	14^2	12^4	3^3
Element type	E	(C_4, C_4^3)	(C'_4, C'^3_4)	(C''_4, C''^3_4)	C_2	(C_3, C_3^2)
Order of class	1	18	18	18	9	8
Number of commuting elements	72	4	4	4	8	9
Σ_1	1	1	1	1	1	1
Σ'_1	1	1	−1	−1	1	1
Σ''_1	1	−1	1	−1	1	1
Σ'''_1	1	−1	−1	1	1	1
Σ_2	2	0	0	0	−2	2
Σ_8	8	0	0	0	0	−1

TABLE VII. Character table for the group $\Sigma(216)$.

Permutation type	1^9	$1^3\,3^2$	$1^3\,3^2$	14^2	126	126	12^4	3^3	3^3	3^3
Element type	E	C_3	C_3^2	(C_4, C_4^3)	C_6	C_6^5	C_2	(C'_3, C'^2_3)	C''_3	C''^2_3
Order of class	1	12	12	54	36	36	9	8	24	24
Number of commuting elements	216	18	18	4	6	6	24	27	9	9
Σ_1	1	1	1	1	1	1	1	1	1	1
$\bar{\Sigma}_1$	1	ω	ω^2	1	ω	ω^2	1	1	ω	ω^2
$\bar{\Sigma}_1{}^*$	1	ω^2	ω	1	ω^2	ω	1	1	ω^2	ω
Σ_2	2	−1	−1	0	1	1	−2	2	−1	−1
$\bar{\Sigma}_2$	2	$-\omega$	$-\omega^2$	0	ω	ω^2	−2	2	$-\omega$	$-\omega^2$
$\bar{\Sigma}_2{}^*$	2	$-\omega^2$	$-\omega$	0	ω^2	ω	−2	2	$-\omega^2$	$-\omega$
Σ_3	3	0	0	−1	0	0	3	3	0	0
Σ_8	8	2	2	0	0	0	0	−1	−1	−1
$\bar{\Sigma}_8$	8	2ω	$2\omega^2$	0	0	0	0	−1	$-\omega$	$-\omega^2$
$\bar{\Sigma}_8{}^*$	8	$2\omega^2$	2ω	0	0	0	0	−1	$-\omega^2$	$-\omega$

$\Sigma(36)$. As we have just indicated, this group has invariant subgroups of order 9 and 18. The commuting elements are C_4 with its own period, C_2 with the period of the corresponding C_4, and C_3 with its own period, and its period multiplied by another C_3 and C_3^2.

Table VI gives the character table for the group $\Sigma(72)$. In addition to invariant subgroups of order 9 and 18, this group also has $\Sigma(36)$ as an invariant subgroup. The commuting elements for C_4 and C_3 are of the same type as in $\Sigma(36)$. C_2, in addition to commuting with its corresponding C_4 period, also commutes with the period of specific elements of C'_4 and C''_4, where $C_4^2 = (C'_4)^2 = (C''_4)^2 = C_2$.

Table VII gives the character table for the group $\Sigma(216)$. In addition to the invariant subgroups of order 9 and 18, this group also has $\Sigma(72)$ as an invariant subgroup. (This fact is not apparent from looking at the generators.) The invariant subgroup of order 9 generates a homomorphism of $\Sigma(216)$ onto the double tetrahedral group, T'.

The discussion of the commuting elements of $\Sigma(216)$ is very much more complicated than for the other Σ groups. The following types of self-consistent elements can be conjectured: Individual elements C_4 and C_6 commute with their own periods. From the orders of the classes we can see that three C_6 elements commute with each C_3 element; four C_6 and three C_4 commute with each C_2. When all possible periods are taken into account, all elements commuting with C_2 are exhausted. Twelve of the elements commuting with a given C_3 are also accounted for; the remaining six elements commuting with a given C_3 are two other C_3 elements, a C'_3 element and their inverses. A given C'_3 in turn commutes, in addition to its period, with three other C'_3 elements and their inverses—as well as three C_3, six C''_3 and their inverses. Finally, a given C''_3, in addition to its own period, commutes with two other C''_3, one C'_3 and their inverses.

C. The "Dihedral-like" Groups

We consider the finite groups first, that is, the cases $\Delta(3n^2)$ and $\Delta(6n^2)$, where n is an integer. The generators of these groups are given in Table I. We note from them that $\Delta(3n^2)$ is a subgroup of $\Delta(6n^2)$. Table VIII gives the group table for $\Delta(6n^2)$ as well as the definition for those element types which are not already defined in Table I. The corresponding group for $\Delta(3n^2)$ can be obtained from Table VIII by suppressing the rows and columns headed B, D, and F. For the finite groups $A(\alpha, \beta)$ is written $A(p, q)$ where

$$\alpha \to \alpha_n(p) = (2\pi/n)p, \quad p = 0, 1, \cdots n-1, \qquad \text{(II 4)}$$
$$\beta \to \beta_m(q) = (2\pi/m)q, \quad q = 0, 1, \cdots m-1,$$

and n and m are fixed integers. The arguments of

TABLE VIII. Group table for the Δ groups.

	$A(\alpha', \beta')$	$B(\alpha', \beta')$	$C(\alpha', \beta')$	$D(\alpha', \beta')$	$E(\alpha', \beta')$	$F(\alpha', \beta')$
$A(\alpha, \beta)$	$A(\alpha_1, \beta_1)$	$B(\alpha_1, \beta_1)$	$C(\alpha_1, \beta_1)$	$D(\alpha_1, \beta_1)$	$E(\alpha_1, \beta_1)$	$F(\alpha_1, \beta_1)$
$B(\alpha, \beta)$	$B(\alpha_1, \beta_2)$	$A(\alpha_1, \beta_2 + \pi)$	$F(\alpha_1, \beta_2)$	$E(\alpha_1, \beta_2 + \pi)$	$D(\alpha_1, \beta_2)$	$C(\alpha_1, \beta_2 + \pi)$
$C(\alpha, \beta)$	$C(\alpha_2, \beta_3)$	$D(\alpha_2 + \pi, \beta_3)$	$E(\alpha_2, \beta_3)$	$F(\alpha_2 + \pi, \beta_3)$	$A(\alpha_2, \beta_3)$	$B(\alpha_2 + \pi, \beta_3)$
$D(\alpha, \beta)$	$D(\alpha_3, \beta_3)$	$C(\alpha_3, \beta_3)$	$B(\alpha_3, \beta_3)$	$A(\alpha_3, \beta_3)$	$F(\alpha_3, \beta_3)$	$E(\alpha_3, \beta_3)$
$E(\alpha, \beta)$	$E(\alpha_3, \beta_2)$	$F(\alpha_3, \beta_2 + \pi)$	$A(\alpha_3, \beta_2)$	$B(\alpha_3, \beta_2 + \pi)$	$C(\alpha_3, \beta_2)$	$D(\alpha_3, \beta_2 + \pi)$
$F(\alpha, \beta)$	$F(\alpha_2, \beta_1)$	$E(\alpha_2 + \pi, \beta_1)$	$D(\alpha_2, \beta_1)$	$C(\alpha_2 + \pi, \beta_1)$	$B(\alpha_2, \beta_1)$	$A(\alpha_2 + \pi, \beta_1)$

$$\alpha_1 = \alpha + \alpha', \quad \beta_1 = \beta + \beta'$$
$$\alpha_2 = \alpha - \alpha' - \beta', \quad \beta_2 = \beta - \alpha' - \beta'$$
$$\alpha_3 = \alpha + \beta', \quad \beta_3 = \beta + \alpha'$$

$$A^{-1}(\alpha, \beta) = A(-\alpha, -\beta)$$
$$B^{-1}(\alpha, \beta) = B(-\alpha, \alpha + \beta + \pi)$$
$$C^{-1}(\alpha, \beta) = E(-\beta, \alpha + \beta)$$
$$D^{-1}(\alpha, \beta) = D(-\beta, -\alpha)$$
$$E^{-1}(\alpha, \beta) = C(\alpha + \beta, -\alpha)$$
$$F^{-1}(\alpha, \beta) = F(\alpha + \beta + \pi, -\beta)$$

The matrices $A(\alpha, \beta)$, $B(\alpha, \beta)$ and $E(\alpha, \beta)$ are defined in Table I.

$$D(\alpha, \beta) = \begin{pmatrix} 0 & e^{i\alpha} & 0 \\ e^{i\beta} & 0 & 0 \\ 0 & 0 & e^{i(\pi-\alpha-\beta)} \end{pmatrix}; \quad C(\alpha, \beta) = \begin{pmatrix} 0 & 0 & e^{i\alpha} \\ e^{i\beta} & 0 & 0 \\ 0 & e^{-i(\alpha+\beta)} & 0 \end{pmatrix}; \quad F(\alpha, \beta) = \begin{pmatrix} 0 & 0 & e^{i\alpha} \\ 0 & e^{i\beta} & 0 \\ e^{i(\pi-\alpha-\beta)} & 0 & 0 \end{pmatrix}$$

other element types are similarly defined. If n and m are not equal, use of the group table shows that

$$\alpha_n \subseteq \alpha_{nm} = (2\pi/nm)p, \qquad \beta_m \subseteq \beta_{nm} = (2\pi/nm)q. \tag{II 5}$$

Hence we need only consider $m = n$. Consider the element $A(p, q)$ for some given n. If p is fixed, q can take on n different values. Therefore there are n^2 different elements $A(p, q)$. Since there are three different element types, A, C, and E, in the $\Delta(3n^2)$ group, we conclude that the order of this group is $3n^2$. For $\Delta(6n^2)$ there are six element types; hence its order is $6n^2$, as is indicated already by the labeling we have used. It follows from the properties of the group table that when n is initially chosen to be an odd integer, the group corresponding to $2n$ is generated. We will therefore restrict the values of n in the $\Delta(6n^2)$ group to even integers.

From the group table it is possible to see how the group breaks up into classes. For example, for $\Delta(6n^2)$ we get

$$\begin{aligned} &A(p, q)A(p', q')A^{-1}(p, q) = A(p', q'), \\ &B(p, q)A(p', q')B^{-1}(p, q) = A(p', -p' - q'), \\ &C(p, q)A(p', q')C^{-1}(p, q) = A(-p' - q', p'), \\ &D(p, q)A(p', q')D^{-1}(p, q) = A(q', p'), \\ &E(p, q)A(p', q')E^{-1}(p, q) = A(q', -p' - q'), \\ &F(p, q)A(p', q')F^{-1}(p, q) = A(-p' - q', q'). \end{aligned} \tag{II 6}$$

Similar expressions can be written for the other element types. From Eqs. (II 6) we can immediately see, for example, that the set of elements $\{A(p', q'), A(p', -p' - q'), A(-p' - q', p'), A(q', p'), A(q', -p' - q'), A(-p' - q', q')\}$ form a class. There are at most six elements in this class, but there may be fewer for particular values of the arguments. The number of classes for $\Delta(3n^2)$ is $\frac{1}{3}(8 + n^2)$ when $\frac{1}{3}n$ is not an integer, and $(8 + \frac{1}{3}n^2)$ when $\frac{1}{3}n$ is an integer. For $\Delta(6n^2)$ the number of classes is $\frac{1}{6}(8 + 9n + n^2)$ when $\frac{1}{3}n$ is not an integer, and $\frac{1}{6}(24 + 9n + n^2)$ when $\frac{1}{3}n$ is an integer. When $\frac{1}{3}n$ is not an integer, the Δ groups are simultaneously subgroups of SU_3 and SU_3/C. When $\frac{1}{3}n$ is an integer, the Δ groups are subgroups only of SU_3. They contain a center which has to be divided out in order to obtain subgroups of SU_3/C.

TABLE IX. Class structure of $\Delta(3n^2)$ for $n = 4$, $p, q = 0, 1, 2, 3$.

$\{A(0, 0)\}$
$\{A(1, 0), A(3, 1), A(0, 3)\}$
$\{A(2, 0), A(2, 2), A(0, 2)\}$
$\{A(3, 0), A(1, 3), A(0, 1)\}$
$\{A(1, 1), A(2, 1), A(1, 2)\}$
$\{A(2, 3), A(3, 2), A(3, 3)\}$
$\{C(0, 0), C(1, 0), C(2, 0), C(3, 0), C(0, 1), C(1, 1), C(2, 1), C(3, 1), C(0, 2), C(1, 2), C(2, 2), C(3, 2), C(0, 3), C(1, 3), C(2, 3), C(3, 3)\}$
$\{E(0, 0), E(1, 0), E(2, 0), E(3, 0), E(0, 1), E(1, 1), E(2, 1), E(3, 1), E(0, 2), E(1, 2), E(2, 2), E(3, 2), E(0, 3), E(1, 3), E(2, 3), E(3, 3)\}$

As an example, we give the class structure of $\Delta(3n^2)$ for $n = 4$ in Table IX.

We use the technique of little groups[17] to find the irreducible representations of the Δ groups.[18] The $\Delta(3n^2)$ group has three 1-dimensional representations when $\frac{1}{3}n$ is not an integer which are generated by

[17] See for example, J. S. Lomont, *Applications of Finite Groups* (Academic Press Inc., New York, 1959), Chap. 5.

[18] For details of the explicit construction of these representations, see W. H. Klink, "Finite and Disconnected Subgroups of SU_3 and their Application to the Elementary Particle Spectrum," dissertation, Johns Hopkins University, Baltimore, Maryland, 1963 (unpublished).

the homomorphism $\Delta(3n^2)/A(p, q) \cong A_3$, where A_3 is the alternating group on three letters, and all $A(p, q)$ are mapped onto $A(0, 0)$, $C(p, q)$ onto $C(0, 0)$, and $E(p, q)$ onto $E(0, 0)$.

We will not concern ourselves with the cases when $\frac{1}{3}n$ is an integer any further. The Δ groups associated with these values of n do not differ significantly from those associated with other values of n. For example, for $\Delta(3n^2)$ and $\frac{1}{3}n$ an integer, there are nine instead of three 1-dimensional representations. But the characters for the 3-dimensional representations of $\frac{1}{3}n$ an integer are given by the same formula, given below, as those for other n.

The expression for the characters of the 3-dimensional representations is

$$\Delta_3^{m_1 m_2}(A(p, q)) = e^{(2\pi i/n)[m_1 p + m_2 q]} + e^{(2\pi i/n)[m_1 q - m_2(p+q)]} + e^{(2\pi i/n)[-m_1(p+q) + m_2 p]}, \quad \text{(II 7)}$$

$$\Delta_3^{m_1 m_2}(C(p, q)) = \Delta_3^{m_1 m_2}(E(p, q)) = 0.$$

The superscripts m_1 and m_2 label different irreducible representations and take on the values $0, 1, \cdots (n-1)$. Not all different values of m_1 and m_2 give inequivalent irreducible representations. Using the orthonormality relations for characters,[19] we see that $(-m_1 + m_2, -m_1)$ and $(-m_2, m_1 - m_2)$ are equivalent to (m_1, m_2). One must refer to Eq. (II 6) to obtain a typical p and q for each class. The total number of 3-dimensional representations when $\frac{1}{3}n$ is not an integer is $\frac{1}{3}(n^2 - 1)$.

Turning to the $\Delta(6n^2)$ group when $\frac{1}{3}n$ is not an integer, we obtain the following results: there are two 1-dimensional and one 2-dimensional irreducible representations generated by the homomorphism $\Delta(6n^2)/\{A(p, q)\} \cong S_3$, and the remaining irreducible representations are 3 and 6 dimensional. The characters of the 3-dimensional representations are given in Eq. (II 7) for element types A, C, and E, except that (m_1, m_2) can only take on the values $(m, 0)$, (m, m), and $(0, m)$. In other words, a single index m replaces m_1 and m_2. For the remaining element types B, D, and F, the characters are

$$\Delta_3^{m_1 m_2 t}(B(p, q)) = (-1)^{t+1} e^{(2\pi i/n)[m_1 p + m_2 q]},$$

$$\Delta_3^{m_1 m_2 t}(D(p, q)) = (-1)^{t+1} e^{(2\pi i/n)[m_1(\frac{1}{3}n - p - q) + m_2 p]}, \quad \text{(II 8)}$$

$$\Delta_3^{m_1 m_2 t}(F(p, q)) = (-1)^{t+1} e^{(2\pi i/n)[m_1 q + m_2(\frac{1}{3}n - p - q)]},$$

where m_1 and m_2 again can be replaced by $(m, 0)$, (m, m), and $(0, m)$, and the label t takes on the values 1 and 2. This additional label appears because the little group associated with the 3-dimensional representations contains a reflection operator. The same linear combination of m_1 and m_2 is equivalent to a given (m_1, m_2) as for (II 6) above. The total number of 3-dimensional irreducible representations when $\frac{1}{3}n$ is not an integer is $2(n-1)$. The characters of the 6-dimensional irreducible representations are

$$\Delta_6^{m_1 m_2}(A(p, q)) = e^{(2\pi i/n)[m_1 p + m_2 q]} + e^{(2\pi i/n)[m_1 p - m_2(p+q)]} + e^{(2\pi i/n)[m_1 q - m_2(p+q)]} + e^{(2\pi i/n)[m_1 q + m_2 p]} + e^{(2\pi i/n)[-m_1(p+q) + m_2 p]} + e^{(2\pi i/n)[-m_1(p+q) + m_2 q]}, \quad \text{(II 9)}$$

$$\Delta_6^{m_1 m_2}(B(p, q)) = \Delta_6^{m_1 m_2}(C(p, q)) = \Delta_6^{m_1 m_2}(D(p, q)) = \Delta_6^{m_1 m_2}(E(p, q)) = \Delta_6^{m_1 m_2}(F(p, q)) = 0.$$

The labels (m_1, m_2) cannot take the values $(m, 0)$, (m, m), or $(0, m)$. The representations labeled $(m_1 - m_2, -m_2)$, $(-m_2, m_1 - m_2)$, (m_2, m_1), $(-m_1 + m_2, -m_1)$, and $(-m_1, -m_1 + m_2)$ are equivalent to (m_1, m_2). The total number of 6-dimensional irreducible representations when $\frac{1}{3}n$ is not an integer is $\frac{1}{6}(n^2 - 3n + 2)$.

The $\lim_{n\to\infty} 2\pi/n$ generates the disconnected groups $\Delta(3\infty^2)$ and $\Delta(6\infty^2)$. They have essentially the same structure as their finite counterparts. Their irreducible representations are given by letting $(2\pi/n)p \to \alpha$ and $(2\pi/n)q \to \beta$. Naturally, statements about order and number of elements in a class no longer hold. The only other alteration we must make in the expression for the finite Δ groups is that the indices (m_1, m_2) labeling the irreducible representations can now take on the values $0, \pm 1, \pm 2, \cdots$. Previously the labelings corresponding to negative integers were redundant and therefore not considered.

III. APPLICATION TO THE PARTICLE SPECTRUM

A. Crystal Groups as Isotopic Spin Groups

Before discussing physical applications of the Σ and Δ groups, we would like to review and expand the arguments given by Case, Karplus, and Yang[10] for the application of finite subgroups of SU_2 to isotopic spin. For purposes of illustration, we consider only the double tetrahedral group T', the

[19] Reference 9, p. 104.

TABLE X. Character table of the double tetrahedral group, T'.

Element type	E	R	(C_2, C_2R)	C_3	C_3R	C_3^2R	C_3^2
Order of class	1	1	6	4	4	4	4
Number of commuting elements	24	24	4	6	6	6	6
Γ_1	1	1	1	1	1	1	1
$\bar{\Gamma}_1$	1	1	1	ω	ω	ω^2	ω^2
$\bar{\Gamma}_1{}^*$	1	1	1	ω^2	ω^2	ω	ω
Γ_2	2	-2	0	1	-1	1	-1
$\bar{\Gamma}_2$	2	-2	0	ω	$-\omega$	ω^2	$-\omega^2$
$\bar{\Gamma}_2{}^*$	2	-2	0	ω^2	$-\omega^2$	ω	$-\omega$
Γ_3	3	3	-1	0	0	0	0

character table of which is given in Table X. The same sort of analysis could be performed on the other double crystal groups.

We propose to analyze $\pi-N$ scattering in detail using T'. We use our knowledge of SU_2 to make the various isotopic multiplet assignments. The characters of Γ_1, Γ_2, and Γ_3 can be obtained by subducing[20] the representations of SU_2 corresponding to $T = 0$, $\frac{1}{2}$, and 1, respectively. Some of the elements related to the T_3 operator in SU_2 (i.e., $e^{iT_3\varphi}$ become elements of T'). Let us arbitrarily choose to relate $e^{iT_3\varphi}$ to the elements of highest symmetry, which in T' are of order three. We can therefore define a T_3 operator corresponding to Γ_1, Γ_2, and Γ_3, and the element C_3 (selected arbitrarily to be an element in the fourth column of Table X) has the form

$$C_3 = e^{\frac{2}{3}\pi i T_3}, \tag{III 1}$$

where T_3 is given by the expression familiar from SU_2:

$$\begin{aligned} T_3(\Gamma_1) &= 1, \\ T_3(\Gamma_2) &= \begin{pmatrix} \frac{1}{2} & \\ & -\frac{1}{2} \end{pmatrix}, \\ T_3(\Gamma_3) &= \begin{pmatrix} 1 & & \\ & 0 & \\ & & -1 \end{pmatrix}. \end{aligned} \tag{III 2}$$

We can thus assign the nucleons to the representation Γ_2 and the pions to Γ_3.

To analyze π–N scattering, we must consider the direct product

$$\Gamma_2 \otimes \Gamma_3 = \Gamma_2 \oplus \bar{\Gamma}_2 \oplus \bar{\Gamma}_2^*. \tag{III 3}$$

We therefore need the isotopic multiplet structure corresponding to $\bar{\Gamma}_2$ and Γ_2^*. In the representation $\bar{\Gamma}_1$

$$C_3 = \omega = e^{\frac{2}{3}\pi i T_3}, \tag{III 4}$$

which implies that $T_3 = 1$. Similarly for $\bar{\Gamma}_1^*$, we have $T_3 = -1$. Since

$$\bar{\Gamma}_2 = \Gamma_2 \otimes \bar{\Gamma}_1, \tag{III 5}$$

we immediately see that the isotopic content of $\bar{\Gamma}_2$ is $(\frac{3}{2}, \frac{1}{2})$. In a similar fashion the isotopic content of $\bar{\Gamma}_2^*$ is $(-\frac{1}{2}, -\frac{3}{2})$. Since T_3 is defined in terms of an element of order three, states having T_3 eigenvalues differing by multiples of three are equivalent. In particular $T_3 = \frac{3}{2}$ and $-\frac{3}{2}$ are equivalent, so one can expect linear combinations of π^+p and π^-n appearing in a single element of a basis. To see that such combinations do in fact appear, we obtain the Clebsch–Gordan coefficients arising from the product representation $\pi \otimes N$. To get the Clebsch–Gordan coefficients, it is necessary to construct nondiagonal elements of the various representations. One C_3 element has already been chosen to be diagonal. Any other C_3 has sufficient mixing properties to determine the Clebsch–Gordan coefficients. The results are

$$\begin{aligned} &\left\{ \begin{aligned} \Gamma_2(\tfrac{1}{2}) &= (\tfrac{2}{3})^{\frac{1}{2}} |\pi^+ n\rangle - \frac{1}{3^{\frac{1}{2}}} |\pi^0 p\rangle, \\ \Gamma_2(-\tfrac{1}{2}) &= \frac{1}{3^{\frac{1}{2}}} |\pi^0 n\rangle - (\tfrac{2}{3})^{\frac{1}{2}} |\pi^- p\rangle, \end{aligned} \right. \\ &\left\{ \begin{aligned} \bar{\Gamma}_2(\tfrac{3}{2}) &= \frac{1}{3^{\frac{1}{2}}} |\pi^+ p\rangle + (\tfrac{2}{3})^{\frac{1}{2}} |\pi^- n\rangle, \\ \bar{\Gamma}_2(\tfrac{1}{2}) &= \frac{\omega}{3^{\frac{1}{2}}} |\pi^+ n\rangle + \omega(\tfrac{2}{3})^{\frac{1}{2}} |\pi^0 p\rangle, \end{aligned} \right. \\ &\left\{ \begin{aligned} \bar{\Gamma}_2^*(-\tfrac{1}{2}) &= (\tfrac{2}{3})^{\frac{1}{2}} |\pi^0 n\rangle + \frac{1}{3^{\frac{1}{2}}} |\pi^- p\rangle, \\ \bar{\Gamma}_2^*(-\tfrac{3}{2}) &= \omega^2(\tfrac{2}{3})^{\frac{1}{2}} |\pi^+ p\rangle - \frac{\omega^2}{3^{\frac{1}{2}}} |\pi^- n\rangle. \end{aligned} \right. \end{aligned} \tag{III 6}$$

As expected, the T' group is adequate to represent isotopic multiplets for which isotopic spin is less than or equal to one, and so the Clebsch–Gordan coefficients for the Γ_2 representation (corresponding to isotopic spin $\frac{1}{2}$) are the usual ones. However, invariance under T' no longer forbids the reaction $\pi^+p \to \pi^-n$. In fact, if we define the scattering amplitude associated with $\bar{\Gamma}_2$ and $\bar{\Gamma}_2^*$ as $\mathfrak{a}(\bar{\Gamma}_2)$ and $\mathfrak{a}(\bar{\Gamma}_2^*)$, respectively, we get

$$\mathfrak{a}(\pi^+p \to \pi^-n) = \frac{(2^{\frac{1}{2}})}{3} [\mathfrak{a}(\bar{\Gamma}_2^*) - \mathfrak{a}(\bar{\Gamma}_2)], \tag{III 7}$$

[20] To subduce means to take away those elements of a group G not in its subgroup H. See Ref. 17, p. 219.

and since this reaction violates charge conservation, we must set it equal to zero. This implies

$$a(\bar{\Gamma}_2) = a(\bar{\Gamma}_2^*), \qquad \text{(III 8)}$$

which can occur in general only if the representations $\bar{\Gamma}_2$ and $\bar{\Gamma}_2^*$ coalesce into a single irreducible representation. If we consider the elements of $\bar{\Gamma}_2(T_3)$ and $\bar{\Gamma}_2^*(T_3)$ listed in Eq. (III 6) to form the basis for a single irreducible representation, a simple unitary transformation will take them into the standard isotopic multiplet form corresponding to isotopic spin $\frac{3}{2}$. Thus, the imposition of charge conservation has served to regenerate the 4-dimensional representation of SU_2. In fact, charge conservation actually regenerates the full SU_2 symmetry from T'. This is not surprising, since charge conservation requires the introduction of group elements corresponding to arbitrary rotations[21] about one axis. When such elements are adjoined to those of T', the full SU_2 group is generated by them.

B. $\Sigma(n)$ Groups as Higher Symmetry Groups

We are ready now to attempt the application of the $\Sigma(n)$ subgroups of SU_3 to the particle spectrum. Since we want to retain the isotopic spin multiplet structure as much as possible, we will consider only those $\Sigma(n)$ groups which have the double crystal groups as subgroups. This condition eliminates immediately all groups but $\Sigma(216)$ from consideration, since a necessary condition for a group H to be a subgroup of a larger group G is that the orders of the elements of H must exist in G. Since O' (the double octahedral group) has elements of order eight, and I' elements of order ten, an inspection of character Tables II–VII shows that neither O' nor I' can be subgroups of any of the Σ groups. Furthermore, one sees that T' could only be a subgroup of $\Sigma(216)$ since it has elements of order six. Using the techniques of Littlewood[15], one can prove that T' is indeed a subgroup of $\Sigma(216)$. The elements of T' arise from all classes of $\Sigma(216)$ except those labeled by the permutation type 3^3.

We are now faced with the problem of assigning particles to the various representations of $\Sigma(216)$ and of selecting commuting elements with which to associate two commuting operators. The elements of T' which we associated with T_3 in Sec. III A are elements of order six in the $\Sigma(216)$ group. But elements of this type commute only with their own periods and therefore no independent quantum number, in addition to the one associated with C_6, can be assigned. We must therefore go to elements of lower order in our attempt to assign quantum numbers. From the list of commuting elements of $\Sigma(216)$ given in Sec. II B, we see that nontrivial commuting elements are at most of order three. To find which elements of order three we can employ, we make use of the fact that some of the representations of $\Sigma(216)$ have the same structure as SU_3, and the possibility therefore exists of subducing[20] representations of SU_3 to get representations of $\Sigma(216)$, just as Γ_1, Γ_2, and Γ_3 were subduced representations of $T = 0$, $\frac{1}{2}$, and 1 in SU_2. The dimensions of the low-lying representations of SU_3 are one, three, six, and eight. The possible corresponding representations in $\Sigma(216)$ are Σ_1, $\bar{\Sigma}_1$, $\bar{\Sigma}_1^*$, Σ_3, Σ_8, $\bar{\Sigma}_8$, and $\bar{\Sigma}_8^*$. Since 1- and 8-dimensional representations of SU_3 have real characters, only Σ_1, Σ_3, and Σ_8 remain as possibilities. We know that in SU_3 there are two 3-dimensional irreducible representations, labeled $(1, 0)_3$ and $(0, 1)_3$, which are complex conjugates of each other, and for which

$$(0, 1)_3 \otimes (1, 0)_3 = (0, 0)_1 \oplus (1, 1)_8, \qquad \text{(III 9)}$$

where $(0, 0)_1$ and $(1, 1)_8$ are the 1- and 8-dimensional representations, respectively.[22] If there is any correspondence between the 3-dimensional representations of SU_3 and $\Sigma(216)$, $\Sigma_3(216)$ must result when both $(0, 1)_3$ and $(1, 0)_3$ are subduced. However,

$$\Sigma_3(216) \otimes \Sigma_3(216) = \Sigma_1 \oplus \bar{\Sigma}_1 \oplus \bar{\Sigma}_1^* \oplus \Sigma_3 \oplus \Sigma_3, \qquad \text{(III 10)}$$

and does not contain $\Sigma_8(216)$. Thus, the 3-dimensional representations of SU_3 and $\Sigma_3(216)$ do not correspond.[23] In fact, from the homomorphism $\Sigma(216)/\{9\} \cong T'$, discussed in Sec. II, it is clear that the 3-dimensional representation comes from a corresponding representation of T', a subgroup of SU_2. On the other hand, comparison of the outer products of $\Sigma_8(216)$ and $(1, 1)_8$ of SU_3,

$$\Sigma_8 \otimes \Sigma_8 = \Sigma_1^S \oplus \Sigma_8^S \oplus (\Sigma_8^S \oplus \bar{\Sigma}_8^S \oplus \bar{\Sigma}_8^{*S} \oplus \Sigma_3^S)$$

[21] If an element C_n, corresponding to a rotation of $(2\pi/n)\,p$ about an axis (p, n integer, $p \leqq n$) is related to some observable quantity (say charge), then it is conserved modulo n. In the limit $n \to \infty$, $(2\pi/n)\,p \to \theta$, this quantity will be conserved exactly.

[22] P. Tarjanne, Physica **105**, (1962).

[23] It may appear puzzling that generators for all the Σ groups given in Table I are 3-dimensional and therefore seem to provide a counterexample to the result we have just obtained. These matrices generate 3-dimensional representations of subgroups of SU_3. In those cases, where the subgroups of SU_3 differ from those of SU_3/C, we must factor out the center to obtain the groups corresponding to $\varphi = 1$. The 3-dimensional representations would then disappear. These considerations do not affect our character tables.

$$\oplus \Sigma_8^A \oplus (\bar{\Sigma}_8^A \oplus \bar{\Sigma}_2^{*A}) \oplus (\bar{\Sigma}_8^{*A} \oplus \bar{\Sigma}_2^A), \quad \text{(III 11)}$$

$$(1, 1)_8 \otimes (1, 1)_8 = (0, 0)_1^S \otimes (1, 1)_8^S \oplus (2, 2)_{27}^S \oplus (1, 1)_8^A \oplus (3, 0)_{10}^A \oplus (0, 3)_{10}^A$$

(where the superscripts S and A stand for symmetric and antisymmetric products), shows that $\Sigma_8(216)$ corresponds to $(1, 1)_8$ of SU_3, as can be seen from the close correspondence of the reduced product representations. We will show later that the representations of SU_3 of higher than eight dimensions, which cannot correspond to irreducible representations of $\Sigma(216)$, become reducible representations of $\Sigma(216)$. For example, $(0, 3)_{10}$ of SU_3 reduces to $\bar{\Sigma}_8^* \oplus \Sigma_2$.

Proceeding with our attempt to identify the quantum numbers, we next subduce $(1, 1)_8$ of SU_3 to obtain $\Sigma_8(216)$. Hypercharge is a diagonal operator in $(1, 1)_8$, with diagonal structure[21] $Y = (0, 1, 1, 0, -1, -1, 0, 0)$. Since we wish to have the operator Y appear in an element of order three, the group element $e^{i\theta_2 Y}$ of SU_3 must have the parameter θ_2 restricted to 0, $\frac{2}{3}\pi$ or $\frac{4}{3}\pi$. We obtain

$$\mathrm{Tr}\,(e^{i\theta_2 Y}) = 4(1 + \cos\theta_2) = \begin{cases} 8, & \theta_2 = 0, \\ 2, & \theta_2 = \frac{2}{3}\pi, \frac{4}{3}\pi. \end{cases} \quad \text{(III 12)}$$

Comparing with the character for $\Gamma_8(216)$, we see that the Y operator must be an element of the $1^3 3^2$ permutation classes. When one attempts the same sort of reasoning with T_3, one discovers that only $e^{i\theta_1 2T_3}$ will give an element of order three, since T_3 has eigenvalues $(1, \frac{1}{2}, -\frac{1}{2}, -1, -\frac{1}{2}, \frac{1}{2}, 0, 0)$ in the $(1, 1)_8$ representation. We obtain

$$\mathrm{Tr}\,(e^{i\theta_1 2T_3}) = 2(\cos 2\theta_1 + 2\cos\theta_1 + 1) = \begin{cases} 8, & \theta_1 = 0, \\ -1, & \theta_1 = \frac{2}{3}\pi, \frac{4}{3}\pi, \end{cases} \quad \text{(III 13)}$$

and therefore the T_3 operator must be an element of the 3^3 permutation classes of Table VII. However, no elements of T' exist in the 3^3 permutation classes. Thus, in spite of the fact that T' is a subgroup of $\Sigma(216)$, we lose the isotopic submultiplet structure. All we are left with is the possibility of assigning two independent quantum numbers with no trace of isotopic submultiplet structure left. Since charge independence has disappeared, we have a great deal of freedom to choose the second quantum number.[24] We arbitrarily pick Q, the charge operator, so that the basis function of the $\Sigma_8(216)$ representation will be

$$|Y, Q, \alpha\rangle_8 = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \\ 0 & -1 \\ -1 & -1 \\ -1 & 0 \\ 0 & 0\ 1 \\ 0 & 0\ 0 \end{pmatrix} = \begin{pmatrix} \Sigma^+ \\ p \\ n \\ \Sigma^- \\ \Xi^- \\ \Xi^0 \\ \Sigma^0 \\ \Lambda \end{pmatrix} = \begin{pmatrix} \pi^+ \\ K^+ \\ K^0 \\ \pi^- \\ K^- \\ \bar{K}^0 \\ \pi^0 \\ \eta \end{pmatrix}. \quad \text{(III 14)}$$

The index α is added to distinguish between Σ^0 and Λ, since total isotopic spin no longer exists to distinguish between these particles. With the assignment of Q as given above, the C_3 element related to Q must be in the permutation classes $1^3 3^2$, and is of the form

$$C_3 = e^{(4\pi i/3)Q}. \quad \text{(III 15)}$$

In order to make charge assignments for the other representations, it is necessary to be more specific about which elements of order three are related to Y and Q. We arbitrarily choose the diagonal elements $e^{\frac{1}{3}\pi i Y}$ to be in the class corresponding to the second column of Table VII. One then can show that the corresponding diagonal element related to Q is in the third column. We can therefore identify immediately the charge and hypercharge content of the following representations:

$$|\ \rangle_1 = (0, 0),$$
$$|\bar{}\rangle_1 = (1, -1),$$
$$|\bar{}\rangle_1^* = (-1, 1),$$

$$|\bar{}\rangle_8 = |\ \rangle_8 \otimes |\bar{}\rangle_1 = \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ -1 & -1 \\ 1 & 1 \\ 0 & 1 \\ 0 & -1 \\ 1 & -1\ 1 \\ 1 & -1\ 0 \end{pmatrix}, \quad \text{(III 16)}$$

[24] There is also no longer any reason why Σ groups other than $\Sigma(216)$ should not be analyzed. For convenience we will continue the analysis with $\Sigma(216)$. The results we obtain will be typical of the other Σ groups.

$$|\bar{\ }\rangle_8^* = |\ \rangle_8 \otimes |\bar{\ }\rangle_1^* = \begin{bmatrix} -1 & -1 \\ 0 & -1 \\ 0 & 1 \\ -1 & 0 \\ 1 & 0 \\ 1 & 1 \\ -1 & 1\ 1 \\ -1 & 1\ 0 \end{bmatrix}. \quad \text{(III 16 cont.)}$$

The charge and hypercharge assignments of the 2- and 3-dimensional representations can only be obtained, from the information given in the character table, after some further manipulations. The assignments are

$$|\ \rangle_2 = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad |\bar{\ }\rangle_2 = \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}, \quad |\bar{\ }\rangle_2^* = \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}, \quad |\ \rangle_3^{\sim} = \begin{bmatrix} 1 & -1 \\ 0 & 0 \\ -1 & 1 \end{bmatrix}. \quad \text{(III 17)}$$

We now carry out a procedure analogous to that of Sec. III A and consider baryon–meson scattering, where both the baryons and mesons are in the Σ_8 representation. Therefore we must consider the direct product $\Sigma_8 \otimes \Sigma_8$ whose reduction is given in Eq. (III 11), and study the Clebsch–Gordan problem for this case. The charge multiplet assignments we have made above are required in this analysis. Because Q and Y are defined in terms of elements of order three, the states having Q and Y eigenvalues which differ by multiples of three are equivalent. Since we have previously seen that these equivalent states are critical for the analysis of the structure of the representations, we will concern ourselves only with the Clebsch–Gordan coefficients involving them. In addition to the two diagonal elements of order three, it is necessary to construct nondiagonal elements for the various representations. We employ the generators $E(0, 0)$ and V^2 of Table I to construct the 8-dimensional elements, where we use the Clebsch–Gordan coefficients[25] related to Eq. (III 9) and the fact that $\Sigma_3(216 \times 3)$ given by the $\Sigma(216)$ generators in Table I corresponds to a subduced representation of $(0, 1)_3$ of SU_3. The element $E(0, 0)$ is sufficient to enable us to calculate the Clebsch–Gordan coefficients for the 2- and 3-dimensional representations, and is obtained by making use of the homomorphism $\Sigma(216)/\{9\} \cong T'$. The Clebsch–Gordan coefficients which result for the equivalent states with labeling ($Y = -1$, $Q = 1$) are

$$\begin{aligned}
|\overline{-1, 1, 1}\rangle_8^{*A} &= [1/2(3^{\frac{1}{2}})](\Sigma^+\overline{K^0} - \Xi^0\pi^+ \\
&\quad + 2pK^0 - 2K^+n - \Sigma^-K^- + \Xi^-\pi^-), \\
|\overline{-1, 1, 0}\rangle_8^{*A} &= \tfrac{1}{2}(-\Sigma^+\overline{K^0} + \Xi^0\pi^+ \\
&\quad - \Sigma^-K^- + \Xi^-\pi^-), \\
|\overline{-1, 1}\rangle_2^{A} &= (1/6^{\frac{1}{2}})(\Sigma^+\overline{K^0} - \Xi^0\pi^+ \\
&\quad - pK^0 + K^+n - \Sigma^-K^- + \Xi^-\pi^-), \\
|\overline{-1, 1, 1}\rangle_8^{*S} &= [1/2(3)^{\frac{1}{2}}](\Sigma^+\overline{K^0} + \Xi^0\pi^+ \\
&\quad - 2pK^0 - 2K^+n + \Sigma^-K^- + \Xi^-\pi^-), \\
|\overline{-1, 1, 0}\rangle_8^{*S} &= \tfrac{1}{2}(-\Sigma^+\overline{K^0} - \Xi^0\pi^+ \\
&\quad + \Sigma^-K^- + \Xi^-\pi^-), \\
|-1, 1\rangle_3^{S} &= (1/6^{\frac{1}{2}})(\Sigma^+\overline{K^0} + \Xi^0\pi^+ \\
&\quad + pK^0 + K^+n + \Sigma^-K^- + \Xi^-\pi^-).
\end{aligned} \quad \text{(III 18)}$$

We can see from the above that invariance under $\Sigma(216)$ allows reactions in which either charge or hypercharge or both are no longer conserved. We wish to demand only that charge conservation hold, and will impose no further condition on hypercharge. In other words, we require only that the reactions $\Sigma^+\bar{K}^0 \to \Sigma^-K^-$ and $\Sigma^+\bar{K}^0 \to \Xi^-\pi^-$ be forbidden. This condition yields, using Eq. (III 18),

$$\mathcal{A}(\bar{\Sigma}_8^{*A}) = \mathcal{A}(\bar{\Sigma}_2^{A}), \qquad \mathcal{A}(\bar{\Sigma}_8^{*S}) = \mathcal{A}(\Sigma_3^{S}), \quad \text{(III 19)}$$

where $\mathcal{A}(\Sigma_n)$ is the scattering amplitude corresponding to the irreducible representation Σ_n. Equations (III 19) hold in general only if the representations, to which the amplitudes correspond, coalesce. In other words $\bar{\Sigma}_8^{*A}$ and $\bar{\Sigma}_2^{A}$ coalesce into a single ten-dimensional representation which in fact is equivalent to $(0, 3)_{10}$ of SU_3. The $\bar{\Sigma}_8^{*S}$ and Σ_3^{S} also coalesce to form a segment of what eventually will become $(2, 2)_{27}$. In order to exhibit how $(2, 2)_{27}$, as well as the other ten-dimensional representation reappears, it is necessary to consider the other overlapping states (e.g., $|1, -1\rangle$). Thus, the imposition of charge conservation alone is sufficient to recover the full SU_3 structure from $\Sigma(216)$ (including the exact conservation of hypercharge).

[25] S. Gasiorowicz, "A Simple Graphical Method in the Analysis of SU_3," ANL-6729 (1963).

C. The Δ Groups as Higher Symmetry Groups

We turn finally to a study of the Δ groups. The charge conservation difficulty we have encountered in the Σ groups will occur also in the Δ groups of finite order. We will therefore consider only the disconnected Δ groups, and, in particular, we look at $\Delta(6\infty^2)$, since it possesses irreducible representations of higher dimension than $\Delta(3\infty^2)$. There is no 8-dimensional irreducible representation in this group, and so we are forced to follow the lines of the Sakata model.[11] We can take the generators $A(\alpha, \beta)$ of Table I as 3-dimensional diagonal elements. Choosing the basis function

$$|\ \rangle_3 = \begin{pmatrix} p \\ n \\ \Lambda \end{pmatrix}, \qquad \text{(III 20)}$$

we have

$$A(\alpha, \beta) = \exp i[(\alpha - \beta)T_3 + (\alpha + \beta)(\tfrac{3}{2}Y - I)], \qquad \text{(III 21)}$$

where the eigenvalues of T_3 and Y can be easily read off from the basis function in Eq. (III 20). Thus the operator T_3 and Y can be successfully defined. In order that the isotopic multiplet substructure also hold for $\Delta(6\infty^2)$, there must exist elements mixing the p–n doublet, i.e., group elements of the form

$$\begin{pmatrix} a & b & 0 \\ -b^* & a^* & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad |a|^2 + |b|^2 = 1, \qquad \text{(III 22)}$$

where neither a nor b can vanish. A glance at Tables I and VIII shows that there are no such elements in either $\Delta(3\infty^2)$ or $\Delta(6\infty^2)$. However, we notice that elements exist in both $\Delta(6\infty^2)$ and $\Delta(3\infty^2)$ which map a particle onto any other particle. Thus, invariance under the infinite Δ groups implies charge symmetry, but not charge independence, together with exact charge and hypercharge conservation.

IV. CONCLUSION

We have seen that although 8- and even ten-dimensional irreducible representations exist for the finite subgroups of SU_3 and we can make the usual particle assignments for baryon and meson octets, we do not find it possible to maintain even a limited isotopic spin invariance. Moreover, as expected, we run into the same difficulty with charge conservation as Case, Karplus, and Yang do for finite subgroups of SU_2. We find, in addition, that hypercharge is not conserved exactly. If charge conservation is imposed as a separate condition, it leads automatically to the conservation of hypercharge as well. In fact the imposition of the additional symmetry of charge conservation serves to regenerate the full SU_3 symmetry. Perhaps this result is not as surprising as it appears at first sight. As has been pointed out before,[26] and reiterated in the present paper, imposition of charge conservation can serve to generate continuous groups of more than one parameter. The charge operator appears in elements of the infinite Δ groups, of SU_2, and of SU_3. Since the Σ groups are not subgroups of the infinite Δ groups, or of the Abelian groups, or of SU_2, charge conservation cannot generate any of these continuous groups. All the representations we have obtained are unitary, as are elements containing the charge operator, and so presumably only compact groups can be generated from the Σ groups by charge conservation. All such subgroups of SU_3 have been enumerated[6] and we have been led to exclude them all. Thus, only the full SU_3 group is left as the one which could be generated by the imposition of charge conservation.

Of the groups considered by us, it is possible to include exact charge and hypercharge conservation from the beginning only in the disconnected groups. But these groups lead only to charge symmetry, and the octet model cannot be accommodated, since the maximum dimension of the irreducible representations is six.

There is always the possibility that some way of assigning charge and hypercharge operators other than by analogy to the full SU_3 group exists and might lead to different results in the physical applications than the ones we have obtained. In any case the detailed analysis of the groups carried out in Sec. II is obviously independent of the physical uses made of it in Sec. III, and may have other applications.

ACKNOWLEDGMENTS

We would like to thank Professor J. C. Ward and Professor G. Washnitzer for helpful discussions.

APPENDIX. THE IRREDUCIBLE REPRESENTATIONS OF SU_3/C

We turn now to a discussion of the continuous group SU_3/C in order to justify our consideration

[26] S. Weinberg, "On the Derivation of Intrinsic Symmetries" (1963) (preprint).

of only finite subgroups of SU_3/C rather than of SU_3 as well. The group SU_3/C has been referred to previously.[12,13] In what follows we will obtain the dimensions of the irreducible representations of SU_3/C. We recall that the center of a group, $\{C\}$, is the set of elements which commute with all elements of the group.[27] The identity is always a trivial element of the center. The center of SU_3 has two other elements, which, in the representation $(1, 0)_3$ can be written

$$\{C\} = \left\{ I = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}, \omega I, \omega^2 I \right\}. \tag{A1}$$

Since the center is an invariant subgroup, SU_3/C is a group which is called the linear fractional or projective group of SU_3. SU_3/C is locally isomorphic to SU_3, since both have the same Lie algebra. The difference between the two groups is that, while the change by 2π of any parameter characterizing an element of SU_3/C yields the element we started with, the parameter must generally be changed by $3 \cdot 2\pi$ for all representations of SU_3 to return to the element we started with. An analogous statement can be made for $SU_2/C \cong O_3$ and SU_2 except that in SU_2 an element is regained if the parameter is changed by $2 \cdot 2\pi$. The irreducible representations of O_3 are the same as some of the irreducible representations of SU_2, but the irreducible SU_2 representations of even dimension do not occur in O_3. We can expect similarly that some of the representations of SU_3 are not representations of SU_3/C. We see which representations are eliminated by calculating the characters of SU_3. Weyl[28] gives the following expression for these:

$$\chi_{\lambda_1,\lambda_2}(\varphi_1, \varphi_2) = p_{\lambda_1+\lambda_2} p_{\lambda_2} - p_{\lambda_1+\lambda_2+1} p_{\lambda_2-1}, \tag{A2}$$

where p_λ is defined by

$$\sum_{\lambda=0}^{\infty} p_\lambda z^\lambda = [(1 - ze^{i\varphi_1})(1 - ze^{i\varphi_2})(1 - ze^{-i(\varphi_1+\varphi_2)})]^{-1}, \tag{A3}$$

and

$$p_{-1} = p_{-2} = \cdots = 0.$$

(λ_1, λ_2) are the weight diagram indices given in Ref. 1. The first few characters are

$$\begin{aligned}
\chi_{(0,0)}(\varphi_1, \varphi_2) &= 1, \\
\chi_{(1,0)}(\varphi_1, \varphi_2) &= e^{i\varphi_1} + e^{i\varphi_2} + e^{-i(\varphi_1+\varphi_2)}, \\
\chi_{(2,0)}(\varphi_1, \varphi_2) &= e^{2i\varphi_1} + e^{2i\varphi_2} + e^{-2i(\varphi_1+\varphi_2)} \\
&\quad + e^{i(\varphi_1+\varphi_2)} + e^{-i\varphi_1} + e^{-i\varphi_2}, \\
\chi_{(1,1)}(\varphi_1, \varphi_2) &= e^{i(2\varphi_1+\varphi_2)} + e^{i(\varphi_1+2\varphi_2)} + e^{i(\varphi_1-\varphi_2)} \\
&\quad + e^{i(-\varphi_1+\varphi_2)} + e^{i(-\varphi_1-2\varphi_2)} + e^{i(-2\varphi_1-\varphi_2)} + 2.
\end{aligned} \tag{A4}$$

Typical of representations of SU_3 which have a center of distinct elements is $(1, 0)_3$, whose center is given by (A 1). Suppose an element g in this representation has the character

$$\mathrm{Tr}\,(g) = \chi_{(1,0)}(\varphi_1, \varphi_2). \tag{A5}$$

The elements ωg and $\omega^2 g$ are distinct. They have the characters

$$\begin{aligned}
\mathrm{Tr}\,(\omega g) &= \chi_{(1,0)}(\varphi_1 + \tfrac{2}{3}\pi, \varphi_2 + \tfrac{2}{3}\pi), \\
\mathrm{Tr}\,(\omega^2 g) &= \chi_{(1,0)}(\varphi_1 + \tfrac{4}{3}\pi, \varphi_2 + \tfrac{4}{3}\pi).
\end{aligned} \tag{A6}$$

Conditions like (A 5) and (A 6) obtain for all representations whose centers have three distinct elements. On the other hand, those representations which have only the identity element in the center (and therefore are representations of SU_3/C) must have characters which are invariant under the transformation

$$\varphi_i \to \varphi_i + \tfrac{2}{3}\pi, \qquad \varphi_i + \tfrac{4}{3}\pi, \tag{A7}$$

where $i = 1$ or 2. For the list of characters given above, only $\chi_{(0,0)}$ and $\chi_{(1,1)}$ satisfy the invariance requirement. The general condition satisfied by representations $(\lambda_1, \lambda_2)_d$ of SU_3/C is precisely the empirical expression given by Sakurai,[14] namely

$$|\lambda_1 - \lambda_2| = 3n, \qquad n = 0, 1, 2 \cdots. \tag{A8}$$

Thus, strictly speaking, the Sakata model and the "eightfold way" are particle assignments associated with different, if closely related, groups.

[27] Reference 17, p. 23.
[28] H. Weyl, *The Classical Groups* (Princeton University Press, Princeton, New Jersey, 1939), Chap. 6, Sec. 5.

On Pais's Charge Correlation Coefficients

J. M. Shpiz
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received 25 February 1964)

An alternative proof of a theorem of Pais which states that the charge correlation coefficients for an n–π system are independent of the row of the representation of the symmetric group is presented in which no detailed knowledge of the representations is used.

CONSIDER an n–π system whose state vector $\varphi_a^{[N]}$ belongs to the ath row of the $[N]$th irreducible representation of the group S_n of permutations of charge indices. Then $\varphi_a^{[N]}$ is some linear combination of state vectors $X_a^{[N]}(m_+m_-m_0)$ which belong to charge partition $m = (m_+m_-m_0)$—that is $X_a^{[N]}(m)$ describes a state in which there are $m_+\pi^+$, $m_-\pi^-$ and $m_0\pi^0$.

We write

$$\varphi_a^{[N]} = \sum_{\text{distinct } (m)} C_a^{[N]}(m) X_a^{[N]}(m) \tag{1}$$

and assume that the state vectors are normalized,

$$(\varphi_a^{[N]} \mid \varphi_a^{[N]}) = (X_b^{[N]} \mid X_b^{[N]}) = 1. \tag{2}$$

Pais[1] has proven that the charge correlation coefficients which are the magnitudes $|C_a^{[N]}(m)|$ are independent of a, by using an explicit construction of the representations. We present here an alternative proof which uses no detailed knowledge of the explicit nature of the irreducible representation of S_n.

The essential feature is that the statement "$X_a^{[N]}(m)$ belongs to the charge partition $(m) = (m_+m_-m_0)$" is an observable one and is independent of the labeling of the individual pions. Hence we may construct a projection operator $\Lambda^{(m)}$ which is unity when acting on a state of charge partition (m) and zero when acting on any other state and is independent of the labels of the individual pions so that $\Lambda^{(m)}$ commutes with permutations of the pion labels. We first prove the theorem for unitary irreducible representations and then extend it to general irreducible representations.

Let P denote a permutation of pion labels, then[2]

$$P\varphi_a^{[N]} = \sum_b \varphi_b^{[N]} D^{[N]}(P)_{ba}. \tag{3}$$

The order of the group is $n!$ and if we denote the dimension of the representation $D^{[N]}$ by $d[N]$ we have

$$\sum_P D^{[N]}(P)^*_{ab} D^{[N]}(P)_{cd} = \frac{n!}{d[N]} \delta_{ac}\delta_{bd}. \tag{4}$$

From which, since $[P, \Lambda^{(m)}] = 0$,

$$|C_a^{[N]}(m)|^2 = (\Lambda^{(m)}\varphi_a^{[N]} \mid \Lambda^{(m)}\varphi_a^{[N]})$$

$$= \frac{1}{n!}\sum_P (\Lambda^{(m)}\varphi_a^{[N]} \mid P^{-1}P \mid \Lambda^{(m)}\varphi_a^{[N]})$$

$$= \frac{1}{n!}\sum_P D^{[N]}(P)^*_{ba} D^{[N]}(P)_{ca} (\Lambda^{(m)}\varphi_b^{[N]} \mid \Lambda^{(m)}\varphi_c^{[N]}),$$

whence

$$|C_a^{[N]}(m)|^2 = \frac{1}{d[N]}\sum_b |C_b^{[N]}(m)|^2. \tag{5}$$

Thus $|C_a^{[N]}(m)|$ is independent of the row index a, for unitary $D^{[N]}$.

Suppose now that $\bar{\varphi}_a^{[N]}$ belongs to a nonunitary irreducible representation; then it is equivalent to a unitary representation

$$\bar{\varphi}_a^{[N]} = \sum_b S_{ab}\varphi_b^{[N]}. \tag{6}$$

The new states of charge partition (m), $\bar{X}_a^{[N]}(m)$ and the new charge correlation coefficients $\bar{C}_a^{[N]}(m)$ are defined as before by

$$\bar{C}_a^{[N]}(m)\bar{X}_a^{[N]}(m) \equiv \Lambda^{(m)}\bar{\varphi}_a^{[N]} = \Lambda^{(m)}\sum_b S_{ab}\varphi_b^{[N]}$$

$$= \sum_b S_{ab} C_b^{[N]}(m) X_b^{[N]}(m). \tag{7}$$

We take $(\bar{\varphi}_a^{[N]} \mid \bar{\varphi}_a^{[N]}) = 1$, and observe that the unitary irreducible nature of $D^{[N]}(P)$ implies $(X_a^{[N]}(m) \mid X_b^{[N]}(m)) = \delta_{ab}$, and use the fact that $|C_b^{[N]}(m)|$ is independent of b, we obtain

$$|\bar{C}_a^{[N]}(m)|^2 = \sum_b |S_{ab}|^2 \, |C_b^{[N]}(m)|^2$$

$$= |C_d^{[N]}(m)|^2 \sum_b |S_{ab}|^2 \tag{8}$$

or, since $\Sigma_b |S_{ab}|^2 = (\bar{\varphi}_b \mid \bar{\varphi}_b) = 1$,

$$|\bar{C}_a^{[N]}(m)|^2 = |C_d^{[N]}(m)|^2 \tag{9}$$

Equation (9) states that the charge correlation coefficients are independent of the basis as well as independent of the row. Q.E.D.

[1] A. Pais, Ann. Phys. 9, 548 (1960). We are here concerned with Theorem E of this paper. Professor Pais has published a sequel to this article in Ann. Phys. 22, 274 (1963).

[2] E. P. Wigner, *Group Theory and Its Applications to the Quantum Mechanics of Atomic Spectra*, English translation by J. J. Griffin (Academic Press Inc., New York, 1959).

Renormalization of Singlet Amplitude in Intermediate-Vector-Boson Theory of Weak Interactions*

M. BOLSTERLI
School of Physics, University of Minnesota, Minneapolis, Minnesota and University of California, Los Alamos Scientific Laboratory, Los Alamos, New Mexico
(Received 18 February 1964)

A simple prescription for renormalizing the singlet amplitude in the intermediate-vector-boson theory of weak interactions is given.

IN a recent paper,[1] it has been shown that the Bethe–Salpeter equation for the triplet amplitude in the intermediate-vector-boson theory of weak interactions has a unique solution in configuration space; however, this configuration-space solution is so singular near the light cone that its Fourier transform does not exist, although there does exist a method of obtaining zero for the triplet amplitude on the mass shell. In this note, a simple consideration of the singlet amplitude leads to a prescription for its renormalization. The results of "peratization" theory[2] are immediately obtained. Unfortunately, the triplet amplitude does not appear to be susceptible to similar treatment.

The Born term for the singlet amplitude is

$$P_+\Lambda P_- i(2\pi)^4 g^2(4 - k^2/M^2)/(k^2 - M^2 + iO), \quad (1)$$

where M is the Boson mass and

$$\Lambda = \tfrac{1}{4}\gamma_\mu \otimes \gamma^\mu, \; P_\pm = \gamma_\pm \otimes \gamma_\pm, \; \gamma_\pm = \tfrac{1}{2}(1 \pm i\gamma_5). \quad (2)$$

The fact that Λ^2 is the singlet projection operator, and that

$$\Lambda^2 P_+ \mathbf{k} \otimes \mathbf{k} P_- \Lambda^2 = k^2 P_+ \Lambda P_- \quad (3)$$

where

$$\mathbf{k} = k\cdot\gamma,$$

has been used in writing (1). The compound vertex (1) can also be written

$$P_+\Lambda P_- \cdot 3i(2\pi)^4 g^2/(k^2 - M^2 + iO) - P_+\Lambda P_- i(2\pi)^4 g^2/M^2. \quad (4)$$

In this form, the first term is an interaction like that mediated by a scalar or pseudoscalar meson, differing in that the vertex contains the factor $P_+\Lambda P_-$ instead of 1 or $\gamma_5 \otimes \gamma_5$. Hence, the first term in (4) is renormalizable in the ordinary sense and can be ignored as far as questions of finiteness are concerned. The second term in (4) is identical with the simple vertex generated by the 4-fermion interaction

$$\mathfrak{L}_{\text{int}} = -\tfrac{1}{4}(g^2/M^2)(\bar\psi\gamma_\mu\gamma_-\psi)(\bar\psi\gamma^\mu\gamma_-\psi). \quad (5)$$

This latter is not renormalizable, nor does it lead to a tractable Bethe–Salpeter equation, as will immediately be seen.

The Bethe–Salpeter equation in the theory with the 4-fermion interaction (5) is (in the notation of Ref. 1)

$$A_\pm^{FF}(p, k, k_0) = i(2\pi)^4(g^2/M^2) \times \left[1 \pm (2\pi)^{-8} \int \frac{(k_1^2 - p^2)A_\pm^{FF}(p, k_1, k_0)\, d^4k_1}{[(p + k_1)^2 - m_i^2][(p - k_1)^2 - m_r^2]}\right] \quad (6)$$

where the final 4-momenta are $p \pm k$ and the initial ones are $p \pm k_0$,

$$A_\pm^{FF}(p, k, k_0) = A_1^{FF}(p, k, k_0) \pm A_2^{FF}(p, -k, k_0), \quad (7)$$

and A_1^{FF} and A_2^{FF} are defined as in Ref. 1. A common factor $P_+\Lambda P_-$ has been removed from all terms in Eq. (6). It follows from Eq. (6) that $A^{FF}(p, k, k_0)$ is independent of k and, therefore, by symmetry, also of k_0. Thus, Eq. (6) becomes

$$A_\pm^{FF}(p) = i(2\pi)^4(g^2/M^2)[1 \pm (2\pi)^{-8} I(p) A_\pm^{FF}(p)], \quad (8)$$

$$I(p) = \int \frac{k^2 - p^2}{[(p + k)^2 - m_i^2][(p - k)^2 - m_r^2]}\, d^4k = \infty. \quad (9)$$

With $I(p) = \infty$, Eq. (8) is meaningless. If the propagators in (9) are regularized so that $D_F^{\text{reg}}(0) < \infty$, then it is clear from (8) that

$$\lim_{M_i \to \infty} A_\pm^{FF}(p, M_i) = 0, \quad (10)$$

where M_i are the regulator masses.

An alternative way of obtaining the result (10)

* Work performed at Los Alamos Scientific Laboratory under the auspices of the U. S. Atomic Energy Commission.
[1] K. Bardacki, M. Bolsterli, and H. Suura, Phys. Rev. **133**, B 1273 (1964).
[2] G. Feinberg and A. Pais, Phys. Rev. **131**, 2724 (1963); and Phys. Rev. **133**, B 477 (1964); also Y. Pwu and T. T. Wu, Phys. Rev. **133**, B 778 (1964).

for $A_{\pm}^{FF}(p)$ is to eliminate it from the very beginning by *subtracting* the 4-fermion interaction (3) from the original Lagrangian containing only the vector-boson interaction. This subtraction procedure must give all the "peratized" results[2] for the singlet amplitude, for example, the 3 instead of 4 in the first term of (2). It is clear from (2) and the remarks immediately following it that the subtraction leads to a renormalizable theory as long as only singlet 2-fermion interactions are considered. The triplet 2-fermion amplitude is unaffected by the subtraction.

The only subtraction procedure that appears to work on the triplet interaction is to subtract the interaction

$$\bar{\psi}\vec{\partial}\psi\varphi \tag{11}$$

with a scalar meson in such a way as to cancel the entire $\mathbf{k} \otimes \mathbf{k}$ term in the numerator of the vector-meson propagator. This would leave a Born term with no triplet part:

$$P_{+}\Lambda P_{-}i(2\pi)^4 g^2 4/(k^2 - M^2 + i0), \tag{12}$$

which is again renormalizable. This corresponds to taking a linear combination of vector and vector-coupled scalar fields, both with the same mass, in such a way that the effective meson propagator is proportional to $g_{\mu\nu}(k^2 - M^2 + i0)^{-1}$.

This procedure would give zero for the triplet amplitude and the previous results for the singlet amplitude (with an adjustment of the unrenormalized coupling constant).

Upper and Lower Bounds on Generalized Fourier Coefficients*

L. M. DELVES
Applied Mathematics Department, The University of New South Wales, Kensington, N. S. W., Australia
(Received 11 October 1963)

Methods are given for obtaining variational upper and lower bounds on the nth Fourier coefficient $C^2_n = |\int g^* \sigma U_n d\tau|^2$ of a function g relative to a sequence of eigenfunctions U_n. The methods differ in their ease of application and in the amount of information required concerning the eigenvalues associated with the U_n. Some illustrative examples are given.

I. INTRODUCTION

We consider the eigenvalue equation

$$[H - \lambda_n \sigma] U_n = 0 \tag{1.1}$$

for eigenvalue λ_n, eigenfunction U_n, and suppose that the operators H, σ are Hermitian and such that the spectrum of λ is bounded from below and has been ordered. We also ignore possible degeneracies in the spectrum. We are not concerned with the detailed form of H and σ, which may be (for instance) finite matrices or differential or integral operators; but we use here the notation of finite matrices and vectors, so that in particular the inner product of two functions V_1 and V_2 is represented by the notation $V_1^\dagger V_2$ rather than $\int V_1^* V_2 \, d\tau$. This notation is chosen only for its brevity and ease of printing; the results we obtain are applicable to a wide class of operators and spaces.

With the conditions stated, the eigenfunctions U_n form a complete set and are orthogonal with weight function σ; we choose the normalization so that

$$U_n^\dagger \sigma U_m = \delta_{nm}. \tag{1.2}$$

Then an arbitrary function g can be expanded in terms of the U_n:

$$g = \sum_{m=1}^{\infty} C_m U_m, \qquad C_n = U_n^\dagger \sigma g. \tag{1.3}$$

The coefficient C_n is called the nth generalized Fourier coefficient of g, with respect to the sequence $\{U_m\}$. For a given function g, the phases of the U_n can be chosen so that C_n is real and positive; we shall assume this has been done. We shall be interested in finding approximations to C_n when the eigenfunctions U_n are not known. This problem has been considered previously by Rayner.[1,2] He gave[1] a neat method of finding an upper bound on C_n^2 given only that λ_n is known. This bound has the advantage that it is *variational*, so that equality can in principle be achieved. He also gave[2] a method for constructing a sequence of upper and lower bounds to C_1^2, which involves only a minimum of knowledge of the eigenvalues, but this method has several disadvantages. The most serious is that it involves the solution of a sequence of equations of the form

$$H\varphi_n - \sigma\varphi_{n-1} = 0. \tag{1.4}$$

These are "simpler" than (1.1) in that they are not eigenvalue equations; nonetheless it will usually be the case that they are not soluble. A further disadvantage is that any given term in the sequence of bounds is of strictly limited accuracy; the sequences converge to C_1^2, but the convergence may be slow. We call such methods "nonvariational," with the implication that they are inferior to variational methods; but the superiority of a variational method depends on the choice of a good trial function, and where this is difficult a nonvariational method may be superior.

The final disadvantage of Rayner's method is that it does not give bounds on any but C_1^2.

We give here several methods of bounding C_n^2. These methods fall into three classes:

(a) We give nonvariational upper and lower bounds on C_n^2 which do not involve the solutions of any differential equations. These bounds are easy to compute, but their closeness to C_n^2 depends on the form of g.

(b) We give a sequence of nonvariational upper and lower bounds for C_n^2 which do involve the solution of a sequence of equations. These methods can be regarded as the generalization to higher coefficients of the methods of Ref 2.

(c) We give a variational *lower* bound on C_n^2.

Examples are given of the use of each principle. One interesting application of these bounds is given

* This research was supported in part by U. S. Air Force Grant No. 62-400 to the University of New South Wales.
[1] M. E. Rayner, Quart. J. Math. 13, 61 (1962).
[2] Reference 1, p. 137.

in a companion paper, in which it is shown that the lower bound can be used to derive bounds on the expectation value $\langle W\rangle_n$:

$$\langle W\rangle_n = U_n^+ W U_n$$

for an arbitrary (bounded) operator W.

II. UPPER AND LOWER BOUNDS ON FOURIER COEFFICIENTS

A. A Variational Upper Bound Due to Rayner

We repeat here for convenience the variational upper bound given by Rayner.[1] For an arbitrary function g, the nth Fourier coefficient C_n [Eq. (1.3)] is bounded by

$$C_n^2 \leq B^+(f)\sigma^{-1}B(f), \tag{2.1}$$

where $B(f) = H(f - g) - \lambda_n\sigma f$. Equality is reached in (2.1) if

$$(H - \lambda_n\sigma)(f - g) = \lambda_n\sigma(g - C_nU_n) \tag{2.1a}$$

and f is an arbitrary function provided that H is Hermitian with respect to the function $f - g$. For example, if H is a differential operator, f must be chosen so that $f - g$ satisfies the same boundary conditions as the U_n.

B. A Sequence of Upper and Lower Bounds for C_n^2

In his second paper,[2] Rayner constructs a sequence of upper and lower bounds for the first Fourier coefficient C_1^2. The sequences involve the solution of a sequence of linear equations, and converge to C_1^2. We give here a similar method by which sequences which converge to C_p^2, $p > 1$, may be constructed. These methods involve only a minimum of information about the eigenvalues λ_n of (1.1).

Suppose that we have an approximation α_p to the eigenvalue λ_p, which is sufficiently good to be closer to λ_p than any other λ_n:

$$|\lambda_n - \alpha|_{\min} = |\lambda_p - \alpha|. \tag{2.2}$$

Then let us construct a sequence of functions g_n defined by

$$g_0 \equiv g,$$
$$(H - \alpha\sigma)g_n + \sigma g_{n-1} = 0, \qquad n \geq 1. \tag{2.3}$$

Then

$$g_m^+\sigma U_n = [1/(\alpha_p - \lambda_n)]g_m^+(-H + \alpha_p\sigma)U_n$$
$$= [1/(\alpha_p - \lambda_n)]g_{m-1}^+\sigma U_n \quad \text{by (2.3).} \tag{2.4}$$

Let us define

$$1/(\alpha_p - \lambda_n) = \beta_n. \tag{2.4a}$$

Then by repetition of this process we find

$$g_m^+\sigma U_n = [1/(\alpha_p - \lambda_n)]^m g_0^+\sigma U_n$$
$$= \beta_n^m C_n. \tag{2.5}$$

We define a sequence of numbers a_m by

$$a_m = g_m^+\sigma g \tag{2.6}$$
$$= \sum_{n=1}^{\infty} \beta_n^m C_n^2.$$

We can give a sequence of upper and lower bounds on C_p^2 in terms of the sequence $\{a_m\}$.

Upper Bound on C_p^4

From the sequence $\{a_m\}$ we can construct a sequence $\{\gamma_m\}$ as follows:

$$\gamma_m = a_m^{m+2}/a_{m+2}^m. \tag{2.7}$$

Then we have the following theorem:

$$C_p^4 \leq \gamma_m \qquad m \text{ even}, \tag{2.8}$$
$$\{\gamma_m\} \to C_p^4 \qquad m \to \infty,$$

so that the γ_m form a convergent sequence of upper bounds. The proof follows the similar proof of Rayner. We note first that the sequence β_n is such that

$$\max\{|\beta_n|\} = \beta_p. \tag{2.9}$$

Substituting (2.6) into the definition (2.7) of γ_m we find for even m

$$\gamma_m^{1/(m+2)} - \gamma_{m+2}^{1/(m+2)} = \frac{(\sum \beta_n^m C_n^2)(\sum \beta_n^{m+4} C_n^2) - (\sum \beta_n^{m+2} C_n^2)^2}{(\sum \beta_n^{m+2} C_n^2)^{m/(m+2)} \sum \beta_n^{m+4} C_n^2}.$$

But for even m the denominator is positive. Moreover,

$$(\sum \beta_n^{m+2} C_n^2)^2 = [\sum \beta_n^{\frac{1}{2}(m+2)} C_n \beta_n^{\frac{1}{2}m} C_n]^2 \leq \sum \beta_n^{m+4} C_n^2 \sum \beta_n^m C_n^2$$

by Cauchy's inequality, so that

$$\gamma_{m+2} \leq \gamma_m, \quad m \text{ even}. \tag{2.10}$$

Moreover, since $\beta_n^2 < \beta_p^2$ for $n \neq p$ we have

$$1 + \frac{1}{C_p^2}\sum{}' \left(\frac{\beta_n}{\beta_p}\right)^{m+2} C_n^2 < 1 + \frac{1}{C_p^2}\sum{}' \left(\frac{\beta_n}{\beta_p}\right)^m C_n^2, \tag{2.11}$$

where $\sum'$ denotes a summation over $n \neq p$. Thus

$$\gamma_m = \frac{C_p^4\{1 + (1/C_p^2)\sum' (\beta_n/\beta_p)^m C_n^2\}^{m+2}}{\{1 + (1/C_p^2)\sum' (\beta_n/\beta_p)^{m+2} C_n^2\}^m} > C_p^4\left\{1 + \frac{1}{C_p^2}\sum{}' \left(\frac{\beta_n}{\beta_p}\right)^m C_n^2\right\}^2. \tag{2.12}$$

That is, $\gamma_m > C_p^4$.

Moreover,

$$\sum{}' \left(\frac{\beta_n}{\beta_p}\right)^m C_n^2 > \left(\frac{\beta_q}{\beta_p}\right)^m \sum C_n^2 = \left(\frac{\beta_q}{\beta_p}\right)^m N, \quad (2.13)$$

where β_q is the larger in modulus of β_{p-1}, β_{p+1} and

$$N = g^+ \sigma g. \quad (2.13a)$$

Hence

$$\gamma_m < C_p^4\left[1 + \frac{1}{C_p^2}\sum{}' (\beta_n)^m C_n^2\right]^{m+2}$$
$$< C_p^4\left[1 + \frac{N}{C_p^2}\left(\frac{\beta_q}{\beta_p}\right)^m\right]^{m+2}.$$

But $(\beta_q/\beta_p)^m < 1$ for all m so that as $m \to \infty$

$$\gamma_m \to C_p^4.$$

We have thus proved that the sequence $\{\gamma_m\}$ converges to C_p^4, while for even m the γ_m form a steadily decreasing sequence of upper bounds for C_p^4.

Lower Bounds on C_p^2

Suppose that we can find a constant M such that

$$|\beta_q/\beta_p| \leq M < 1, \quad (2.14a)$$

and suppose that m is even and such that

$$N(m+2)M^m/\gamma_m^{\frac{1}{2}} < 1. \quad (2.14b)$$

In these formulas, β_q and N are defined in (2.13a).

Then if we define δ_m through

$$\delta_m = \tfrac{1}{2}\gamma_m[1 + (1 - N(m+2)M^m/\gamma_m^{\frac{1}{2}})^{\frac{1}{2}}], \quad (2.15)$$

we have the convergent inequality

$$C_p^2 > \delta_m, \qquad \delta_m \to C_p^2 \qquad m \to \infty. \quad (2.16)$$

Proof:

We have

$$\gamma_m = C_p^4 \frac{[1 + (1/C_p^2)\sum{}' (\beta_n/\beta_p)^m C_n^2]^{m+2}}{[1 + (1/C_p^2)\sum{}' (\beta_n/\beta_p)^{m+2} C_n^2]^m}$$
$$< C_p^4\left[1 + \frac{1}{C_p^2}\sum{}' \left(\frac{\beta_n}{\beta_p}\right)^m C_n^2\right]^{m+2}$$
$$< C_p^4[1 + (N/C_p^2)(\beta_q/\beta_p)^m]^{m+2} \quad \text{[by (2.13)]}$$
$$< C_p^4[1 + NM^m/C_p^2]^{m+2} \quad \text{[by (2.14)]}$$
$$C_p^2 > \gamma_m^{\frac{1}{2}}[1 + NM^m/C_p^2]^{\frac{1}{2}(-m-2)}$$
$$> \gamma_m^{\frac{1}{2}}[1 - (m+2)NM^m/2C_p^2]$$

by the second mean value theorem, so that

$$C_p^4 - C_p^2\gamma_m^{\frac{1}{2}} + NM^m(m+2)\gamma_m^{\frac{1}{2}}/2 > 0,$$

whence as in Ref. 1 it follows that if

$$N(m+2)M^m/\gamma_m^{\frac{1}{2}} < 1,$$
$$C_p^2 > \tfrac{1}{2}\gamma_m^{\frac{1}{2}}[1 + (1 - NM^m(m+2)/\gamma_m^{\frac{1}{2}})^{\frac{1}{2}}]. \quad (2.17)$$

Moreover, the convergence of the right-hand side of (2.17) to C_p^2 as $m \to \infty$ is immediate since the γ_m are bounded below by C_p^4 and $M < 1$.

Alternative Lower and Upper Bounds on C_p^2

The lower bound (2.17) involves a minimum of information concerning the λ_i, but is inconvenient to compute. We can find a simpler lower bound and an alternative upper bound, as follows.

Consider the sequence $\{\epsilon_m\}_\pm$ defined by

$$\epsilon_{m\pm} = \frac{[a_{2m+1} - \beta_{p\pm1}a_{2m}]}{(\beta_p)^{2m}(\beta_p - \beta_{p\pm1})}. \quad (2.18)$$

From (2.6) this can be written in the form

$$\epsilon_{m\pm} = C_p^2 + \sum_n{}' \left(\frac{\beta_n}{\beta_p}\right)^{2m} \frac{(\beta_n - \beta_{p\pm1})}{(\beta_p - \beta_{p\pm1})} C_n^2. \quad (2.18a)$$

Now in the summand we have, for all n, $|\beta_n| < |\beta_{p\pm1}|$ so that the numerator is positive (negative) according as $\beta_{p\pm1}$ is positive (negative).

Moreover, if

$\alpha_p > \lambda_p$, we have $\beta_p - \beta_{p\pm1} > 0$, $\beta_{p+1} < 0$, $\beta_{p-1} > 0$

while if

$\alpha_p < \lambda_p$, we have $\beta_p - \beta_{p\pm1} < 0$, $\beta_{p-1} > 0$, $\beta_{p+1} < 0$. (2.19)

We therefore have the inequalities

$$\epsilon_{m+} \geq C_p^2 \geq \epsilon_{m-}, \qquad \alpha_p > \lambda_p, \quad (2.19a)$$
$$\epsilon_{m-} \geq C_p^2 \geq \epsilon_{m+}, \qquad \alpha_p < \lambda_p. \quad (2.19b)$$

Moreover, in each case

$$\epsilon_{m+} \to \epsilon_{m-} \to C_p^2, \qquad m \to \infty. \quad (2.20)$$

The bounds (2.19), but not (2.20) are maintained if we have appropriate bounds on the λ_n, and hence the β_n; for an upper bound, we require lower bounds on the denominators in $\epsilon_{m\pm}$, that is on β_p and $\beta_p - \beta_{p\pm1}$, while we require an upper bound on the numerator, that is, a lower bound on $\beta_{p\pm1}$. For lower bounds on C_p^2, these bounds must be reversed.

C. Simple Upper and Lower Bounds on C_p^2

The bounds of the previous paragraph have the grave disadvantage that they involve the solution of Eqs. (2.3). In most cases this will be impossible. We can get bounds on C_p^2 which do not involve (2.3); the accuracy of these bounds is limited, and de-

pends upon the function g, but in many cases may be sufficient.

Upper Bound

The upper bound in question is trivial; we have, for any p

$$C_p^2 \leq \sum C_p^2 = g^+\sigma g. \tag{2.21}$$

Lower Bounds

We give first a lower bound on C_1^2. Consider the quantity

$$\frac{\lambda_2 g^+\sigma g - g^+Hg}{\lambda_2 - \lambda_1} = C_1^2 + \sum_3^\infty \frac{(\lambda_2 - \lambda_n)C_n^2}{\lambda_2 - \lambda_1} < C_1^2. \tag{2.22}$$

We thus have an immediate lower bound if λ_2 and λ_1 are known. Further, if λ_{1T}, λ_{2T} are such that

$$\lambda_{2T} - \lambda_{1T} \geq \lambda_2 - \lambda_1, \qquad \lambda_{2T} \leq \lambda_2,$$

we have

$$\frac{\lambda_{2T} g^+\sigma g - g^+Hg}{\lambda_{2T} - \lambda_{1T}} < C_1^2. \tag{2.22a}$$

Equation (2.22a) is due originally to Eckart.[3]

For arbitrary C_p, $p > 1$, we can construct a bound as follows:

$$\frac{g^+(H - \lambda_{p-1}\sigma)\sigma^{-1}(H - \lambda_{p+1}\sigma)g}{(\lambda_{p-1} - \lambda_p)(\lambda_{p+1} - \lambda_p)}$$

$$= C_p^2 + \sum{}' C_n^2 \frac{(\lambda_{p-1} - \lambda_n)(\lambda_{p+1} - \lambda_n)}{(\lambda_{p-1} - \lambda_p)(\lambda_{p+1} - \lambda_p)} < C_p^2, \tag{2.23}$$

where the inequality follows since every term in the sum is negative or zero.

This bound is again maintained if we have lower bounds on λ_{p-1} and λ_{p+1}, and an upper bound on the denominator.

The bounds of this paragraph are clearly of limited accuracy for a given function g. They can be expected to give close bounds only if the coefficient C_p is much larger than all the others. There is one situation in which this will (hopefully) be the case; and this is when g is itself a *trial function*[4] for U_p, the eigenfunction belonging to λ_p. In this case these bounds (2.21)–(2.23) are variational bounds, in the sense that if

$$g = AU_n + \epsilon$$

then

$$[C_p^2]_{\text{bound}} = C_p^2 + O(\epsilon^2), \tag{2.24}$$

and equality is reached when $g = AU_n$. Thus the bounds give one way of choosing a good trial function: we choose the form of g to maximize C_p^2 for a given normalization. In the next paragraph we further utilize these crude bounds to provide a variational lower bound on C_p^2 for an arbitrary function g.

D. Variational Lower Bound on C_p^2

Suppose that we want a lower bound on $C_p^2 = (g^+\sigma u_p)^2$ for some function g, and that we have already calculated an upper bound C_t^2 on C_p^2, from, for instance, Eq. (2.1):

$$C_t^2 \geq C_p^2.$$

Suppose further that we have a trial function U_t for U_p, with overlap integral a_p:

$$a_p = U_t^+\sigma U_p,$$

and that we have calculated a lower bound a_t^2 on a_p^2 from, for instance, Eq. (2.23):

$$a_t^2 \leq a_p^2. \tag{2.25b}$$

Since we assume all our Fourier coefficients to be real and positive, we have

$$C_t \geq C_p, \qquad a_t < a_p. \tag{2.25c}$$

We note that this is a restriction on the class of functions g and U_t. While it is always possible to choose the phase of U_p so that either C_p *or* a_p is real and positive, it may not be possible to choose them simultaneously so. However, this restriction is trivial; we can always choose U_n (and therefore U_t) real so that a_p is real; then if g is complex

$$g = g_x + ig_I,$$

we can bound separately the Fourier coefficients of g_x and g_I, supposing only that we know their sign. If the sign is negative, we consider the function $-g$.

Given (2.25c) we can proceed as follows. The function $\bar{g}$,

$$\bar{g} = \alpha U_t - g, \tag{2.26a}$$

has a pth Fourier coefficient d_p

$$d_p = \bar{g}^+\sigma U_p = \alpha a_p - C_p. \tag{2.26b}$$

We choose

$$\alpha = C_t/a_t \geq C_p/a_p, \tag{2.26c}$$

so that

$$d_p > 0. \tag{2.27}$$

We now construct an upper bound d_t^2 on d_p^2 by any

[3] C. Eckart, Phys. Rev. **36**, 878 (1930).

[4] This case has been considered in detail by H. F. Weinberger, J. Res. Natl. Bur. Std. (U. S.) **64B**, 217 (1960).

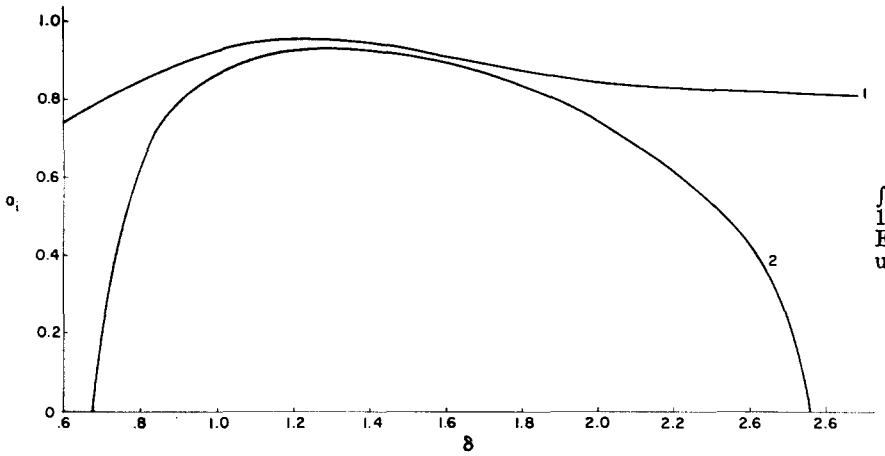


FIG. 1. The first Fourier coefficient $\int g u_0\, dt$ for the function $g = 2\delta^{\frac{1}{2}} x e^{-\delta x}$. 1—exact; 2—lower bound a: from Eq. (2.22). The line $a_i = 1$ gives an upper bound.

of the above methods

$$d_t^2 > d_p^2$$

and therefore

$$d_t > d_p = \alpha a_p - C_p, \tag{2.28}$$

from which we find

$$C_p > \alpha a_p - d_t > C_t - d_t \tag{2.29}$$

which is the required lower bound.

III. AN EXAMPLE

Let us consider the sequence of functions U_n given by the one-sided harmonic oscillator:

$$\left[\frac{d^2}{dx^2} - x^2 + \lambda_n\right] U_n = 0, \tag{3.1}$$

$$U_n(0) = U_n(\infty) = 0, \qquad \int_0^\infty U_n^2\, dx = 1.$$

The eigenvalues λ_n are given by

$$\lambda_n = 2n + 1 \qquad n = 1, 2, 3, \cdots, \tag{3.2}$$

while the lowest eigenfunction $U_1(x)$ is

$$U_1(x) = 2\pi^{-\frac{1}{4}} x e^{-\frac{1}{2}x^2}. \tag{3.3}$$

We define a trial function $U_t(x)$ for U_1 by

$$U_t(x) = 2\delta^{\frac{3}{2}} x e^{-\delta x} \tag{3.4}$$

and search for bounds on the first Fourier coefficient of the function g defined by

$$g(x) = 2\delta^{\frac{3}{2}} x^3 e^{-\delta x}. \tag{3.5}$$

This choice of g is motivated by the application of these bounds in a companion paper.

With the notation

$$\int U_t U_1\, dx = a_1, \qquad \int g U_1\, dx = C_1, \tag{3.6}$$

we can construct bounds on a_1 and C_1 as follows:

(*A*) *A lower bound on* a_1^2. The coefficient a_1^2 cannot exceed the value 1 given by Eq. (2.21).

A lower bound a_t^2 on a_1^2 is given by Eq. (2.22) with $\lambda_2 = 7$, $\lambda_1 = 3$. This bound is plotted as a function of the parameter δ in Fig. 1, together with the exact value which can easily be evaluated in terms of the incomplete gamma function. We see that, as expected, the agreement is poor for extreme values of δ, but rather good for δ such that U_t is expected to approximate U_1 reasonably well. In fact, if we choose δ by the criterion $1 - a_t^2 =$ minimum, we find

$$\delta = 3^{\frac{1}{2}}, \tag{3.7}$$

which is the same value as obtained by minimizing $\int U_t\, HU_t\, dx = \lambda_t$. Indeed if we assume λ_1 and λ_2 are known, this will always be so if we use (2.22).

(*B*) *Upper bounds on* C_1^2. We can calculate a crude upper bound C_{u1}^2 on C_1^2 from Eq. (2.1). We can calculate a variational upper bound C_{uv}^2 on C_1^2 from Eq. (2.1) where f is a trial function satisfying the condition $f(0) = f(\infty) = 0$.

We take

$$f = Ag \tag{3.8}$$

with A a variational parameter. The integrals are easily carried out, as is the evaluation of the exact C_1^2. Minimizing the functional (2.1) with respect to A gives the best bound C_{uv}^2 for a function f of the form (3.8); this bound is plotted as a function of δ in Fig. 2, together with the nonvariational bound C_{u1}^2 and the exact value C_1^2. We see that the variational bound is somewhat better than C_{u1}^2 for all δ.

(*C*) *Lower bounds on* C_1^2. Similarly we can calculate a nonvariational lower bound C_{l1}^2 from (2.22)

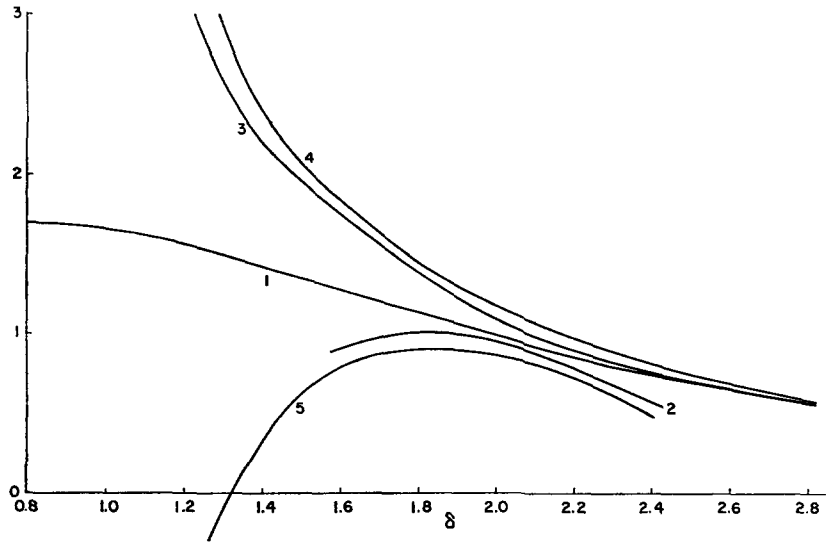


FIG. 2. The first Fourier coefficient $\int g u_0\, dx$ for the function $g = 2\delta^{\frac{1}{2}} x^{\frac{1}{2}} e^{-\delta x}$. 1—exact; 2—lower bound from Eq. (2.22); 3—variational bound from Eq. (3.8); 4—upper bound from Eq. (2.21); 5—variational lower bound from Eq. (3.10).

and a variational lower bound by the technique described by Eqs. (2.26) to (2.29). We define

$$\bar{g} = \alpha U_t - g,$$

where g is given by (3.5), and U_t by (3.4), and for a given δ we define α by

$$\alpha = C_{u_1}/a_1, \tag{3.9}$$

where C_{u_1} is the variational upper bound on C_1 obtained above, and a_1 the lower bound found previously on a_1 and plotted in Fig. 1. With this choice of $\bar{g}$ we can find an upper bound d_t on $\alpha a_1 - C_1$ from (2.21), and hence a lower bound on C_1 from (2.29). With the choice of trial functions f in (2.21),

$$f = AU_t = 2\delta^{\frac{1}{2}} A x e^{-\delta x}, \tag{3.10}$$

we find after minimizing on A the results plotted in Fig. 2. The variational bound C_{l_1} obtained is *worse* than the nonvariational bound C_{l1}, so that clearly (3.10) is a poor trial function.

ACKNOWLEDGMENTS

I am grateful to Professor J. M. Blatt for helpful discussions of this work, and to the referee for bringing the work of Weinberger to my attention.

Upper and Lower Bounds on the Matrix Elements of an Arbitrary Bounded Operator

L. M. DELVES
Applied Mathematics Department, The University of New South Wales, Kensington, N. S. W., Australia
(Received 11 October 1963)

Methods are given for constructing variational upper and lower bounds on the matrix element $W_{mn} = \int U_m^* W U_n d\tau$ of an arbitrary bounded operator W between two eigenstates U_m, U_n of a Hamiltonian H. Numerical examples of the method are given.

I. INTRODUCTION

THIS paper is concerned with the following problem. Given a Hamiltonian H and weight function σ defining a sequence of eigenfunctions U_n,

$$(H - \lambda_n \sigma) U_n = 0, \tag{1.1}$$

for which the eigenfunction cannot be found exactly; and given an arbitrary operator W, construct the best possible approximation to the matrix element $W_{mn} = \langle m| W |n\rangle$. In a recent series of papers[1-3] variation principles for W_{mn} were set up for Hermitian W and their use illustrated.

These principles give an approximation $[W_{mn}]$, accurate to $O(\epsilon^2)$ if approximations to U_n of $O(\epsilon)$ are known. However, they suffer from the defect that they yield no information as to their accuracy, even the sign of the error being unknown. We here give methods of constructing upper and lower bounds on W_{mn}, for bounded but not necessarily Hermitian W. These methods depend on those given in the previous paper[4] for bounding Fourier coefficients; if variational bounds on the appropriate Fourier coefficients are constructed, the bounds upon W_{mn} are also variational, so that the advantages of the previous variational formulations are retained. In addition, a single bound determines the sign of the error, and furthermore ensures that a more complicated trial function gives a more accurate (rather than just a different) result; while with some extra labor, both bounds can be obtained, thus bounding the error.

The results we obtain are applicable to a wide class of operators; for simplicity of printing, we use the notation of finite matrices, so that inner products will be written

$$\langle m |W| n\rangle = U_m^+ W U_n.$$

II. UPPER AND LOWER BOUNDS ON THE DIAGONAL MATRIX ELEMENTS OF A SOLUBLE POSITIVE OPERATOR

To see that the calculation can be reduced to that of bounding a set of Fourier coefficients, we first consider the case of a positive Hermitian operator W, and assume that W is simple enough that all of its eigenfunctions φ_n and eigenvalues μ_n with respect to some positive Hermitian weight function ω are known. That is, we write

$$W\varphi_n = \mu_n \omega \varphi_n, \qquad \varphi_n^+ \omega \varphi_n = 1, \tag{2.1}$$

and assume that the set $\{\varphi_n\}$ and $\{\mu_n\}$ $(\mu_n > 0)$ are known. We look at the element $W_{pp} = U_p^+ W U_p$, and expand U_p in terms of the φ_n:

$$U_p = \sum C_{pn}\varphi_n, \qquad U_p^+ W U_p = \sum \mu_n C_{pn}^2,$$
$$C_{pn} = \varphi_n^+ \omega U_p. \tag{2.2}$$

This expansion is possible since W is Hermitian.

For a positive operator W, all the μ_n are positive, so that given upper (lower) bounds C_{Tn}^2 on C_{pn}^2 we have upper (lower) bounds on $U_p^+ W U_p$:

$$U_p^+ W U_p \gtrless \sum \mu_n C_{Tn}^2 \quad \text{if} \quad C_{Tn}^2 \lesseqgtr C_{pn}^2. \tag{2.3}$$

Now we can derive both upper and lower bounds on C_{pn}^2 using any of the techniques given in Ref. 4, since we can write, using the Hermitian character of σ, ω,

$$C_{pn} = (\sigma^{-1}\omega\varphi_n)^+ \sigma U_p \tag{2.4}$$

which is the pth Fourier coefficient with respect to the set $\{U_n\}$ of the known function

$$g_n = \sigma^{-1}\omega\varphi_n.$$

The most commonly occuring case is $\sigma = \omega = 1$, in which case $g_n = \varphi_n$.

Other Bounded Operators

If the operator W is not positive, but is bounded from below, we can consider instead the operator

$$\bar{W} = W - \alpha\omega, \qquad \alpha < \mu_1, \tag{2.5}$$

[1] L. M. Delves, Nucl. Phys. 41, 497 (1963).
[2] L. M. Delves, Nucl. Phys. 45, 313 (1963).
[3] L. M. Delves, Math. Comp. (to be published).
[4] L. M. Delves, J. Math. Phys. 5, 1055 (1964).

where the (negative) α is a lower bound on the lowest eigenvalue of W. Then $\bar{W}$ is a positive operator whose matrix elements are trivially related to those of W:

$$U_m^+ \bar{W} U_n = U_m^+ W U_n + \alpha U_m^+ \omega U_n,$$

and we can therefore bound the diagonal matrix element of $\bar{W}$ as above. To obtain bounds on $U_m^+ W U_n$, we need also bounds on $U_m^+ \omega U_n$. These can also be obtained as above, since ω is assumed positive. Moreover, in many cases it will be trivial to evaluate this term, since if

$$\omega = \sigma,$$

we have

$$U_m^+ \omega U_n = \delta_{mn}. \tag{2.6}$$

III. AN ALTERNATIVE LOWER BOUND ON DIAGONAL ELEMENTS

The methods of the previous paragraph are limited in usefulness by the need to know the eigenfunctions of W. We give here an alternative lower bound which does not require this knowledge. We consider positive operators W; the extension to bounded operators is as before. However, we do not need to assume W is Hermitian.

For any arbitrary function q, we have

$$q^+ W q \geq 0. \tag{3.1}$$

We take

$$q = U_p - U_{pT} \tag{3.1a}$$

where U_{pT} is a normalized trial function for U_p. Then expanding (3.1) we have

$$U_p^+ W U_p \geq U_p^+ W U_{pT} + U_{pT}^+ W U_p - U_{pT}^+ W U_{pT}. \tag{3.2}$$

We therefore have a lower bound on $U_p^+ W U_p$:

$$U_p^+ W U_p \geq C_{1T} + C_{2T} - U_{pT}^+ W U_{pT}, \tag{3.3}$$

where C_{1T}, C_{2T} are lower bounds on the pth Fourier coefficients of the functions

$$g_1 = \sigma^{-1} W U_{pT}, \qquad g_2 = \sigma^{-1} W^+ U_{pT}, \tag{3.4}$$

and

$$C_{1T} \leq g_1^+ \sigma U_p, \qquad C_{2T} \leq g_2^+ \sigma U_p.$$

For Hermitian W we have $C_{1T} = C_{2T}^*$, and both can be chosen real.

IV. AN ALTERNATIVE UPPER BOUND ON THE DIAGONAL ELEMENT

We can also construct an upper bound for $U_p^+ W U_p$ without knowing the eigenvalue of W, if we can find a soluble *comparison operator* X. Suppose that we have an operator X for which $U_p^+ X U_p$ is *known*, or for which an upper bound is known (for instance, X may be a soluble operator so that the method of Sec. II can be used). Suppose further that we have the operator relation

$$X \geq W. \tag{4.1}$$

That is to say, for any function g, $g^+ X g \geq g^+ W g$. This equation is satisfied, for instance, if X commutes with W and if each eigenvalue of X is greater than the corresponding eigenvalue of W.

We can then define the positive operator Y,

$$Y = X - W, \tag{4.2}$$

and find a lower bound Y_2 on the matrix elements of Y by the methods of Sec. III:

$$U_p^+ Y U_p \equiv U_p^+ X U_p - U_p^+ W U_p \geq Y_L. \tag{4.3}$$

Then if an upper bound X_u is known for $U_p^+ X U_p$,

$$U_p^+ X U_p \leq X_u, \tag{4.4}$$

we have at once

$$U_p^+ W U_p \leq U_p^+ X U_p - Y_L \leq X_u - Y_L. \tag{4.5}$$

Of course, it may be difficult to find a suitable comparison operator X. However, the method does *not* depend upon the difference between X and W being small.

V. BOUNDS ON OFF-DIAGONAL ELEMENTS

We can similarly construct bounds on the off-diagonal elements $U_p^+ W U_q$. If the eigenfunctions and eigenvalues of W are known, we can extend the method of Sec. II easily:

$$U_p^+ W U_q = \sum \lambda_n a_n^+ b_n, \tag{5.1}$$

where

$$\begin{aligned} a_n^+ &= U_p^+ \omega \varphi_n = U_p^+ \sigma (\sigma^{-1} \omega \varphi_n), \\ b_n &= \varphi_n^+ \omega U_q = (\varphi_n^+ \omega \sigma^{-1}) \sigma U_q, \end{aligned} \tag{5.2}$$

so that upper (lower) bounds on $U_p^+ W U_q$ can be constructed given upper (lower) bounds on the Fourier coefficients a_n and b_n. Note however that the methods of Ref. 4 bound directly a_n^2 and b_n^2, so that it is necessary to know the relative signs of a_n, b_n. We assume here that they are real; if W and H are real, it is always possible to choose the U_n and φ_n real so that this is the case.

We can also extend the method of Sec. III to off-diagonal elements, provided that we first have bounds on the diagonal elements. Suppose that we

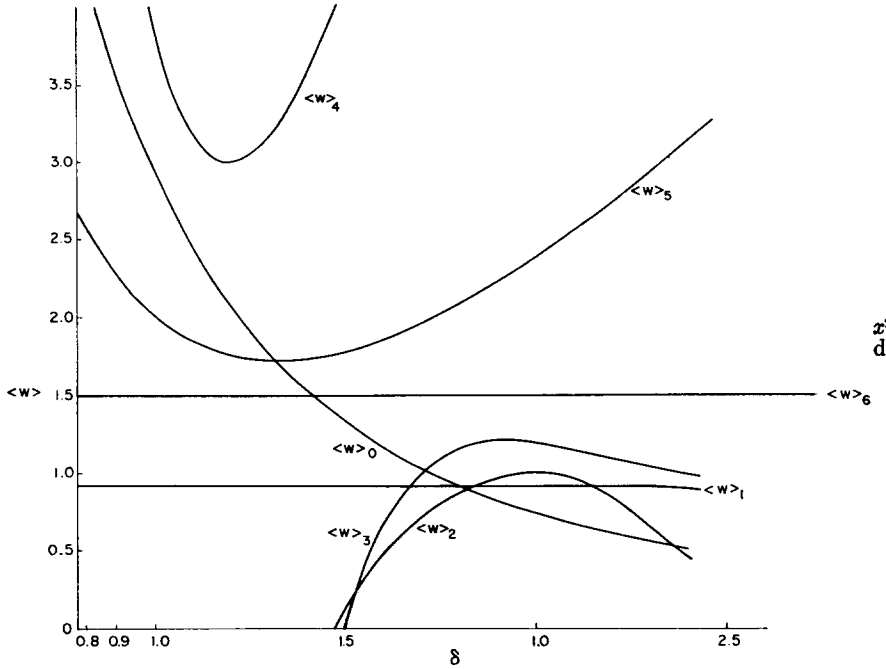


FIG. 1. Estimates of $\langle W\rangle$ for $W = x^2$. The various approximations are defined in Sec. 6.

have trial functions U_{pT}, U_{qT} for U_p, U_q; then, if we define

$$g = U_p + U_q - U_{pT} - U_{qT}, \tag{5.3}$$

the relation (3.1) gives for Hermitian W

$$\begin{aligned} U_p^+WU_q \geq\ & -\tfrac{1}{2}[U_p^+WU_p + U_q^+WU_q] \\ & + U_p^+W(U_{mT} + U_{nT}) + U_q^+W(U_{mT} + U_{nT}) \\ & - \tfrac{1}{2}[(U_{pT} + U_{qT})^+W(U_{pT} + U_{qT})]. \end{aligned} \tag{5.4}$$

Hence we obtain a lower bound on $U_p^+WU_q$ if we have *upper* bounds on $U_q^+WU_p$ and $U_q^+WU_q$ and *lower* bounds on the Fourier coefficients $U_p^+W(U_{mT} + U_{nT})$ and $U_q^+W(U_{mT} + U_{nT})$.

Finally, we can obtain *upper* bounds on $U_p^+WU_q$ if we have a soluble comparison operator V. If $X = V - W$ is a positive operator we can use the method indicated by (5.4) to obtain a bound X_{Lpq} on $U_p^+XU_q$:

$$U_p^+XU_q \equiv U_p^+VU_q - U_p^+WU_q \geq X_{Lpq} \tag{5.5}$$

and if we have an upper bound V_{upq} on $U_p^+VU_q$, then

$$V_{upq} > U_p^+VU_q, \quad U_p^+WU_q \leq V_{upq} - X_{Lpq}. \tag{5.6}$$

VI. A NUMERICAL EXAMPLE

As an example of these methods we consider the one-sided harmonic oscillator described by

$$HU_n(x) = (-d^2/dx^2 + x^2)U_n(x) = \lambda_nU_n(x), \tag{6.1}$$

$$U_n(0) = U_n(\infty) = 0,$$

for which $\lambda_1 = 3$, $\lambda_2 = 7$, $\lambda_n = 4n - 1$, and ask for bounds on the expectation value of the potential energy x^2. This problem was considered in Ref. 4, and variational estimates given for $U_1^+x^2U_1 \equiv \langle W\rangle$; however, no estimate of the error could be made there. The variational estimate was

$$\langle W\rangle_1 = 0.9204 \tag{6.2a}$$

while the exact value is

$$\langle W\rangle_E = 1.5. \tag{6.2b}$$

In addition, a zeroth estimate W_0 was defined in Ref. 1. This is obtained by choosing a trial function U_T for φ_1

$$U_T = 2\gamma^{\frac{3}{2}}xe^{-\gamma x} \tag{6.3}$$

and varying the parameter γ to give the best bound on λ_1. We find

$$\lambda_1 < \langle H\rangle_T = \gamma^2 + 3/\gamma^2$$

giving the minimum $\lambda_T = 2\sqrt{3}$, $\gamma_T = \sqrt{3}$, and

$$\langle W\rangle_0 \equiv U_T^+x^2U_T = 1.732. \tag{6.2c}$$

Neither $\langle W\rangle_0$ nor $\langle W\rangle_1$ will in general give bounds upon $\langle W\rangle$; so that even the sign of the error is in principle not known. We can get upper and lower bounds on $\langle W\rangle$ by using the techniques of this paper, as follows.

Lower Bounds

For this choice of W the lower bound (3.3) reads

$$\langle W\rangle = U_1^+xU_1 \geq 2C_T - U_T^+x^2U_T \tag{6.4}$$

where C_T satisfies the inequality

$$C_T \leq U_1^+ g, \qquad g = x^2 U_T. \tag{6.5}$$

With the choice (6.3) for U_T, the function g is that considered in the previous paper (Ref. 4); and this paper gives two lower bounds C_{LV} and C_{L1}, (Fig. 1, Ref. 4) which satisfy (6.5). For a given value of the variational parameter γ, these bounds define through (6.4) lower bounds $\langle W\rangle_2$ and $\langle W\rangle_3$ on $\langle W\rangle$, which are plotted in Fig. (1), together with the zeroth approximation $\langle W\rangle_0$.

Upper Bounds

For any Hamiltonian H, a suitable comparison potential for the potential V is H itself, since

$$H = T + V > V$$

satisfies Eq. (4.1); and moreover we can easily find upper and lower bounds on $\langle H\rangle$:

$$\langle H\rangle \leq U_T^+ H U_T = \langle H\rangle_T, \tag{6.6a}$$

$$\langle H\rangle \geq \langle H\rangle_T - \frac{U_T^+ H^2 U_T - \langle H\rangle^2}{\lambda_2 - \langle H\rangle}. \tag{6.6b}$$

The lower bound (6.6b) is that due to Temple.[5] We can therefore use the procedures of Sec. 4 to find upper bounds on $\langle W\rangle$: we set

$$W = x^2, \qquad X = H = -d^2/dx^2 + x^2,$$
$$Y = X - W = -d^2/dx^2;$$

and then a lower bound on $\langle Y\rangle$ leads to an upper bound on $\langle W\rangle$. For the simple harmonic oscillator we can find such a bound quite readily, since we have from the virial theorem

$$\langle T\rangle = \langle V\rangle = \tfrac{1}{2}\langle H\rangle.$$

We can therefore use the trial function (6.3) and Eq. (6.6b) to give a lower bound on $\langle Y\rangle$, and hence an upper bound $\langle W\rangle_4$ on $\langle W\rangle$. Moreover, we have the alternative upper bound:

$$\langle W\rangle \leq \langle W\rangle_5 = \tfrac{1}{2}\langle H\rangle_T. \tag{6.7}$$

The upper bounds are plotted as a function of γ in Fig. 1. It is seen that $\langle W\rangle_5$ is a much better bound than $\langle W\rangle_4$. This is to be expected, as $\langle W\rangle_4$, in common with all the methods given in this paper, involves the expectation value of H^2, which is much more sensitive to the form of the trial function than $\langle H\rangle$.

Summary

Taking the best upper and lower bounds, we find

$$1.22 \leq \langle x^2\rangle \leq 1.732. \tag{6.8}$$

DISCUSSION

We have given here methods of finding variational upper and lower bounds on the matrix elements of an arbitrary (bounded) operator W. The methods seem practicable, in the sense that the amount of work involved is not prohibitive. However it is necessary to evaluate $\langle H^2\rangle$, so that the work is much greater than that needed for the variational estimates given previously,[1-3] and the results are much more sensitive to the trial functions used.

[5] G. Temple Proc. Roy. Soc. (London) A119, 276 (1928).

Many-Channel Bargmann Potentials*

JOSEPH R. COX
Department of Physics, University of Miami, Coral Gables, Florida
(Received 17 January 1964)

A generalization of the Bargmann potentials to include the case of nonrelativistics-wave scattering of a particle by a target having a finite number of discrete excited states is presented. This generalization allows the explicit construction of a large class of many-channel S-matrices meromorphic on their energy Riemann surfaces as well as the explicit construction of the corresponding potential matrices. A two-channel example is treated in detail.

1. INTRODUCTION

THE Bargmann potentials[1-4] furnish instructive nonrelativistic models of two-particle elastic scattering processes. For any one particular partial wave these potentials yield both an explicitly solvable Schrödinger equation and an S matrix which is a rational function of momentum. Thus one can often construct a potential which gives rise to the desired number and energies of bound states and resonances simply by making the appropriate choice of poles and zeros of the S matrix.

The main purpose of this paper is to present a generalization of the Bargmann potentials which includes the case of inelastic as well as elastic scattering. For the many-channel problem of s-wave scattering of a particle by a target having a finite number of discrete excited states, those solutions of the coupled Schrödinger equations which result in S matrices meromorphic in energy are given explicitly in terms of a set of descriptive parameters. The corresponding potential matrices which describe the interparticle forces are also given in terms of the same parameters. The relationship between these parameters, the S matrix, and the bound states of the system is discussed.

A short account is given in Sec. 2 of those features of the many-channel formalism introduced by Newton[5] which will be needed in the sections which follow. The notation is essentially the same as that used in footnote 6. In Sec. 3 the generation, according to LeCouteur and Newton, of all the elements of the S-matrix from the Fredholm determinant of the integral equation for the physical wavefunction is described. Then the statements[6] concerning the relation of the zeros of the Fredholm determinant to the bound states and resonances of the system when the channel momenta are considered as independent variables are converted to the corresponding statements which hold when the energy conservation relation between the channel momenta is imposed. In Sec. 4, with the aid of the contents of Secs. 2 and 3, those results concerning the many-channel Bargmann potentials mentioned above in the preceding paragraph are obtained. Finally, in Sec. 5, a two-channel example is considered.

2. PRELIMINARIES

The basic equation of interest in this paper is the s-wave Schrödinger equation[6]

$$-\psi''(K,r) + V(r)\psi(K,r) = K^2\psi(K,r) \qquad (2.1)$$

for the $n \times n$ matrix $\psi(K, r)$, where n is the number of channels. Each column of $\psi(K, r)$ is itself a solution of (2.1) and is distinguished from the other column solutions by its boundary condition. $V(r)$ is the $n \times n$ symmetric potential matrix. It will be restricted here to depend only on r. K in (2.1) is the diagonal $n \times n$ matrix of the channel wavenumbers $k_1, \cdots, k_n$:

$$K_{ij} = k_i\delta_{ij}, \qquad i, j = 1, \cdots, n. \qquad (2.2)$$

Conservation of energy $\hbar^2k_1^2/2\mu$ of the system requires that

$$k_1^2 = k_j^2 + \Delta_j^2, \qquad j = 1, \cdots, n, \qquad (2.3)$$

where μ is the (common) channel reduced mass, and $\hbar^2\Delta_j^2/2\mu$, $\Delta_j \geq 0$, is the threshold energy of the jth channel, i.e., it is the energy difference between the ground state and the $(j-1)$th excited state of the target.

Provided only that all the elements of the potential matrix $V(r)$ have finite first absolute moments, there exists,[6] at least for real energies above the highest threshold, an $n \times n$ matrix solution $F(-K, r)$ of (2.1) defined by

$$\lim_{r\to\infty} e^{-iKr}F(-K,r) = 1, \qquad (2.4)$$

* Based on part of a dissertation submitted in 1962 to Indiana University in partial fulfillment of the requirements for the Ph.D. degree in physics.

[1] V. Bargmann, Rev. Mod. Phys. 21, 488 (1949).
[2] W. R. Theis, Z. Naturforsch. 11a, 889 (1956).
[3] R. G. Newton, J. Math. Phys. 1, 345 (1960).
[4] L. D. Faddeyev (transl. from the Russian by B. Seckler) J. Math. Phys. 4, 72 (1963).
[5] R. G. Newton, Ann. Phys. 4, 29 (1958).
[6] R. G. Newton, J. Math. Phys. 2, 188 (1961).

and, for all energies, an $n \times n$ matrix solution $\phi(K, r)$ of (2.1) defined by

$$\phi(K, 0) = 0, \qquad \phi'(K, 0) = 1. \tag{2.5}$$

$\phi(K, r)$, when expressed in terms of $F(K, r)$ and $F(-K, r)$, is[7]

$$\phi(K, r) = (\tfrac{1}{2}i)[F(K, r)K^{-1}F^{T}(-K) - F(-K, r)K^{-1}F^{T}(K)], \tag{2.6}$$

where $F(K)$, the generalized Jost matrix function,[5] is related to $F(K, r)$ by

$$F(-K) = F(-K, 0). \tag{2.7}$$

The physical wavefunction $\psi(K, r)$ which solves (2.1) may be written in terms of $\phi(K, r)$ and $F(K)$:

$$\psi(K, r) = \phi(K, r)F^{T^{-1}}(-K)K. \tag{2.8}$$

The resulting S matrix is[8]

$$S(K) = K^{\frac{1}{2}}F^{-1}(-K)F(K)K^{-\frac{1}{2}}. \tag{2.9}$$

The S matrix (2.9) is symmetric.[5] This is the expression of the reciprocity theorem, and it is a consequence of the assumption of a symmetric potential matrix $V(r)$. Also, the open-channel submatrix of the S matrix (2.9) is unitary.[5] This is the expression of conservation of current in the open channels.

3. S-MATRIX AND FREDHOLM DETERMINANT

An interesting connection between the elements of the S matrix (2.9) and the determinant $f(K) \equiv f(k_1, \cdots, k_n)$ of the generalized Jost matrix function (2.7) is[6,9]

$$S_{\alpha\alpha} = f(k_1, \cdots, -k_\alpha, \cdots)/f(K), \tag{3.1}$$

$$S^2_{\alpha\beta} = S_{\alpha\alpha}S_{\beta\beta} - f(k_1, \cdots, -k_\alpha, -k_\beta, \cdots)/f(K), \tag{3.2}$$

where $\alpha \neq \beta$ and $\alpha, \beta = 1, \cdots, n$. If, for example, $f(K)$ is known in analytic form, then the whole S matrix may be constructed from it by using (3.1) and (3.2). In this case the study of the S matrix is thereby reduced to the study of the single function $f(K)$.

$f(K)$ also has several other important properties. Newton[6] has shown that if the second as well as the first absolute moments of all elements of $V(r)$ exist, then

$$f(K) \equiv \det F(-K) \tag{3.3}$$

is also the Fredholm determinant of the coupled integral equations for the physical wavefunction (2.8), and that consequently it is a regular analytic function in the whole upper half of the complex plane of each channel momentum $k_1, \cdots, k_n$ when all the k's are treated as independent variables. Let us now deduce from this analyticity property of $f(K)$ and from other results[6] concerning the zeros of $f(K)$ the corresponding statements which hold when the energy-conservation relation (2.3) between the k's is imposed.

Eliminating $k_2, \cdots, k_n$ in favor of k_1 from $f(K)$ by means of (2.3) and considering k_1 a complex number, we are led to associate with $f(K)$ a k_1 Riemann surface consisting of 2^n half-planes, each distinguished from the others by its particular combination of signs of the imaginary parts of all the nk's, and having branch points at $k_1 = \pm\Delta_2, \cdots, \pm\Delta_n$ which for the positive (upper) sign correspond to the threshold energies of the second through nth channel, respectively.

On this k_1 Riemann surface, then, $f(K)$ is a regular analytic function in the entire half-plane characterized by Im $k_1, \cdots, k_n > 0$. It can have no zeros in this half-plane except possibly on the imaginary axis. Zeros on the imaginary axis give rise to the conventional "all channels closed" bound states of the system. Furthermore, $f(K)$ cannot have any zeros on the real k_1 axis bounding the half-plane Im $k_1, \cdots, k_n > 0$ if all channels are open ($k_1 > \Delta_n$) or if $k_1 < -\Delta_n$, but it can have zeros there in pairs symmetric about the origin and possibly at the origin if $i < n$ channels are open and the remainder are closed ($0 < k_1 < \Delta_n$). Of these zeros of $f(K)$ the ones which lie on the positive real k_1 axis give rise to bound states "embedded in the continuum," which have the property that if the forces which produce them are altered slightly, then any such zero in the interval $\Delta_i < k_1 < \Delta_{i+1}$ will in general move downward onto that half-plane characterized by Im $k_1, \cdots, k_i < 0$, Im $k_{i+1}, \cdots, k_n > 0$ and cause a resonance.

4. MANY-CHANNEL BARGMANN POTENTIALS

Let q be any positive integer. Let B be a constant diagonal matrix of order nq with elements

$$B_{pp'} = \delta_{pp'}b_p, \qquad p, p' = 1, \cdots, q, \tag{4.1}$$

where the b_p are diagonal matrices of order n whose elements $b_p^{(i)}\ \delta_{ii}$ are connected by

$$(b_p^{(1)})^2 = (b_p^{(i)})^2 - \Delta_i^2, \qquad j = 1, \cdots, n, \tag{4.2}$$

where Δ_i is the same as in (2.3). If any of the $b_p^{(i)}$'s

[7] The matrix transpose is indicated by the superscript "T."

[8] For the relation of the S matrix to the cross section, see footnote 5.

[9] K. J. LeCouteur, Proc. Roy. Soc. (London) **A256**, 115 (1960).

are complex numbers, then their complex conjugates must also be included in B. Always we require

$$\operatorname{Re} b_p^{(i)} > 0. \tag{4.3}$$

Let A be a constant nonsingular $nq \times nq$ matrix with elements

$$A_{pp'} = \delta_{pp'}a_p, \tag{4.4}$$

where the a_p are real symmetric $n \times n$ matrices. Furthermore, let the elements of A and B also be such that the $nq \times nq$ matrix $Y(r)$ with elements

$$Y_{pp'}(r) = 1\delta_{pp'} - (b_p + b_{p'})^{-1}e^{-(b_p+b_{p'})r}a_{p'}, \tag{4.5}$$

where 1 is the $n \times n$ unit matrix, is nonsingular for all $r \geq 0$. The elements of A and B are subject to no other restrictions than those imposed above. Finally, let $E(r)$ be an $nq \times n$ matrix with elements

$$E_p(r) = e^{-b_p r}. \tag{4.6}$$

The main result of this paper is that the solution of (2.1) satisfying (2.4) with

$$V(r) = 2E^{\mathrm{T}}(r)Y^{\mathrm{T}-1}(r)(AB + BA)Y^{-1}(r)E(r) \tag{4.7}$$

is

$$F(-K, r) = [1 + iE^{\mathrm{T}}(r)AY^{-1}(r)(1K + iB)^{-1}E(r)]e^{iKr}. \tag{4.8}$$

From (4.8) it is a simple matter to compute both the physical wavefunction with the aid of (2.6)–(2.8), and the S matrix with the aid of (2.7) and (2.9) or (3.1)–(3.3). Thus, when $V(r)$ is given by (4.7), all quantities of physical interest are obtainable in closed form. Furthermore, as was stated in Sec. 3, the zeros of the $f(K)$'s so obtained give directly the bound states and include the resonances.

A direct proof that (4.8) satisfies (2.1) with (4.7) will now be given. First it is necessary to establish that

$$B^2A - AB^2 = AK^2Y(r) - Y^{\mathrm{T}}(r)K^2A. \tag{4.9}$$

This is done by making explicit use of (2.3). From (4.1) and (4.2) it is apparent that

$$B^2_{pp'} = [(1b_{p'}^{(1)\prime})^2 + D^2]\delta_{pp'},$$

where

$$D_{ij} \equiv \Delta_i\delta_{ij}, \qquad i, j = 1, \cdots, n.$$

Thus it follows with the aid of (4.4) that

$$(B^2A - AB^2)_{rs} = \delta_{rs}(D^2a_s - a_sD^2), \qquad r, s = 1, \cdots, q. \tag{4.10}$$

On the other hand, (2.3) may be used to obtain

$$(1K^2)_{pp'} = (1k_1^2 - D^2)\delta_{pp'}, \qquad p, p' = 1, \cdots, q,$$

where K is given by (2.2). It then follows with the aid of (4.4) and (4.5) that

$$(AK^2Y(r) - Y^{\mathrm{T}}(r)K^2A)_{rs} = \delta_{rs}(D^2a_s - a_sD^2), \qquad r, s = 1, \cdots, q. \tag{4.11}$$

(4.9) follows immediately from (4.10) and (4.11).

Next we write (4.8) in the form

$$F(-K, r) = (1 + W)e^{iKr},$$

where

$$W \equiv iE^{\mathrm{T}}(r)AY^{-1}(r)(1K + iB)^{-1}E(r), \tag{4.12}$$

and upon inserting it into (2.1) we find

$$V(r) = M(1 + W)^{-1}, \tag{4.13}$$

where

$$M \equiv K^2W - WK^2 + 2iW'K + W''. \tag{4.14}$$

The remainder of the proof consists in showing that the right side of (4.13) is identical with (4.7). Hence, although it appears to be K-dependent, it is not.

Differentiating (4.12) with respect to r, we obtain

$$W' = iE^{\mathrm{T}}AY^{-1}(1K + iB)^{-1}E' + iE^{\mathrm{T}\prime}AY^{-1}(1K + iB)^{-1}E - iE^{\mathrm{T}}AY^{-1}Y'Y^{-1}(1K + iB)^{-1}E. \tag{4.15}$$

(4.1), (4.4), and (4.5) imply

$$Y' = B(1 - Y) + (1 - Y)A^{-1}BA, \tag{4.16}$$

and (4.1), (4.6) imply

$$E' = -BE. \tag{4.17}$$

Using (4.16), (4.17) and remembering that $B = B^{\mathrm{T}}$, we obtain from (4.15)

$$W' = -iE^{\mathrm{T}}AY^{-1}(B + A^{-1}BA)Y^{-1} \times (1K + iB)^{-1}E. \tag{4.18}$$

Differentiating (4.18) and again using (4.16), (4.17) yields

$$W'' = iE^{\mathrm{T}}AY^{-1}[2(B + A^{-1}BA)Y^{-1} \times (B + A^{-1}BA) - B(B + A^{-1}BA) - (B + A^{-1}BA)A^{-1}BA]Y^{-1}(1K + iB)^{-1}E. \tag{4.19}$$

Using (4.4), (4.5) and remembering that $A = A^{\mathrm{T}}$, we obtain

$$YA^{-1} = A^{-1}Y^{\mathrm{T}}, \tag{4.20}$$

and using (4.6) in addition we obtain

$$Y' = EE^{\mathrm{T}}A. \tag{4.21}$$

We now use (4.12), (4.16), (4.17), (4.20), and (4.21) to write (4.7) as

$$V(r)(1 + W) = 2iE^{T}AY^{-1}(B + A^{-1}BA) \times [-iY^{-1}(1K + iB) - Y^{-1}B - A^{-1}BAY^{-1} + Y^{-1}(B + A^{-1}BA)Y^{-1}](1K + iB)^{-1}E. \quad (4.22)$$

On the other hand, use of (4.12), (4.14), (4.18), and (4.19) yields

$$M = iK^{2}E^{T}AY^{-1} \times (1K + iB)^{-1}E - iE^{T}AY^{-1}(1K + iB)^{-1}EK^{2} + 2E^{T}AY^{-1}(B + A^{-1}BA)Y^{-1}(1K + iB)^{-1}KE + iE^{T}AY^{-1}[2(B + A^{-1}BA)Y^{-1} \times (B + A^{-1}BA) - B(B + A^{-1}BA) - (B + A^{-1}BA)A^{-1}BA]Y^{-1}(1K + iB)^{-1}E, \quad (4.23)$$

where the fact that K commutes with E has been used. (4.13) is true if and only if the right sides of (4.22) and (4.23) are equal. Direct comparison and use of (4.20) reveals that they are equal if

$$E^{T}Y^{T-1}[(B^{2}A - AB^{2}) - (AK^{2}Y - Y^{T}K^{2}A)]Y^{-1} \times (1K + iB)^{-1}E \equiv 0,$$

which, according to (4.9), is indeed the case.

We have thus shown that (4.8) solves (2.1) when the potential matrix is (4.7). That (4.8) also satisfies (2.4) follows easily from (4.1)–(4.6).

The generalized Jost matrix function obtained from (4.8) by using (2.7) is

$$F(-K) = 1 + i\sum_{p=1}^{q} N_{p}(K + ib_{p})^{-1}, \quad (4.24)$$

where the $n \times n$ matrix N_{p} is defined as

$$N_{p} = \sum_{p'=1}^{q} [AY^{-1}(0)]_{p'p}. \quad (4.25)$$

For $n = 1$, (4.24) becomes a rational function of the single variable k_1, and, according to (2.9), so does the S matrix. The potentials (4.7) thus reduce for $n = 1$ to the single-channel Bargmann potentials,[1-4] and in this sense they represent for $n > 1$ a generalization of such potentials to the many-channel case.

Since all elements of (4.7) clearly[10] have finite first and second absolute moments, it follows from the discussion of Sec. 3 that the Fredholm determinant $f(K)$ resulting from (4.24) upon use of (3.3) can have no poles in that half-plane of its k_1 Riemann surface characterized by Im $k_1, \cdots, k_n > 0$. In order to see how this comes about we note from (4.24) that all poles of $f(K)$ arise as the result of zeros of a typical factor of the form

$$\det(K + ib) = \prod_{i=1}^{n}(k_i + ib^{(i)}),$$

which by (4.3) can never be zero unless Im $k_i < 0$ for at least one k_i. Thus, Im $k_1, \cdots, k_n > 0$ is precisely that one of the 2^n half-planes on which no pole can occur. Furthermore, since the $q^{(i)}$'s are connected by (4.2), it follows that unless the numerator of $f(K)$ has one or more zeros coincident with the zeros of $\det(K + ib)$, $f(K)$ has a pole in all half-planes except Im $k_1, \cdots, k_n > 0$ if it has a pole in one of them, and the multiplicity of the pole in a particular half-plane equals the number of k_i's whose imaginary parts are negative there. Counting each pole according to its multiplicity, then, $f(K)$ must have an integer multiple of $\frac{1}{2}n2^n$ poles unless its numerator contains one or more coincident zeros. When $n = 2$, for example, the zeros of the term $\det(K + ib)$ will in general give rise to a single pole at $k_1 = -ib^{(1)}$, $k_2 = ib^{(2)}$ on Im $k_1 < 0$, Im $k_2 > 0$ and at $k_1 = ib^{(1)}$, $k_2 = -ib^{(2)}$ on Im $k_1 > 0$, Im $k_2 < 0$, a double pole at $k_1 = -ib^{(1)}$, $k_2 = -ib^{(2)}$ on Im k_1, $k_2 < 0$, and, of course, no poles on Im k_1, $k_2 > 0$.

It is evident from (3.3) and (4.24) that the zeros of $f(K)$ and hence, as was explained in Sec. 3, the bound states and resonances of the system are to be obtained by finding, on the appropriate Riemann surface, the zeros of the determinant of a qth-degree polynomial in the diagonal matrix K with $n \times n$ matrix coefficients which are formed from the N_{p}'s and b_{p}'s. This is a complicated task in even the simplest of coupled-channel cases, as is indicated in the example of the next section. However, the general two-channel problem may be reduced to the more familiar one of finding the zeros of a polynomial in the single variable t by introducing the transformation $k_1(t) = \Delta_2(1 + t^2)(1 - t^2)^{-1}$, $k_2(t) = 2\Delta_2 t(1 - t^2)^{-1}$, where Δ_2 is the same as in (2.3). This transformation maps the double-sheeted two-channel Riemann surface onto the t plane in a one-to-one manner.[11]

On the basis of what is known about the single-channel Bargmann potentials, the important work of Marchenko concerning the inverse problem of scattering theory,[4] and its partial generalization to the many-channel case,[11] we conjecture that the positive integer q and the matrix parameters A and B

[10] Each element is everywhere bounded in r and vanishes exponentially as $r \to \infty$.

[11] See Ph.D. dissertation of Joseph R. Cox, Indiana University, 1962 (unpublished).

which were introduced at the beginning of this section have the following significance. q is the number of poles of the S matrix $S(K)$ as determined by (3.1)–(3.3) and (4.24) which lie in the half-plane Im $k_1, \cdots, k_n > 0$. The positions of these poles are given by $K = ib_p$, where the b_p's are the diagonal elements of B [see (4.1)]. If $K = ib_p$ corresponds to a bound state of the system, then a_p, the associated diagonal element of A [see (4.4)] may, apart from the restrictions imposed at the beginning of Sec. 4, be chosen arbitrarily without altering the S matrix or the energies of the bound states.[12] If, on the other hand, $K = ib_p$ does not correspond to a bound state of the system, then the associated a_p is i times the residue at $K = ib_p$ of $k_1K^{-\frac{1}{2}}\,S(K)K^{-\frac{1}{2}}$.

5. A TWO-CHANNEL EXAMPLE

The simplest two-channel S matrix of Sec. 4 is obtained by setting $n = 2$ and choosing $q = 1$. According to (4.1)–(4.4) we may then write

$$B = \begin{bmatrix} b_1 & 0 \\ 0 & b_2 \end{bmatrix}, \quad A = \begin{bmatrix} a_1 & a_3 \\ a_3 & a_2 \end{bmatrix}, \quad \text{and} \quad b_1^2 = b_2^2 - \Delta^2,$$

where all elements of A and B are real, $b_1, b_2 > 0$, and $\Delta \equiv \Delta_2$, $b_1, b_2 \equiv b^{(1)}, b^{(2)}$, $a_1, a_2, a_3 \equiv a_{11}, a_{22}, a_{12}$. It is convenient to define $h_1(r)$, $h_2(r)$, $l(r)$ by

$$h_{1,2}(r) = a_{1,2} - (2b_{2,1})^{-1}e^{-2b_{2,1}r} \det A,$$

$$l(r) \equiv \det Y(r) = 1 - a_1(2b_1)^{-1}e^{-2b_1r} - a_2(2b_2)^{-1}e^{-2b_2r} + (4b_1b_2)^{-1}e^{-2(b_1+b_2)r} \det A.$$

As mentioned in Sec. 4, A must be chosen so that $\det A \neq 0$, and A, B so that $\det Y(r) \neq 0$ for all $r \geq 0$.

The generalized Jost function, as obtained with the aid of (4.5), (4.24), and (4.25), is then

$$F(-K) = \begin{bmatrix} (k_1 + i\alpha_1)(k_1 + ib_1)^{-1} & i\beta(k_2 + ib_2)^{-1} \\ i\beta(k_1 + ib_1)^{-1} & (k_2 + i\alpha_2)(k_2 + ib_2)^{-1} \end{bmatrix}, \tag{5.1}$$

where α_1, α_2, β are related to the elements of A and B by

$$\alpha_{1,2} = b_{1,2} + h_{1,2}(0)l^{-1}(0), \qquad \beta = a_3l^{-1}(0).$$

From (3.3) and (5.1) we obtain the Fredholm determinant:

$$f(k_1, k_2) = \frac{(k_1 + i\alpha_1)(k_2 + i\alpha_2) + \beta^2}{(k_1 + ib_1)(k_2 + ib_2)}. \tag{5.2}$$

It is evident from (5.2) that finding the zeros of $f(k_1, k_2)$ in even this "simple" example is already equivalent to solving a fourth-degree algebraic equation in k_1 or k_2. We do not do this here, but instead we merely state that it can be shown indirectly that $f(k_1, k_2)$ as given by (5.2) can have no zeros in Im $k_1, k_2 > 0$.[11] Consequently, there can be no s-wave bound states.

Equations (5.2), (3.1), and (3.2) then give the S matrix:

$$S_{11} = \frac{[(k_1 - i\alpha_1)(k_2 + i\alpha_2) - \beta^2](k_1 + ib_1)}{[(k_1 + i\alpha_1)(k_2 + i\alpha_2) + \beta^2](k_1 - ib_1)},$$

$$S_{22} = \frac{[(k_1 + i\alpha_1)(k_2 - i\alpha_2) - \beta^2](k_2 + ib_2)}{[(k_1 + i\alpha_1)(k_2 + i\alpha_2) + \beta^2](k_2 - ib_2)}, \tag{5.3}$$

$$S_{12} = S_{21} = \frac{(k_1k_2^{-1})^{\frac{1}{2}}\beta(k_2 + ib_2)}{[(k_1 + i\alpha_1)(k_2 + i\alpha_2) + \beta^2](k_1 - ib_1)}.$$

It is apparent from (5.3) that the two channels become uncoupled if a_3 (and hence β) $\to 0$:

$$S_{11} = \frac{(k_1 - i\alpha_1)(k_1 + ib_1)}{(k_1 + i\alpha_1)(k_1 - ib_1)},$$

$$S_{22} = \frac{(k_2 - i\alpha_2)(k_2 + ib_2)}{(k_2 + i\alpha_2)(k_2 - ib_2)}, \tag{5.4}$$

$$S_{12} = S_{21} = 0.$$

S_{11} in (5.4) is in itself a well-known S matrix for the one-channel case; it is the "effective range" S matrix when there are no s-wave bound states.[1] That is, if $S_{11} \equiv e^{2i\delta}$, then

$$k_1 \cot \delta = [\alpha_1b_1/(b_1 - \alpha_1)] + k_1^2(b_1 - \alpha_1)^{-1},$$

$$\alpha_1, b_1 > 0.$$

The real symmetric potential matrix which produces the S matrix (5.3) is, according to (4.5)–(4.7),

$$V(r) = -2\frac{d}{dr}\left[l^{-1}(r)\begin{bmatrix} h_1(r)e^{-2b_1r} & a_3e^{-(b_1+b_2)r} \\ a_3e^{-(b_1+b_2)r} & h_2(r)e^{-2b_2r} \end{bmatrix}\right].$$

The physical wavefunction is easily obtained in the manner described in Sec. 4.

ACKNOWLEDGMENTS

I wish to thank Professor Roger G. Newton for his suggestion of the thesis topic which led to the results contained in this paper and for numerous stimulating discussions.

[12] This freedom of choice of those a_p's corresponding to bound states gives rise to a family of potential matrices all of which have the same S matrix and discrete energy levels, a result which is well-known in the single-channel case.[1]

Existence of Scattering Solutions for the Schrödinger Equation

JOHAN G. BELINFANTE*
Physics Department, University of Pennsylvania, Philadelphia, Pennsylvania
(Received 19 December 1963; final manuscript received 10 March 1964)

The Fredholm alternative for the Schrödinger (Lippmann–Schwinger) equation is established for potentials with finite first moment, using the Ascoli selection theorem and Banach space methods.

1. INTRODUCTION

THE purpose of this paper is to supply a simple proof of the known result that the Fredholm alternative applies to the Lippmann–Schwinger equation for potentials with finite first moment, and thereby follows the existence of solutions of the Lippmann–Schwinger equation and the existence of the resolvent operator for real, positive nonzero energies.

Existence theorems are crucial in the proof of dispersion relations. Khuri[1] established the existence theorem by the use of classical Fredholm theory, while Hunziker[2] has established the same results by the use of Banach space methods. Here we follow the abstract methods used by Hunziker.

The actual application of these methods to the proof of dispersion relations is slightly complicated owing to the necessity of introducing not one, but an infinite set of Banach spaces. Thus in this respect it might appear simpler to follow Weinberg's recent suggestion[3] to use Hilbert space methods for the proof of dispersion relations, in which only one Hilbert space is required. It should perhaps be noted, however, that Hilbert space methods, while suitable for complex energies, fail in the interesting physical limit of real, positive energies, and there is some problem to prove that the physical scattering amplitude exists as the boundary value of the scattering amplitude defined for complex energies.[4]

One of the main charms of the more conservative Banach space method used in the present paper, on the other hand, is that it continues to be applicable even in the physical case of real and positive energies, and allows the use of non-square-integrable wavefunctions, such as plane waves, which are singular from the point of view of the Hilbert space theory.

The existence theory presented below is based on the Banach–Schauder theory, using the Ascoli–Arzela selection theorem to establish the compactness of the kernel GV of the Lippmann–Schwinger equation. We should point out however that ultimately one encounters the same inequalities to be proved, whether one uses the classical Fredholm theory or the abstract Banach space approach. Our Lemma 3 below, for example, is equally useful in the more classical proofs as in the abstract proof. The classical methods have a great advantage from a calculational point of view; the abstract approach, however, does have a definite value in concentrating on certain questions of limits, and especially the distinction between pointwise and uniform limits.

The main technical concern of the present paper is the proof of the compactness of the kernel GV of the Lippmann–Schwinger equation. This was greatly simplified by the discovery of a simple proof of the inequality in Lemma 3, which is the direct analog of Appendix III of the classic paper by Jost and Païs.[5] In their method of proof, elliptic functions had to be used. We find that this is unnecessary and that an elegant proof can be given with elementary methods.

2. MATHEMATICAL PRELIMINARIES

The Banach–Schauder theory establishes the Fredholm alternative for all compact operators in an arbitrary Banach space. In the Banach space of bounded and continuous functions, defined throughout three-dimensional Euclidean space, and provided with the uniform norm, which we consider in the present paper, there is a simple criterion for com-

* Supported in part by the U. S. Atomic Energy Commission.

[1] N. Khuri, Phys. Rev. **107**, 1148 (1957); S. Gasiorowicz and H. P. Noyes, Nuovo Cimento **10**, 78 (1958); T. Regge, Nuovo Cimento **8**, 671 (1958); A. Klein and C. Zemach, Ann. Phys. (N. Y.) **7**, 440 (1959).

[2] W. Hunziker, Helv. Phys. Acta. **34**, 593 (1961).

[3] S. Weinberg, Phys. Rev. **131**, 440 (1963); A. Grossmann and T. T. Wu, J. Math. Phys. **2**, 710 (1961).

[4] The Yukawa potential and the square-well potentials obviously have finite first moments; they also satisfy the square-integrability condition imposed by S. Weinberg (Ref. 3), and it can even be shown that they satisfy the more complicated condition $\int d^3\mathbf{x} \int d^3\mathbf{y}\, |V(\mathbf{x})|\, |V(\mathbf{y})|\, |\mathbf{x}-\mathbf{y}|^{-2} < \infty$, which is required in the Hilbert space theory of A. Grossmann and T. T. Wu (Ref. 3). There is no reason to prefer the requirement of a finite first moment over other conditions except on the basis of simplicity.

[5] R. Jost and A. Païs, Phys. Rev. **82**, 840 (1951).

pactness, namely, the Ascoli–Arzela selection theorem. These facts are our main tools.[6]

Let us recall some elementary definitions and facts. A set is called compact if every open covering of the set contains a finite subcovering. A set is called sequentially compact if every infinite sequence contained in the set contains a convergent subsequence. In a Banach space, and more generally in any metric space, a sequentially compact set is also compact, and conversely.[7]

Let the norm of a vector ψ in our Banach space be denoted $||\psi||$. An operator A (linear transformation) is bounded if the operator norm,

$$||A|| = \sup_{\psi} \frac{||A\psi||}{||\psi||}, \tag{1}$$

exists and is finite. A bounded operator is called compact (or completely continuous) if it maps a set of vectors whose norms are uniformly bounded into a compact set of vectors. The unit operator in an infinite-dimensional Banach space is not compact. In a Hilbert space, a compact operator can be approximated arbitrarily closely in norm by a finite-rank operator, a fact which has been exploited by Weinberg and is the basis for his quasiparticle method.[3] In our case we deal with a nonseparable Banach space, and there is apparently no corresponding result available.[8]

3. CONDITIONS ON THE POTENTIAL AND ON THE WAVEFUNCTIONS

It is interesting to remark that the boundary conditions imposed on the wavefunctions in quantum mechanics play as an important a role as the Schrödinger equation itself. Indeed, the boundary conditions play the role in wave mechanics which is the analog of the Bohr–Sommerfeld quantization rules of the old quantum theory.

The Schrödinger equation in its differential form introduces the unnecessary complications of questions of differentiability, which we here avoid in the usual way by considering the Lippmann–Schwinger integral equation form of the Schrödinger equation. We consider the existence theory of the Lippmann–Schwinger equation ($\hbar = 2m = 1$),

$$\langle \mathbf{x} \mid \mathbf{k}_{\text{in}} \rangle = e^{i\mathbf{k}\cdot\mathbf{x}} - \frac{1}{4\pi} \int d^3\mathbf{y} \, \frac{e^{ik|\mathbf{x}-\mathbf{y}|}}{|\mathbf{x}-\mathbf{y}|} \, V(y) \langle \mathbf{y} \mid \mathbf{k}_{\text{in}} \rangle, \tag{2}$$

for the case of real $\mathbf{k}$ and subject to the physical constraint $k = |\mathbf{k}|$. For the results of the present paper, the generalizations to complex $\mathbf{k}$, k, in fact present a fairly minor problem.[2] The potential $V(\mathbf{y})$ will be taken to be spherically symmetric, but may be discontinuous. We demand always that $y\,|V(y)|$ be Lebesgue-integrable on $[0, \infty)$ and that

$$\int_0^\infty y\, dy\, |V(y)| < \infty. \tag{3}$$

Any potential which satisfies these conditions will, for the sake of brevity, be called a normal potential. It is known[9] that a normal potential cannot support an infinite number of bound states. The Coulomb potential is excluded.

We discuss the existence of bounded and continuous solutions $\langle \mathbf{x} \mid \mathbf{k}_{\text{in}} \rangle$. We introduce the Banach space C of all bounded and continuous functions $\psi(\mathbf{x})$ with the uniform norm

$$||\psi|| = \sup_{\mathbf{x}} |\psi(\mathbf{x})|. \tag{4}$$

The completeness axiom in this Banach space is simply the theorem that the uniform limit of a sequence of bounded continuous functions is again a continuous function.

We may write (2) in an obvious symbolic notation[10] as

$$\psi = \psi_0 + GV\psi. \tag{5}$$

It is immediately clear that a plane wave is an admissible choice for ψ_0 since a plane wave is a bounded and continuous function, with unit norm.[11]

The main theorem to be proved below is the

[6] A simple and useful account of this theory, quite adequate for the present paper, has been given in B. Epstein, *Partial Differential Equations* (McGraw-Hill Book Company, Inc., New York, 1962), Chap. 4.

[7] In Ref. 6, the concept of sequential compactness is used exclusively. The equivalence of this concept with ordinary compactness in the case of a metric space is proved for example in A. E. Taylor, *Introduction to Functional Analysis* (Chapman and Hall, Ltd., London, 1958), pp. 70–71.

[8] I. Maddaus, Bull. Am. Math. Soc. **44**, 279 (1938); R. S. Phillips, Trans. Am. Math. Soc. **48**, 516 (1940).

[9] V. Bargmann, Proc. Natl. Acad. Sci. U. S. **38**, 961 (1952); J. Schwinger, Proc. Natl. Acad. Sci. U. S. **47**, 122 (1961).

[10] We may note that V itself need not even be an operator in our Banach space because $V(y)$ may be a discontinuous function of y, but nevertheless GV is always an operator. The Green's function serves to smooth out any discontinuities the potential may have in a rather remarkable fashion.

[11] The behavior of the potential as $r \to \infty$ plays an essential role in dispersion theory. Roughly, the situation is as follows: if the potential decreases as $e^{-\mu r}$ for large r, then the scattering amplitude has a pole when the momentum transfer $(\mathbf{k} - \mathbf{k})^2$ is equal to $-\mu^2$, and the partial-wave amplitudes have a left-hand cut with branch point at energy k^2 equal to $\frac{1}{4}(-\mu^2)$. If one cuts off the potential after some finite range R, then although the physical scattering is not altered appreciably, the analytic properties change drastically. The left-hand cut in the partial-wave amplitudes then disappears and is replaced by an essential singularity at infinite energy. Presumably a similar phenomenon will take place if one tries to carry through the box normalization technique used in earlier versions of scattering theory. Cf. for example, A. O. Barut and K. H. Ruei, J. Math. Phys. **2**, 181 (1961).

known theorem[12] that the operator GV is a compact operator on the Banach space C. Our purpose of proving this result here again is the hope that the simplifications introduced here will lead these methods to find greater application in quantum mechanics.

Lemma 1. For a normal potential $V(y)$, the operator GV is a bounded operator.

Proof: Let ψ be any function in C. Then

$$|GV\psi(\mathbf{x})| \leq \frac{1}{4\pi} \int d^3\mathbf{y} \, \frac{1}{|\mathbf{x}-\mathbf{y}|} \, |V(y)| \, ||\psi||. \tag{6}$$

Imagine expanding $|\mathbf{x}-\mathbf{y}|^{-1}$ in a Legendre series in $P_l(\hat{x}\cdot\hat{y})$. Then it is clear that an angular integration yields

$$\int d\Omega_{\hat{y}} \, \frac{1}{|\mathbf{x}-\mathbf{y}|} = 4\pi \min\left\{\frac{1}{x}, \frac{1}{y}\right\} \leq \frac{4\pi}{y}. \tag{7}$$

Then (6) becomes

$$||GV\psi|| \leq ||\psi|| \int_0^\infty y \, dy \, |V(y)|. \tag{8}$$

Thus GV is a bounded operator with operator norm, defined by (1), given by

$$||GV|| \leq \int_0^\infty y \, dy \, |V(y)|. \tag{9}$$

Q.E.D.

Some remarks are in order. In order to prove that GV is an operator at all, we need to establish that $(GV\psi)(\mathbf{x})$ is a continuous function as well as being bounded. Later on we, in fact, prove continuity, as well as the even stronger statement of equicontinuity. As a matter of fact, one can even show that GV maps bounded functions into continuous functions. To avoid repetition, but at some sacrifice to the logic, we therefore defer the question of continuity until later. We may also note at this point that (9) gives a sufficient criterion for the convergence of the (Born) perturbation series, $1 + GV + (GV)^2 + \cdots$. The Born series converges if the potential $V(y)$ satisfies

$$\int_0^\infty y \, dy \, |V(y)| < 1. \tag{10}$$

The condition (10) is not a necessary condition for the convergence of the Born series; in fact it is a familiar fact that the Born series converges at high energies even when (10) is violated.[13] We may also note that (10) can only be true if there are no bound states present.[9]

[12] A. Y. Povzner, Mat. Sbornik 32, 109 (1953); T. Ikebe, Arch. Ratl. Mech. Anal. 5, 1 (1960); F. M. Odeh, J. Math. Phys. 2, 794 (1961).

[13] A. Klein and C. Zemach, Nuovo Cimento 10, 1078 (1958).

4. PROOF OF COMPACTNESS OF THE KERNEL GV

The proof of compactness is given after a few simple lemmas are established. In Lemma 2 we approximate a given normal potential V by suitably chosen cut-off potentials V_δ. For a cut-off potential V_δ, the compactness of GV_δ is proved readily. The compactness of GV for all normal potentials then follows by a standard theorem[14] which asserts that the limit (in norm) of a sequence of compact operators is again compact.[15]

Lemma 2. For any normal potential $V(\mathbf{x})$, and any $\delta > 0$, there exist positive numbers M, $R < \infty$ such that the cut-off potential, defined by

$$V_\delta(x) = \begin{cases} V(x) & \text{if } |V(x)| < M \text{ and } x < R, \\ 0 & \text{otherwise,} \end{cases} \tag{11}$$

satisfies

$$||GV - GV_\delta|| < \delta \tag{12}$$

for all k.

Proof: This follows immediately from the requirement that $y \, |V(y)|$ be Lebesgue-integrable, together with the methods used in Lemma 1.

Q.E.D.

In applying the Ascoli–Arzela selection theorem as a criterion for compactness in our Banach space C, we must of course use it in a form suitable for functions defined in the (unbounded) three-dimensional Euclidean space. Sufficient for our purposes is the following statement of this theorem (for proof cf. Ref. 6).

Selection Theorem. If $K \subset C$ is a set of bounded and continuous functions, such that:

(1) there exists a uniform bound M such that, for all $\psi \in K$,

$$||\psi|| < M; \tag{13}$$

(2) the functions $\psi \in K$ are equicontinuous (defined below);

(3) there exists a uniform bound M', and a radius R such that

$$|\psi(\mathbf{x})| < M/|\mathbf{x}| \tag{14}$$

for all $|\mathbf{x}| > R$ and all $\psi \in K$;

then the set K is compact, that is, from every se-

[14] The proof is a simple application of the famous diagonal sequence method; cf. Ref. 6.

[15] We use in the present paper always the norm topology for operator limits.

quence in K one can select a uniformly convergent subsequence.

The equicontinuity condition is defined as follows: The functions $\psi \in K$ are equicontinuous if given any $\epsilon > 0$, $\exists\, \delta > 0$, with δ independent of both $\mathbf{x}$, ψ, such that for all $\psi \in K$ and all $\mathbf{x}$, we have

$$|\psi(\mathbf{x}) - \psi(\mathbf{x}')| < \epsilon \tag{15}$$

whenever $|\mathbf{x}' - \mathbf{x}| < \delta$. The condition (13) will follow from Lemma 1. Condition (14) is easily proved in our case by the use of Lemma 2.[16]

It remains therefore to prove the equicontinuity condition in the case of a cutoff potential. For the proof of equicontinuity we employ the following lemma for which a new and simple proof is given.

Lemma 3. Let Σ denote the sphere $|\mathbf{y}| < R$ and let $\mathbf{x}$, $\mathbf{x}'$ be arbitrary vectors. Then

$$\int_\Sigma d^3\mathbf{y} \left| \frac{1}{|\mathbf{x} - \mathbf{y}|} - \frac{1}{|\mathbf{x}' - \mathbf{y}|} \right| \leq 4\pi R\, |\mathbf{x} - \mathbf{x}'|. \tag{16}$$

Proof: From the identity

$$\frac{1}{|\mathbf{x} - \mathbf{y}|} - \frac{1}{|\mathbf{x}' - \mathbf{y}|} = \int_0^1 d\xi \frac{d}{d\xi} \frac{1}{|\mathbf{x}\xi + \mathbf{x}'(1 - \xi) - \mathbf{y}|} \tag{17}$$

it follows, by carrying out the derivative, that

$$\left| \frac{1}{|\mathbf{x} - \mathbf{y}|} - \frac{1}{|\mathbf{x}' - \mathbf{y}|} \right| \leq |\mathbf{x} - \mathbf{x}'| \int_0^1 d\xi \frac{1}{|\mathbf{x}\xi + \mathbf{x}'(1 - \xi) - \mathbf{y}|^2}. \tag{18}$$

Let us now denote

$$r = |\mathbf{x}\xi + \mathbf{x}'(1 - \xi) - \mathbf{y}|, \tag{19}$$

and define an auxiliary function

$$f(r) = \min\left\{ \frac{1}{r^2}, \frac{1}{R^2} \right\} \leq \frac{1}{R^2}. \tag{20}$$

Then

$$\int_\Sigma d^3\mathbf{y} \frac{1}{r^2} \leq (\tfrac{4}{3}\pi)R + \int_\Sigma d^3\mathbf{y} \left| \frac{1}{r^2} - f(r) \right|. \tag{21}$$

Now let Σ' be the sphere $r < R$ (cf. Fig. 1). Then $r^{-2} - f$ vanishes outside Σ'. The integral on the right of (21) may be extended from $\Sigma \cap \Sigma'$ to all of Σ' without affecting the inequality. But

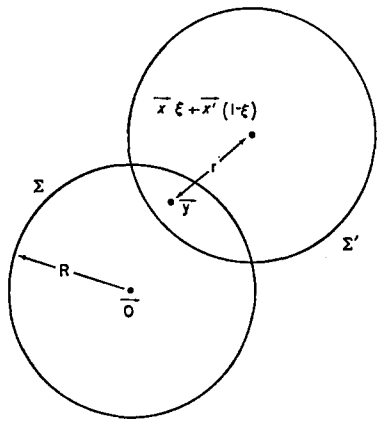


FIG. 1. Auxiliary spheres used in the proof of the inequality,[16] from which equicontinuity follows.

$$\int_{\Sigma'} d^3\mathbf{y} \left| \frac{1}{r^2} - f(r) \right| = 4\pi \int_0^R r^2\, dr \left(\frac{1}{r^2} - \frac{1}{R^2} \right) = (\tfrac{8}{3}\pi)R. \tag{22}$$

From (21) and (22) it then follows that

$$\int_\Sigma d^3\mathbf{y} \frac{1}{r^2} \leq (\tfrac{4}{3}\pi)R + (\tfrac{8}{3}\pi)R = 4\pi R. \tag{23}$$

Q.E.D.

Theorem. For any normal potential $V(\mathbf{x})$, the kernel GV of the Lippmann–Schwinger equation is a compact operator in the Banach space C of bounded and continuous functions.

Proof: It is sufficient to consider the case of a cutoff potential. For a cutoff potential V_δ, it is easily proved by using (7), with x, y interchanged, that

$$|GV_\delta\psi(\mathbf{x})| < MR^3\, ||\psi||/3\, |\mathbf{x}| \tag{24}$$

for all $|\mathbf{x}| > R$, where M and R are defined by Eq. (11). In the proof of equicontinuity there is some algebra involved in removing the exponential factors in the Green's function. One uses the inequality

$$\left| \frac{e^{ik|\mathbf{x}-\mathbf{y}|}}{|\mathbf{x} - \mathbf{y}|} - \frac{e^{ik|\mathbf{x}'-\mathbf{y}|}}{|\mathbf{x}' - \mathbf{y}|} \right| \leq \left| \frac{1}{|\mathbf{x} - \mathbf{y}|} - \frac{1}{|\mathbf{x}' - \mathbf{y}|} \right| + |k| \frac{|\mathbf{x} - \mathbf{x}'|}{|\mathbf{x}' - \mathbf{y}|}. \tag{25}$$

Then an application of (7) and Lemma 3 leads immediately to the result

$$|GV_\delta\psi(\mathbf{x}) - GV_\delta\psi(\mathbf{x}')| \leq M\, ||\psi||\, R[1 + \tfrac{1}{2}(|k|\, R)]\, |\mathbf{x} - \mathbf{x}'|. \tag{26}$$

[16] The author has also considered the use of similar methods to discuss the partial-wave Schrödinger (Fredholm) equation obtained by expanding (2) in a Legendre series in $P_l(\hat{k}\cdot\hat{x})$. Here one encounters a difficulty in the case of real energies in that there is no analog to Eq. (14). However, the proof still goes through in the case of complex k in the upper half-plane.

Now let $\{\psi\}$ be a set of vectors in C such that $||\psi|| < N$, say. Then the set $\{GV_\delta\psi\}$ is a compact set by the selection theorem. The uniform boundedness of the vectors $GV_\delta\psi$ follows from the Lemma 1; the uniform boundedness as $|\mathbf{x}| \to \infty$ follows from (24), and the equicontinuity of the functions $(GV_\delta\psi)(\mathbf{x})$ follows from (26). Hence the operator GV_δ is compact.

If $V(\mathbf{x})$ is a normal potential then the operator GV is the limit (in norm) of a sequence of operators GV_δ, by Lemma 2, each of which is compact. Hence GV itself is compact.

Q.E.D.

This theorem has also been given by Ikebe.[12] From the Banach–Schauder theorem now follows the Fredholm alternative, and also the existence of the resolvent operator $[1 - GV]^{-1}$ for real, positive energies.[17]

Fredholm Alternative: Either the Lippmann–Schwinger equation (2) has, for each real $\mathbf{k}$, a bounded and continuous solution $\langle \mathbf{x} \mid \mathbf{k}_{\text{in}} \rangle$, or else the corresponding homogeneous equation $\psi = GV\psi$ has a nontrivial bounded and continuous solution.

In fact, the homogeneous equation $\psi = GV\psi$ cannot have a solution for real $\mathbf{k}$ except for zero energy. This follows at once by expanding in partial waves and applying elementary facts about Jost functions.[18]

For complex $\mathbf{k}$, k, of course there may exist solutions (bound states) lying at negative energies. One can show that the solutions of the homogeneous equation for negative (nonzero) energy are moreover square-integrable, because such solutions are exponentially damped as $\mathbf{x} \to \infty$.[19]

[17] Cf. Ref. 6.

[18] L. D. Faddeyev (translated by B. Seckler) J. Math. Phys. **4**, 72 (1963); R. G. Newton, J. Math. Phys. **1**, 319 (1960).

[19] The spectrum has been discussed in Refs. 2, 9, and 12.

APPENDIX. THE SOMMERFELD RADIATION CONDITION

In this appendix we consider the behavior of the wavefunction $\langle \mathbf{x} \mid \mathbf{k}_{\text{in}} \rangle$ for large $|\mathbf{x}|$. It will be necessary here to assume that $V(y)$ has finite moments up to fourth order ($n = 1, 2, 3, 4$),

$$\int_0^\infty y^n\, dy\, |V(y)| < \infty. \tag{27}$$

Theorem. Let $\psi \in C$ and let $V(y)$ have finite first four moments. Then

$$\lim_{x\to\infty} x\, |GV\psi(\mathbf{x}) - M(e^{ikx}/x)| = 0 \tag{28}$$

for real k, where M is the scattering amplitude, defined by

$$M = \frac{-1}{4\pi} \int d^3\mathbf{y} e^{-ik\hat{x}\cdot\mathbf{y}} V(y)\psi(\mathbf{y}). \tag{29}$$

Proof: The condition (27) assures the existence of the integral (29). From the exact identity

$$|\mathbf{x} - \mathbf{y}| - x + \hat{x}\cdot\mathbf{y} = [y^2 - (x - |\mathbf{x} - \mathbf{y}|)^2]/2x, \tag{30}$$

it follows that

$$||\mathbf{x} - \mathbf{y}| - x + \hat{x}\cdot\mathbf{y}| \leq y^2/2x, \tag{31}$$

and hence, using $|e^{i\theta} - 1| \leq \theta$,

$$\left|\frac{e^{ik|\mathbf{x}-\mathbf{y}|}}{|\mathbf{x} - \mathbf{y}|} - \frac{e^{ikx}}{x} e^{-ik\hat{x}\cdot\mathbf{y}}\right| \leq \frac{y}{x}\left\{\frac{1}{|\mathbf{x} - \mathbf{y}|} + \frac{|k|\, y}{2x}\right\}. \tag{32}$$

Consequently,

$$\left|GV\psi(\mathbf{x}) - M\frac{e^{ikx}}{x}\right| \leq \frac{||\psi||}{x^2} \int_0^\infty y^3\, dy\, |V(y)| \left\{1 + \frac{|k|\, y}{2}\right\}. \tag{33}$$

Q.E.D.

On Trapped Trajectories in Brownian Motion*

JULIUS L. JACKSON†
University of Leiden, Leiden, The Netherlands
AND
SAM R. CORIELL‡
National Bureau of Standards, Washington, D. C.
(Received 17 January 1964)

The Brownian motion of an ion in the attractive field of an infinite line charge is investigated with reference to the existence of trapped trajectories, i.e., trajectories which never depart from the neighborhood of the center of attraction. It is shown that there is a discontinuity in the nature of the Brownian motion as a function of the line charge density, marked by the appearance of trapped trajectories when the line charge density becomes greater than a critical value.

I. INTRODUCTION

IN this paper we will discuss a specific example of a most interesting aspect of the Brownian motion of particles in attractive force fields—the existence of trapped trajectories. That is, in an attractive force field, depending on its nature, there is the possibility that certain trajectories of particles undergoing Brownian motion will approach the center of attraction and never depart from the immediate neighborhood of the center. This can only occur if the attractive potential is singular but, as is seen from the example we treat, an infinite attractive potential does not necessarily give rise to trapped trajectories.

The particular case which we discuss is that of a particle undergoing Brownian motion in the field of an infinite line charge, i.e., the electrostatic potential $\Psi = (2\sigma^*/\epsilon) \ln \rho$, where σ^* is the charge per unit length along the line and ϵ is the dielectric constant of the medium. We define the reduced charge per unit length $\sigma = (e\sigma^*/\epsilon kT)$, where e is the absolute value of the electron charge, k is the Boltzmann constant and T is the absolute temperature. The reason for this choice of the potential is twofold. Firstly, this example is related to a particular model for the diffusion of labeled ions in polyelectrolyte solutions. Secondly, the Brownian motion in this case exhibits a most interesting "phase change," such that when σ is less than unity there are no trapped trajectories but for $\sigma > 1$, there is the possibility of trajectories never leaving the immediate neighborhood of the line charge.

In a note in another journal[1] we have called attention to this "phase change" with the suggestion that it might be related to the notion of associated counterions in polyelectrolyte solutions. There we called attention to the existence of a singularity in the first transit time function of an ion at $\sigma = 1$. The first transit time function discussed is the mean time for an ion initially within a cylinder of radius ρ_0 to impinge on the surface of the cylinder. We asserted without proof that the singularity was due to the onset of trapped trajectories. Here we give the proof and discuss some of the details of the Brownian motion for this example.

The discussion presented is largely in terms of the function $W(\mathbf{r}, t)$ defined in a closed region V bounded by a surface S, such that its value is the probability that a Brownian particle which is at the point $\mathbf{r}$ at $t = 0$ has impinged upon the surface at least once within the time t. It can be shown that $W(\mathbf{r}, t)$ is a solution of the partial differential equation[2]

$$D(\nabla^2 W - \nabla\Phi\cdot\nabla W) = (\partial W/\partial t), \qquad (1)$$

where D is the diffusion constant of the medium and the reduced potential $\Phi(\mathbf{r})$ is given by

$$\Phi(\mathbf{r}) = e\Psi(\mathbf{r})/kT. \qquad (2)$$

From the definition of $W(\mathbf{r}, t)$, we see that the boundary condition is

$$W(\mathbf{r}, t) = 1 \qquad (\mathbf{r} \text{ on } S) \qquad (3)$$

and the initial condition is

$$W(\mathbf{r}, 0) = 0 \qquad (\mathbf{r} \in V). \qquad (4)$$

* Supported in part by U. S. Atomic Energy Commission, Division of Research.

† Fulbright Guest Research Grantee on leave of absence from the National Bureau of Standards.

‡ National Research Council–National Bureau of Standards Postdoctoral Research Associate.

[1] J. L. Jackson and S. R. Coriell, J. Chem. Phys. **40**, 1460 (1964).

[2] L. Pontrjagin, A. Andronow, and A. Witt, Zh. Eksperim. i Teor. Fiz. **3**, 172 (1933). A description of this method may also be found in M. Leontowitch, *Statistical Mechanics* (State Press for Technical and Theoretical Literature, Moscow, 1944), or in S. Lifson and J. L. Jackson, J. Chem. Phys. **36**, 2410 (1962). Essentially, the same result has been obtained by a somewhat different method by G. Klein, Proc. Roy. Soc. (London) **A211**, 431 (1952).

From Eq. (1), one may obtain an equation for the mean first transit time function $\bar{t}(\mathbf{r})$ which is the mean time for a particle starting at $\mathbf{r}$ to touch the surface S for the first time. The result is

$$D(\nabla^2\bar{t} - \nabla\Phi\cdot\nabla\bar{t}) = -1 \tag{5}$$

with $\bar{t}(\mathbf{r}) = 0$, $\mathbf{r}$ on S.

II. PRELIMINARY REMARKS

If we now apply Eq. (5) to an infinite right cylinder of radius ρ_0 with a reduced potential $\Phi = 2\sigma \ln \rho$ we obtain

$$\bar{t}(\rho) = (\rho_0^2 - \rho^2)/4D(1 - \sigma). \tag{6}$$

We note that for $\sigma < 1$, the result is sensible but as σ approaches unity $\bar{t}(\rho)$ approaches infinity everywhere in the cylinder and one obtains a meaningless negative result for $\sigma > 1$. Furthermore, one may show directly that this singular behavior at $\sigma = 1$ is not an artifact resulting from improper manipulations with a singular potential. To show this one need only apply Eq. (5) to the same region with a finite, cutoff potential, i.e.,

$$\Phi = 2\sigma \ln \rho, \qquad \alpha \le \rho \le \rho_0, \tag{7a}$$

$$\Phi = 2\sigma \ln \alpha, \qquad 0 \le \rho \le \alpha. \tag{7b}$$

In this case one obtains for the first transit time

$$\bar{t}_\alpha(\rho) = [1/4D(1 - \sigma)][\rho_0^2 - \rho^2 + \sigma(\rho^2 - \alpha^2) - \alpha^{2(1-\sigma)}(\rho_0^{2\sigma} - \alpha^{2\sigma})], \qquad \rho \le \alpha, \tag{8a}$$

$$\bar{t}_\alpha(\rho) = [1/4D(1 - \sigma)][\rho_0^2 - \rho^2 - \alpha^{2(1-\sigma)} \times (\rho_0^{2\sigma} - \rho^{2\sigma})], \qquad \rho \ge \alpha. \tag{8b}$$

Clearly as α approaches zero this result approaches the result of Eq. (6) for $\sigma < 1$ and becomes infinite for $\sigma > 1$.

Seeing that this is a real effect, we would then like to understand it in terms of the details of the Brownian motion. Indeed in what follows we will show that the explanation of this divergence is that for $\sigma > 1$, permanently trapped trajectories are possible. In terms of the function $W(\mathbf{r}, t)$, this assertion states that the function

$$\bar{W}(\mathbf{r}) \equiv \lim_{t\to\infty} W(r, t) \tag{9}$$

will *not* be equal to unity everywhere within V for $\sigma > 1$ but will have values less than unity corresponding to the fact that there is then a nonzero probability of a Brownian trajectory never getting out of V. From Eq. (1) we see that the function $\bar{W}(\mathbf{r})$ must be a solution of

$$\nabla^2\bar{W} - \nabla\bar{W}\cdot\nabla\phi = 0. \tag{10}$$

The solution of Eq. (10) for $\Phi = 2\sigma \ln \rho$ may be written

$$\bar{W}(\rho) = 1 + C[1 - (\rho/\rho_0)^{2\sigma}]. \tag{11}$$

We see that there is no explicit criterion for choosing the constant C and further, no way of telling from the form of the solution that C should be different from zero only when $\sigma \ge 1$. All that can be said unambiguously from Eq. (11) is the rather obvious conclusion that for $\sigma < 0$, one must choose $C = 0$, as otherwise $\bar{W}(\rho)$ is not finite at $\rho = 0$.

It is therefore clear that, enlightenment must come from the time-dependent equation itself, Eq. (1). However, even using the time-dependent equation, if one attempts to solve it for the singular potential, $2\sigma \ln \rho$, one encounters the same ambiguity as occurred in the discussion of Eq. (11). One does not know how to choose the time-independent part of the solution, which is equivalent to writing $\bar{W}(\rho)$.

The procedure which we follow to clarify the situation is to solve the time-dependent equation for $W(\mathbf{r}, t)$, Eq. (1), for the cutoff potential, Eqs. (7a) and (7b), and to investigate the long-time behavior of $W(\mathbf{r}, t)$ in the limit as the cutoff radius α approaches zero. For a finite potential, the time-independent part must always be unity and the solution may always be written as

$$W(\rho, t) = 1 + \sum_{i=1}^{\infty} C_i W_i(\rho) e^{-\lambda_i t}, \tag{12}$$

where the $W_i(\rho)$ are the eigenfunctions of the differential equation which is obtained, the λ_i the corresponding (positive) eigenvalue, and the C_i are coefficients determined by the initial conditions.

In the following section we will show that for $\sigma < 1$ all the λ_i remain finite as $\alpha \to 0$, while for $\sigma > 1$ the lowest eigenvalue λ_1 approaches zero as $\alpha \to 0$. Thus for $\sigma < 1$ we have $\bar{W}(\rho) = 1$, whereas for $\sigma > 1$ we have $\bar{W}(\rho)$ approaching $1 + C_1 W_1$.

III. CALCULATIONS

The differential equation satisfied by the $W_i(\rho)$ of Eq. (12) is

$$\frac{d^2W_i}{d\rho^2} + \frac{1}{\rho}\frac{dW_i}{d\rho} + \frac{\lambda_i}{D}W_i = 0 \qquad 0 \le \rho \le \alpha, \tag{13a}$$

$$\frac{d^2W_i}{d\rho^2} + \frac{1 - 2\sigma}{\rho}\frac{dW_i}{d\rho} + \frac{\lambda_i}{D}W_i = 0, \qquad \alpha \le \rho \le \rho_0, \tag{13b}$$

with the conditions that $W_i(\rho_0) = 0$ and $W_i(0)$

is finite. To satisfy the differential equation in the neighborhood of the point $\rho = \alpha$ it is necessary for W_i and its first derivative to be continuous (otherwise a δ-function-type singularity would be introduced in the differential equation at $\rho = \alpha$). Calling $\mu_i = (\lambda_i/D)^{\frac{1}{2}}$ one may write for the solution of the above equations

$$W_i = J_0(\mu_i\rho), \tag{14a}$$

$$W_i = a_1^{(i)}\rho^\sigma J_\sigma(\mu_i\rho) + a_2^{(i)}\rho^\sigma J_{-\sigma}(\mu_i\rho). \tag{14b}$$

The coefficients $a_1^{(i)}$ and $a_2^{(i)}$ and the eigenvalues μ_i are determined from the conditions

$$J_0(\mu_i\alpha) = a_1^{(i)}\alpha^\sigma J_\sigma(\mu_i\alpha) + a_2^{(i)}\alpha^\sigma J_{-\sigma}(\mu_i\alpha), \tag{15}$$

$$J_0'(\mu_i\alpha) = \sigma\alpha^{\sigma-1}[a_1^{(i)}J_\sigma(\mu_i\alpha) + a_2^{(i)}J_{-\sigma}(\mu_i\alpha)] + \alpha^\sigma[a_1^{(i)}J_\sigma'(\mu_i\alpha) + a_2^{(i)}J_{-\sigma}(\mu_i\alpha)], \tag{16}$$

and

$$a_1^{(i)}J_\sigma(\mu_i\rho_0) + a_2^{(i)}J_{-\sigma}(\mu_i\rho_0) = 0. \tag{17}$$

We are interested in the behavior of the solution as α approaches zero. We therefore expand the Bessel functions in Eqs. (15) and (16) and keep the leading terms in α in each case. The resulting equations are

$$1 = a_1^{(i)}\mu_i^\sigma\alpha^{2\sigma}/[2^\sigma(\sigma)!] + a_2^{(i)}2^\sigma\mu_i^{-\sigma}/[(-\sigma)!], \tag{18a}$$

$$-1 = a_1^{(i)}4\mu_i^{\sigma-2}\alpha^{2(\sigma-1)}/[2^\sigma(\sigma-1)!] - a_2^{(i)}2^\sigma\mu_i^{-\sigma}/[(-\sigma+1)!]. \tag{18b}$$

The leading terms in the solution are then

$$a_1^{(i)} = 2^{\sigma-2}\mu_i^{2-\sigma}\sigma!/(1-\sigma)\alpha^{2(\sigma-1)}, \tag{19a}$$

$$a_2^{(i)} = (-\sigma)!\,\mu_i^\sigma/2^\sigma. \tag{19b}$$

For $\sigma < 1$, the coefficient $a_1^{(i)}$ approaches zero as α approaches zero. The solution in the region $(\alpha \leq \rho \leq \rho_0)$ then approaches $J_{-\sigma}(\mu_i\rho)$. All of the eigenvalues, the solutions of

$$J_{-\sigma}(\mu_i\rho_0) = 0 \tag{20}$$

are then finite as α approaches zero and $W(\rho, t)$ approaches unity uniformly for t going to infinity.

On the other hand when $\sigma > 1$, the coefficient $a_1^{(i)}$ approaches infinity and the eigenfunctions in the region $(\alpha \leq \rho \leq \rho_0)$ are dominated by $J_\sigma(\mu_i\rho)$. In this case, the lowest eigenvalue approaches zero as α approaches zero. Assuming that this is so and using Eqs. (19a) and (19b) in Eq. (17), upon expanding the Bessel functions, one obtains for the single lowest eigenvalue

$$\lambda_1 = 4D(\sigma-1)\alpha^{2(\sigma-1)}/\rho_0^{2\sigma}. \tag{21}$$

Thus in the limit as $\alpha \to 0$, the time constant associated with the eigenfunction $W_1(\rho)$ approaches zero and the limiting form of $\bar{W}(\rho)$, the probability that a particle initially at ρ will ultimately touch the surface $\rho = \rho_0$ is

$$\lim_{\alpha\to 0}\bar{W}(\rho) = 1 + C_1W_1(\rho). \tag{22}$$

Inserting the results of Eqs. (19a), (19b), and (21) into Eqs. (14a) and (14b) one obtains for the limiting form of the first eigenfunction

$$W_1 = 1, \qquad 0 \leq \rho \leq \alpha, \tag{23a}$$

$$W_1 = 1 - (\rho/\rho_0)^{2\sigma}, \qquad \alpha \leq \rho \leq \rho_0. \tag{23b}$$

These expressions for W_1 are sufficient for our purposes although from their form, the continuity of the function and its first derivative are not apparent. The reason for this is that we have dropped terms of order $(\alpha/\rho_0)^{2\sigma}$ compared to terms of order unity. To verify the continuity of the solution and its derivative, one need only take the second term in the expansion of $J_0(\mu_1\,\rho)$ and include the next term in $a_2^{(1)}$, viz.

$$a_2^{(1)} = [(-\sigma)!/2^\sigma]\mu_1^\sigma[1 - \tfrac{1}{4}\mu_1^2\alpha^2]. \tag{24}$$

To find the limiting form of $\bar{W}_1(\rho)$ which is the "escape probability" for the logarithmic potential when $\sigma > 1$ we must compute C_1. To do this we make use of the orthogonality condition of the W_i:

$$\int_0^\alpha W_iW_k\rho\,d\rho + \alpha^{2\sigma}\int_\alpha^{\rho_0} W_iW_k\rho^{1-2\sigma}\,d\rho = K_i\,\delta_{ik} \tag{25}$$

which can be shown from Eqs. (13a) and (13b). We then have

$$-\int_0^\alpha W_1\rho\,d\rho - \alpha^{2\sigma}\int_\alpha^{\rho_0} W_1\rho^{1-2\sigma}\,d\rho = C_1\left\{\int_0^\alpha W_1^2\rho\,d\rho + \alpha^{2\sigma}\int_\alpha^{\rho_0} W_1^2\rho^{1-2\sigma}\,d\rho\right\}. \tag{26}$$

The leading term in the limit as $\alpha \to 0$ is then

$$C_1 = -1. \tag{27}$$

The escape probability is then

$$\bar{W}(\rho) = (\rho/\rho_0)^{2\sigma}. \tag{28}$$

It should be noted in conclusion that this result could well have been expected in that when there are trapped trajectories possible one would expect that the escape probability would approach zero as ρ goes to zero.

The Brownian motion of ions in an attractive logarithmic field is thus formally similar to a multiplicative stochastic process, e.g., a chain reaction.

In a chain reaction above the critical value, the relevant statistical question is what is the probability of a given chain extinguishing (the average chain length diverges), and below the critical value one may ask for the average chain length. So in the present problem, above the critical value of σ, the average time of first transit diverges and one may find the probability of an ion escaping, while below the critical value one may compute the average first-transit time.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 8 AUGUST 1964

The Representation of Canonical Variables as the Limit of Infinite Space Volume: The Case of the BCS Model*

HIROSHI EZAWA†
Department of Physics and Astronomy, University of Maryland, College Park, Maryland
(Received 11 February 1964)

We examine the possibility of obtaining the representation suitable for a given Hamiltonian through a limiting procedure. The model employed is that of the Bardeen–Cooper–Schrieffer's superconductivity theory, for which the infinite-volume limit of the Wightman functions gives the right representation. This method is found to be helpful in analyzing the so-called symmetry breaking solutions. In particular, it serves for throwing some light to the Haag's work on the gauge property of the BCS model as well as to the Nambu's dynamical model of elementary particles. What is meant by the limit of infinite volume is studied with regard to the convergence of the Hamiltonian operator.

1. INTRODUCTION

ONE of the most interesting features of the quantum theory of fields is that the canonical commutation (anticommutation) relations allow many representations inequivalent to each other and, moreover, that the Hamiltonian often restricts the representation to be employed.[1-3] To find the representation suitable for a given Hamiltonian is thus of a primary importance and yet very difficult in general. An interesting suggestion seems to have been put forward by the recent works of Araki and Woods for the infinite free Bose gas[4] and of Araki and Wyss for the Fermi gas[5]; they have presented a natural method to construct the non-Fock representations.

The Fock representation[6] (the name of Jordan–Wigner representation[7] might be more appropriate to the fermion case, but we prefer the unified name for the unified concept) is the traditional one in quantum field theory and the basis vectors here are characterized by a finite number of states occupied by the particles. According to the Gårding–Wightman classification[8] there are continuously many non-Fock representations inequivalent to each other and, of course, to Fock's, whose basis vectors are featured by the infinitely many background states occupied by the particles. But, not much has been known about these representations.

The idea of Araki and others is that, in the case of infinite volume, a gas with given density contains infinitely many particles and would naturally lead to the non-Fock representations. Their tools in this investigation are the Wightman functions which are known to characterize the representation.[9] Taking as a vacuum the state with a given distribution of particles in momentum space, they calculated the Wightman functions first in the case of the finite volume V, where, the total number of particles being finite for a given density ρ, the Fock representation works. Then, they took the limit

* This work was supported in part by the National Science Foundation and the U. S. Air Force Office of Scientific Research.

† On leave of absence from the Department of Physics, University of Tokyo, Tokyo, Japan.

[1] L. Van Hove, Physica **18**, 145 (1952); O. Miyatake, J. Inst. Polytech., Osaka City Univ. **2A**, 89 (1952), **3A**, 145 (1952); see also H. Ezawa, Progr. Theoret. Phys. (Kyoto) **30**, 545 (1963), Y. Kato and N. Mugibayashi, *ibid.* **30**, 409 (1963).

[2] F. Coester and R. Haag, Phys. Rev. **117**, 1137 (1960).

[3] H. Araki, J. Math Phys. **1**, 492 (1960).

[4] H. Araki and E. J. Woods, J. Math. Phys. **4**, 637 (1963).

[5] H. Araki and W. Wyss (preprint).

[6] V. Fock, Z. Phys. **75**, 622 (1932); J. M. Cook, Trans. Am. Math. Soc. **74**, 222 (1953).

[7] P. Jordan and E. P. Wigner, Z. Phys. **47**, 631 (1928).

[8] L. Gårding and A. S. Wightman, Proc. Natl. Acad. Sci. U. S. **40**, 617, 622 (1954); A. S. Wightman and S. S. Schweber, Phys. Rev. **98**, 812 (1955).

[9] A. S. Wightman, Phys. Rev. **101**, 860 (1956).

In a chain reaction above the critical value, the relevant statistical question is what is the probability of a given chain extinguishing (the average chain length diverges), and below the critical value one may ask for the average chain length. So in the present problem, above the critical value of σ, the average time of first transit diverges and one may find the probability of an ion escaping, while below the critical value one may compute the average first-transit time.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 8 AUGUST 1964

The Representation of Canonical Variables as the Limit of Infinite Space Volume: The Case of the BCS Model*

HIROSHI EZAWA†
Department of Physics and Astronomy, University of Maryland, College Park, Maryland
(Received 11 February 1964)

We examine the possibility of obtaining the representation suitable for a given Hamiltonian through a limiting procedure. The model employed is that of the Bardeen–Cooper–Schrieffer's superconductivity theory, for which the infinite-volume limit of the Wightman functions gives the right representation. This method is found to be helpful in analyzing the so-called symmetry breaking solutions. In particular, it serves for throwing some light to the Haag's work on the gauge property of the BCS model as well as to the Nambu's dynamical model of elementary particles. What is meant by the limit of infinite volume is studied with regard to the convergence of the Hamiltonian operator.

1. INTRODUCTION

ONE of the most interesting features of the quantum theory of fields is that the canonical commutation (anticommutation) relations allow many representations inequivalent to each other and, moreover, that the Hamiltonian often restricts the representation to be employed.[1-3] To find the representation suitable for a given Hamiltonian is thus of a primary importance and yet very difficult in general. An interesting suggestion seems to have been put forward by the recent works of Araki and Woods for the infinite free Bose gas[4] and of Araki and Wyss for the Fermi gas[5]; they have presented a natural method to construct the non-Fock representations.

The Fock representation[6] (the name of Jordan–Wigner representation[7] might be more appropriate to the fermion case, but we prefer the unified name for the unified concept) is the traditional one in quantum field theory and the basis vectors here are characterized by a finite number of states occupied by the particles. According to the Gårding–Wightman classification[8] there are continuously many non-Fock representations inequivalent to each other and, of course, to Fock's, whose basis vectors are featured by the infinitely many background states occupied by the particles. But, not much has been known about these representations.

The idea of Araki and others is that, in the case of infinite volume, a gas with given density contains infinitely many particles and would naturally lead to the non-Fock representations. Their tools in this investigation are the Wightman functions which are known to characterize the representation.[9] Taking as a vacuum the state with a given distribution of particles in momentum space, they calculated the Wightman functions first in the case of the finite volume V, where, the total number of particles being finite for a given density ρ, the Fock representation works. Then, they took the limit

* This work was supported in part by the National Science Foundation and the U. S. Air Force Office of Scientific Research.

† On leave of absence from the Department of Physics, University of Tokyo, Tokyo, Japan.

[1] L. Van Hove, Physica **18**, 145 (1952); O. Miyatake, J. Inst. Polytech., Osaka City Univ. **2A**, 89 (1952), **3A**, 145 (1952); see also H. Ezawa, Progr. Theoret. Phys. (Kyoto) **30**, 545 (1963), Y. Kato and N. Mugibayashi, *ibid.* **30**, 409 (1963).

[2] F. Coester and R. Haag, Phys. Rev. **117**, 1137 (1960).

[3] H. Araki, J. Math Phys. **1**, 492 (1960).

[4] H. Araki and E. J. Woods, J. Math. Phys. **4**, 637 (1963).

[5] H. Araki and W. Wyss (preprint).

[6] V. Fock, Z. Phys. **75**, 622 (1932); J. M. Cook, Trans. Am. Math. Soc. **74**, 222 (1953).

[7] P. Jordan and E. P. Wigner, Z. Phys. **47**, 631 (1928).

[8] L. Gårding and A. S. Wightman, Proc. Natl. Acad. Sci. U. S. **40**, 617, 622 (1954); A. S. Wightman and S. S. Schweber, Phys. Rev. **98**, 812 (1955).

[9] A. S. Wightman, Phys. Rev. **101**, 860 (1956).

$V \to \infty$ of the Wightman functions keeping ρ constant, finding in fact that the limit Wightman functions describe a series of inequivalent representations, each one specified by the particle density ρ and the distribution function. In some cases, the representation thus arrived at was found to be reducible.

This limiting procedure of constructing new representations reminds us of the situation in the BCS theory of superconductivity,[10] where, once the Hamiltonian is reduced to the so-called BCS Hamiltonian in the approximation to take only the Cooper-pair interaction[11] into account, the Hamiltonian can be replaced by a quadratic form of the field operators so that it can be diagonalized by a simple principal-axis transformation. There are three interesting points: (i) The principal-axis transformation is actually possible within the Fock representation so far as the volume[12] V of the superconductor is finite, but the limit $V \to \infty$ of the transformation does not exist (see Sec. 4 of this paper). (ii) The quadratic substitution for the BCS Hamiltonian turns out to be exact[13] in the limit $V \to \infty$. Since the Fock representation of the electron field does no longer work at $V \to \infty$, we have to find some other representation, which, we may hope, will be obtained by the limiting procedure of Araki and others.[4,5] To begin with we take a finite volume V and the Fock representation of electron field, within which the principal-axis transformation is possible to yield the physical vacuum. With this physical vacuum as a cyclic vector, we can calculate the Wightman functions for the finite V. Then, keeping the density of electrons constant, we take the limit $V \to \infty$ to see what representation will come out. We show in Sec. 5 that this method can successfully be applied, indeed. However, this is not so rewarding in BCS case because we can find the suitable representation in a much simpler way. The interesting question here is whether this method can be so generalized as to apply to other cases.

The reward comes from the third point of interest: (iii) In spite of the gauge invariance of the original BCS Hamiltonian, its quadratic substitute does not share the invariance, although the substitution can be justified exactly at $V \to \infty$. A similar situation arises in Nambu's theory on the origin of the mass of elementary particles,[14] in which he has shown that, despite the γ_5 invariance of his Hamiltonian, the physical Dirac particle can have nonzero mass. The analogy with the BCS theory is very close,[15,16] and the mass of Nambu theory corresponds to the energy gap in the BCS theory. In connection with the noninvariance property, Nambu[15] has pointed out that the physical vacuum is highly degenerate and that the different vacuum states are transmutable to each other by the γ_5-gauge transformation, although his argument is merely a heuristic one. If so, by a linear combination of the gauge-noninvariant vacua we may construct the invariant vacuum. It has been pointed out by Haag that the representations of field operators with the gauge-invariant vacuum are reducible and that the quadratic substitution for the original Hamiltonian is allowed only in the irreducible representations (Haag's Lemma[17]). In addition, it is shown that the operator inducing the gauge transformation, or the number operator of electrons, does not exist due to the infinite volume filled up with electrons of a given density. Haag's argument is based upon an Ansatz about the form of Wightman functions which are intended to describe the gauge-invariant representations. Since his method is so ingenious and the result so fascinating, we are driven to look for the underlying physical and mathematical picture. In addition, one may wonder if the Wightman functions of Haag are in fact proper to the BCS Hamiltonian; this would entail the question as to what is meant by the infinite-volume limit of the Hamiltonian. Thus, it is interesting to see if the limiting procedure of Araki and others leads us to the Ansatz of Haag. As we see in Sec. 6, this is in fact the case. It should be mentioned here that the positivity condition, an essential condition for the system of Wightman functions, is automatically guaranteed in this approach. Also, the reducibility of the gauge-invariant representations is exhibited as quite a natural consequence.

It is a different task to see if the Wightman functions thus obtained by the limiting procedure $V \to \infty$ are actually proper to the limit Hamiltonian. We obtain the affirmative answer in Sec. 4 and so,

[10] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **108**, 1175 (1957); N. N. Bogoliubov, V. V. Tolmachev, and D. V. Shirkov, *A New Method in the Theory of Superconductivity* (English translation, Consultants Bureau, Inc., New York, 1959).

[11] L. N. Cooper, Phys. Rev. **104**, 1189 (1956).

[12] What we mean by "volume" becomes clear below.

[13] N. N. Bogoliubov, Physica, Suppl. 1 **26**, (1960).

[14] Y. Nambu and G. Jona-Lasinio, Phys. Rev. **122**, 345 (1961); M. Baker and S. L. Glashow, Phys. Rev. **128**, 2462 (1962); see also Y. Nambu, Phys. Rev. **117**, 648 (1960).

[15] H. Umezawa, S. Kamefuchi, and Y. Takahashi, Ann. Phys. (to be published).

[16] M. Suzuki, Progr. Theoret. Phys. (Kyoto) **30**, 138, 625 (1963).

[17] R. Haag, Nuovo Cimento **25**, 287 (1962). As to the generalization to relativistic case (especially to Nambu's), see Ref. 15.

taking the above result into account, we can conclude that the Wightman functions assumed by Haag are the only ones suitable for the BCS model. The argument as to the limit Hamiltonian goes as follows. First of all, we have to ask what is meant by the limit $V \to \infty$ of the Hamiltonian. An answer may be obtained by such an argument given by Araki for a Bose field[3] that a system of Wightman functions determines the Hamiltonian proper to itself. In this paper, however, we want to ask the question in terms of the operator convergence.

When we speak of the volume V in quantum field theory it is usual to confine the field within a box of volume V and to construct the creation and the annihilation operators associated with a system of testing functions orthonormal in V, taking their Fock representation. Then, in order to ask the volume change effects, it may seem quite natural to use the scale transformation[18] which enables us to visualize the effects of volume change, say from V to V', within one Hilbert space, e.g., the Fock space associated with V. But, the transformation is known to be nonexistent unless both V and V' are finite. This formalism is not adequate for asking the limit $V' \to \infty$.

Then, in Sec. 2 we make a formalism such that the operators belonging to the volume V are regarded as a subalgebra (the restriction to V) of the algebra of the field operators extending over the whole space; the representation of the former is thus embedded in the representation of the latter. This formalism is used in the following sections.

In Sec. 3, we digress to sharpen the condition for Haag's Lemma.[17] In Sec. 4 we investigate how the quadratic Hamiltonian, which is restricted to V, and its eigenvectors behave in the limit $V \to \infty$. These sections constitute a preparatory step towards the investigation of the representation problems sketched above.

2. RESTRICTIONS OF OPERATORS TO A FINITE VOLUME

The first step in our program described in the introduction is to construct the field operators restricted to a finite volume V together with their representation. We take up an example of the BCS theory of superconductivity.[10]

[18] When a box $-l < x_k < l$, $k = 1, 2, 3$, is expanded into $-l' < x_k' < l'$, the field $\psi(\mathbf{x})$ in the first box undergoes a transformation into $\psi'(\mathbf{x}') = (V/V')^{\frac{1}{2}} \psi(\mathbf{x})$, $(x_k' = (l'/l)\, x_k$, $V = 8l^3$, $V' = 8l'^3)$ so that $\{\psi'(\mathbf{x}'), \psi'^*(\mathbf{y}')\} = \delta(\mathbf{x}' - \mathbf{y}')$ follows from $\{\psi(\mathbf{x}), \psi^*(\mathbf{y})\} = \delta(\mathbf{x} - \mathbf{y})$. This is the scale transformation for fermion field. It does not make sense if $V' \to \infty$. See H. Ezawa, Y. Tomozawa, and H. Umezawa, Nuovo Cimento 5, 810 (1957).

The basic object is, of course, the algebra generated by the smeared-out electron fields (nonrelativistic):

$$\psi_i(f) = \int f^*(\mathbf{x})\psi_i(\mathbf{x})\, d\mathbf{x} \qquad (i = 1, 2), \tag{2.1}$$

where $\psi_i(\mathbf{x})$'s are two different (different in spin direction: $i = 1$, spin up; $i = 2$, spin down) nonrelativistic electron fields extending to infinity, and the function $f(\mathbf{x})$ is assumed to be square-integrable,

$$||f||^2 = \int |f(\mathbf{x})|^2\, d\mathbf{x} < \infty. \tag{2.2}$$

The function $f(\mathbf{x})$ may have a tail extending to infinity, too, and the integral above is taken over the whole space. The set of such functions constitutes the Hilbert space $L_2(\infty)$ and we can take a denumerable set of orthonormal functions $\{f_n(\mathbf{x})\}$ as a basis. Then we have

$$\begin{aligned} \{\psi_i(f_n), \psi_j(f_m)\} &= \{\psi_i^*(f_n), \psi_j^*(f_m)\} = 0, \\ \{\psi_i(f_n), \psi_j^*(f_m)\} &= \delta_{ij}\delta_{nm}. \end{aligned} \tag{2.3}$$

The underlying assumption in restricting $f(\mathbf{x})$ to $L_2(\infty)$ is that every operator of physical interest is expressed in terms of the fields smeared-out by them. In fact, the interaction Hamiltonian, for example, can be expressed at least formally as follows:

$$\begin{aligned} &\int \psi^*(\mathbf{x})\psi^*(\mathbf{y})U(\mathbf{x} - \mathbf{y})\psi(\mathbf{x})\psi(\mathbf{y})\, d\mathbf{x}\, d\mathbf{y} \\ &= \sum_{n,n',m,m'} \tilde{U}_{n,n',m,m'} \psi^*(f_n)\psi^*(f_{n'})\psi(f_m)\psi(f_{m'}), \end{aligned} \tag{2.4}$$

where

$$\begin{aligned} &\tilde{U}_{n,n',m,m'} \\ &\equiv \int f_n^*(\mathbf{x}) f_{n'}^*(\mathbf{y}) U(\mathbf{x} - \mathbf{y}) f_m(\mathbf{x}) f_{m'}(\mathbf{y})\, d\mathbf{x}\, d\mathbf{y}, \end{aligned} \tag{2.5}$$

and we have omitted the spin indices.

Now, we introduce the volume V to which the operators should be restricted. Let's take it to be a cube of edge $2l$:

$$V : -l < x_k < l, \qquad (k = 1, 2, 3). \tag{2.6}$$

Correspondingly, the restriction of $f(\mathbf{x})$ is defined as

$$\hat{f}(\mathbf{x}) = \begin{cases} f(\mathbf{x}), & \mathbf{x} \in V, \\ 0, & \text{otherwise}, \end{cases} \tag{2.7}$$

and also the restriction of $\psi_i(f)$ as $\psi_i(\hat{f})$. Although the restriction of the system $\{f_n(\mathbf{x})\}$ is no longer orthonormal, we still have the closure property

$$\sum_n f_n^*(\mathbf{x}) f_n(\mathbf{y}) = \delta(\mathbf{x}-\mathbf{y}), \qquad \text{if } \mathbf{x}, \mathbf{y} \in V, \tag{2.8}$$

and the relation (2.4), for example, remains true when the integration domains of both (2.4) and (2.5) are restricted to V. The restriction $\mathcal{H}(V)$ of Hamiltonian $\mathcal{H}$ is defined by integrating its density over V. The restriction $\mathcal{H}(V)$ can be expressed in terms of the restricted fields only.

The above procedure does not imply that the physical system under consideration (superconductor) should be confined in V. In fact in the physical system there are interactions between inside and outside of V also. We are just anticipating that the $\mathcal{H}(V)$ should be a good "approximation" to $\mathcal{H}$ if V is sufficiently large, the implication of which will be the main subject of this paper.

In order to treat the field operators restricted to V, it is most suitable to take the system

$$f_p(\mathbf{x}) = \begin{cases} V^{-\frac{1}{2}} \exp[+i\mathbf{p}\mathbf{x}], & \text{for } \mathbf{x} \in V, \\ 0, & \text{otherwise}, \end{cases} \tag{2.9}$$

as a basis of smearing functions, where $\mathbf{p} = (n_1, n_2, n_3)(\pi/l)$ with $n_k = 0, \pm 1, \pm 2, \cdots$. Then the smeared-out fields

$$a_{ip} \equiv \psi_i(f_p), \qquad a_{ip}^* \equiv \psi_i^*(f_p) \tag{2.10}$$

satisfy the anticommutation relations similar to (2.3).

Now, we are in the position to speak of the representations of those fields introduced above. Suppose that an irreducible representation in some space $\mathfrak{H}$ is given for the fields $\psi_i(f)$ smeared out in the infinite volume, then we can have a representation induced for the fields $\psi_i(g)$ restricted to the finite volume V since $\{g(\mathbf{x})\}$ is a subset of $\{f(x)\}$; or more explicitly through

$$\psi_i(g) = \sum_n (g, f_n)\psi_i(f_n). \tag{2.11}$$

However, the representation of $\psi_i(g)$ thus embedded in the representation of $\psi_i(f)$ is no more irreducible because, for example, $\psi_i^*(h)\ \psi_i^*(h')$ commutes with $\psi_i(g)$ if the support of $h(\mathbf{x})$ and $h'(\mathbf{x})$ does not intersect with V. We must deal with such embedding in the following, in particular in Sec. 4. Because of the reducibility just mentioned, the transformation (2.11) cannot be unitary.

To conclude this section, we have to add that, in any representation of the algebra of $\psi_i(f)$'s, we have

$$||\psi_i(f)|| \leq ||f|| \qquad (i = 1, 2), \tag{2.12}$$

due to the anticommutation relations.

3. A CONDITION FOR HAAG'S LEMMA

Throughout this paper, we are concerned with the BCS model of superconductivity.[10,17] The restriction to V of the Hamiltonian is given by

$$\mathcal{H}(V) = \mathcal{H}_0(V) + \mathcal{H}_I(V),$$

$$\mathcal{H}_0(V) \equiv \int_V \sum_{i=1}^{2} \frac{1}{2m} \nabla\psi_i^*(\mathbf{x}) \cdot \nabla\psi_i(\mathbf{x})\, d\mathbf{x}, \tag{3.1}$$

$$\mathcal{H}_I(V) \equiv \frac{1}{V} \int_V \psi_1^*(\mathbf{x})\psi_2^*(\mathbf{x}+\xi) \times \psi_2(\mathbf{y}+\mathbf{n})\psi_1(\mathbf{y})v(\xi, \mathbf{n})\, d\mathbf{x}\, d\mathbf{y}\, d\xi\, d\mathbf{n}.$$

Since the interaction is limited to the Cooper pair[11] (the pair of electrons whose momenta and spins are just the reverse of each other), the phase volume is reduced and we have a factor $1/V$ in front of the interaction Hamiltonian. This factor $1/V$ plays an important role in the following arguments.

Haag has pointed out[17] that, in the limit $V \to \infty$, the interaction Hamiltonian can be substituted for by a quadratic expression,

$$\mathcal{H}_I(V) \equiv \int_V [\Delta(\xi)\psi_1^*(\mathbf{x})\psi_2^*(\mathbf{x}+\xi) + \Delta^*(\xi)\psi_2(\mathbf{x}+\xi)\psi_1(\mathbf{x})]\, d\mathbf{x}\, d\xi + c \text{ number}, \tag{3.2}$$

as long as one is confined to an irreducible representation of $\psi_i(f)$'s (Haag's Lemma). The function $\Delta(\xi)$ is seen to be given by the vacuum expectation value of $B(\xi)$ in (3.3) below, and it can be different depending on which one of the degenerate vacuum states is used to form the expectation value; in other words, the function $\Delta(\xi)$ is representation-dependent. An integral equation can be obtained if we compute the vacuum expectation value of $B(\xi)$ by making use of the transformation (4.2) in the next section.[19]

Now, our aim is to give a condition on $v(\xi, \mathbf{n})$ that guarantees Haag's Lemma mentioned above. Following Haag, we first calculate the commutation relation

$$[\mathcal{H}_I(V), \psi_1(f)] = \frac{1}{V} \int_V A(\xi)B(\xi)\, d\xi,$$

where

$$A(\xi) \equiv \int_V f^*(\mathbf{z})\psi_2^*(\mathbf{z}+\xi)\, d\mathbf{z}, \tag{3.3}$$

$$B(\xi) \equiv \int v(\xi, \mathbf{n})\psi_2(\mathbf{y}+\mathbf{n})\psi_1(\mathbf{y})\, d\mathbf{y}\, d\mathbf{n},$$

[19] As to the solution of this integral equation, see M. Kitamura, Progr. Theoret. Phys. (Kyoto) **30**, 435 (1963) and Ref. 10.

to show that $B(\xi)$ can actually be regarded as c number, and therefore that the commutator (3.3) is linear in field operator. In fact, in the commutation relation,

$$[[\mathfrak{H}_I(V), \psi_1(f)], R] = \frac{1}{V}\int_V A(\xi)[B(\xi), R]\,d\xi + \frac{1}{V}\int_V [A(\xi), R]B(\xi)\,d\xi,$$

with R = any one of $\psi_i^*(g)$ and $\psi_i(g)$ $(i = 1, 2)$, we can show that the first integral vanishes as $V \to \infty$ if $v(\xi, \mathbf{n})$ is integrable,

$$\int |v(\xi, \mathbf{n})|\,d\xi\,d\mathbf{n} < \infty, \tag{3.4}$$

and therefore that $B(\xi)$ can be regarded as c number within any irreducible representation.

The proof is trivial for the cases $R = \psi_i(g)$ since $[B, R] = 0$. In the case of $R = \psi_2^*(g)$, we have $[B(\xi), \psi_2^*(g)]$

$$= \int_V v(\xi, \mathbf{n})g(\mathbf{y} + \mathbf{n})\psi_1(\mathbf{y})\,d\mathbf{y}\,d\mathbf{n} \equiv D(\xi),$$

and then the desired result follows from

$$\left\|\frac{1}{V}\int_V A(\xi)[B(\xi), R]\,d\xi\right\| \le \frac{1}{V}\int_V \|A(\xi)D(\xi)\|\,d\xi \le \frac{1}{V}\|f\|\int_V \|D(\xi)\|\,d\xi \tag{3.5}$$

because, due to our assumption (3.4),

$$\int_V \|D(\xi)\|\,d\xi \le \int_V d\xi\,d\mathbf{n}\,|v(\xi, \mathbf{n})|\cdot\left\|\int_V d\mathbf{y}g(\mathbf{y} + \mathbf{n})\psi_1(\mathbf{y})\right\| \le \|g\|\int_V d\xi\,d\mathbf{n}\,|v(\xi, \mathbf{n})| < \infty. \tag{3.6}$$

Here, we have made repeated uses of the boundedness of $\psi_i(f)$, (2.12). The same proof applies for the case $R = \psi_1^*(g)$. In the same way we can show that, under the condition (3.4) the commutation relation $[\mathfrak{H}_I(V), \psi_2(f)]$ is linear in ψ_1^*, etc.

In conclusion, a sufficient condition for Haag's Lemma mentioned above is given by Eq. (3.4).

4. THE FOCK REPRESENTATION

A

In this section we analyze

$$\left.\begin{aligned} K'(V) &= \mathfrak{H}_0(V) + \mathfrak{H}_I'(V) - \mu N(V) \\ \text{and its limit at } V &\to \infty, \text{ where} \\ N(V) &= \sum_{i=1}^{2}\int_V \psi_i^*(\mathbf{x})\psi_i(\mathbf{x})\,d\mathbf{x} \end{aligned}\right\}, \tag{4.1}$$

and the chenical potential μ is introduced to fix the electron density. It is convenient hereafter to call $K'(V)$ "Hamiltonian."

For a while we use the Fock representation of a_{ip}'s in (2.10), within which $K'(V)$ will actually be shown to be diagonalized. In fact, the Bogoliubov transformation

$$\begin{aligned} \gamma_{1\mathbf{p}} &= a_{1\mathbf{p}}\cos\theta_{\mathbf{p}} + a_{2-\mathbf{p}}^*\sin\theta_{\mathbf{p}}, \\ \gamma_{2-\mathbf{p}} &= -a_{1\mathbf{p}}^*\sin\theta_{\mathbf{p}} + a_{2-\mathbf{p}}\cos\theta_{\mathbf{p}} \end{aligned} \tag{4.2}$$

leads us to the diagonal $K'(V)$,

$$K'(V) = \sum_{\mathbf{p}} E(\mathbf{p})(\gamma_{1\mathbf{p}}^*\gamma_{1\mathbf{p}} + \gamma_{2\mathbf{p}}^*\gamma_{2\mathbf{p}}), \tag{4.3}$$

where the zero-point energy is removed and

$$E(\mathbf{p}) = [\epsilon^2(\mathbf{p}) + \tilde{\Delta}(\mathbf{p})^2]^{\frac{1}{2}}; \quad \epsilon(\mathbf{p}) = \mathbf{p}^2/2m - \mu, \tag{4.4}$$

$$\begin{aligned} \cos\theta_{\mathbf{p}} &= \frac{\tilde{\Delta}}{[(E-\epsilon)^2 + \tilde{\Delta}^2]^{\frac{1}{2}}}, \\ \sin\theta_{\mathbf{p}} &= \frac{E-\epsilon}{[(E-\epsilon)^2 + \tilde{\Delta}^2]^{\frac{1}{2}}}. \end{aligned} \tag{4.5}$$

The Fourier transform $\tilde{\Delta}(\mathbf{p})$ of $\Delta(\xi)$ is assumed to be real so that the Bogoliubov transformation should take a simple form as given in (4.2) and, in particular, in (4.6) below.[20] Now, the only question about the diagonalizability of $K'(V)$ is whether the vacuum of γ_{ip}, $|\Omega(V)\rangle$, actually belongs to the Fock space of a_{ip} or not. In the following we shall show that the answer is affirmative as far as $V < \infty$.

Let us write the transformation (4.2) in the form

$$\gamma_{ip} = \exp[-\theta_{\mathbf{p}}T_{\mathbf{p}}]a_{ip}\exp[\theta_{\mathbf{p}}T_{\mathbf{p}}] \quad (i = 1, 2), \tag{4.6}$$

with

$$T_{\mathbf{p}} \equiv a_{1\mathbf{p}}^*a_{2-\mathbf{p}}^* + a_{1\mathbf{p}}a_{2-\mathbf{p}}.$$

Taking the relation $T_{\mathbf{p}}^3 = -T_{\mathbf{p}}$ into account, we can readily see

$$\left.\begin{aligned} &\exp[-\theta_{\mathbf{p}}T_{\mathbf{p}}] = -T_{\mathbf{p}}\sin\theta_{\mathbf{p}} + T_{\mathbf{p}}^2(1 - \cos\theta_{\mathbf{p}}) + 1, \\ &T_{\mathbf{p}}^2 = (N_{1\mathbf{p}} - N_{2-\mathbf{p}})^2 - 1; \quad N_{ip} = N_{ip}^2 = a_{ip}^*a_{ip}, \end{aligned}\right\} \tag{4.7}$$

thus confirming the equivalence between (4.2) and (4.6).

The physical vacuum $|\Omega(V)\rangle$ of (4.3) is obtained, at least formally, from the bare vacuum $|0\rangle$ of a_{ip}'s through the repeated application of the transformation (4.7):

[20] In case that $\tilde{\Delta}(p)$ appears to be complex, we have only to adjust the phase of a_{ip} to recover its reality. The gauge transformation of this kind exists within the Fock representation.

$$|\Omega(V)\rangle = (\prod_{\mathbf{p}} \exp[-\theta_{\mathbf{p}} T_{\mathbf{p}}])\,|0\rangle$$
$$= (\prod_{\mathbf{p}} [\cos\theta_{\mathbf{p}} - a_{1\mathbf{p}}^{*} a_{2-\mathbf{p}}^{*} \sin\theta_{\mathbf{p}}])\,|0\rangle. \quad (4.8)$$

Here, however, the convergence of the infinite product over $\mathbf{p}$ should be examined. For this purpose we restrict the product up to the momentum $|\mathbf{p}| = P$ and P' to ask whether or not the Cauchy condition[21]

$$\|(\prod_{|\mathbf{p}|<P} \exp[-\theta_{\mathbf{p}} T_{\mathbf{p}}])\,|0\rangle$$
$$- (\prod_{|\mathbf{p}|<P'} \exp[-\theta_{\mathbf{p}} T_{\mathbf{p}}])\,|0\rangle\| \longrightarrow_{P,P'\to\infty} 0 \quad (4.9)$$

holds. According to Eq. (4.7), the left-hand side turns out to be $2\{\exp[\sum_{P\le|\mathbf{p}|<P'} \log\cos\theta_{\mathbf{p}}] - 1\}$ showing that the necessary and sufficient condition for (4.9) is given by the convergence

$$|\sum_{\mathbf{p}} \log\cos\theta_{\mathbf{p}}| < \infty. \quad (4.10)$$

This condition will be satisfied if, as is expected from its nature as Fourier transform, $\tilde{\Delta}(\mathbf{p})$ does not increase as $|\mathbf{p}| \to \infty$, because $\log\cos\theta_{\mathbf{p}}$ is then approximated by $-\frac{1}{2}|\tilde{\Delta}(\mathbf{p})/\epsilon(\mathbf{p})|^2 \sim o(1/|\mathbf{p}|^4)$. Assuming this to be true, we are assured that the physical vacuum $|\Omega(V)\rangle$ as well as the excited states $\prod_k \gamma_{1\mathbf{p}_k}^{*} \prod_l \gamma_{2\mathbf{p}_l}^{*}\,|\Omega(V)\rangle$ exists in the Fock space of $a_{i\mathbf{p}}$'s. The transformation $\prod_{\mathbf{p}} \exp[-\theta_{\mathbf{p}} T_{\mathbf{p}}]$ is unitary.

B

Now, we want to see what will happen in the limit $V \to \infty$. Two things are to be investigated, $\lim_{V\to\infty} |\Omega(V)\rangle$ and the $\lim_{V\to\infty} K'(V)$, preferably in the sense of strong convergence. Let us study the former first. A catastrophe is expected however, since the condition (4.10) is violated as $V \to \infty$:

$$\sum_{\mathbf{p}} \log\cos\theta_{\mathbf{p}}$$
$$\sim \frac{V}{(2\pi)^3} \int \log\cos\theta_{\mathbf{p}}\, d\mathbf{p} \to -\infty. \quad (4.11)$$

It must be recalled here that the representation space $\mathfrak{H}_a(V)$ of $a_{i\mathbf{p}}$'s, the fields restricted to V, changes as V increases in such a way that the limit $V \to \infty$ cannot be described by any unitary transformation. Within the representation of $a_{i\mathbf{p}}$'s used above, we cannot ask the motion of $|\Omega(V)\rangle$ induced by the volume increase $V \to \infty$.[22]

Then let us consider the Fock representation of $a_{i\mathbf{p}}$'s as embedded in the Fock representation of the fields $\psi_i(f_n)$'s in the infinite volume. The space $\mathfrak{H}_a(\infty)$ of the latter is simply spanned by the bases (see Sec. 2),

$$\psi_1^{*}(f_n)\psi_1^{*}(f_{n'}) \cdots \psi_2^{*}(g_m)\psi_2^{*}(g_{m'}) \cdots |0\rangle, \quad (4.12)$$

of which the vacuum $|0\rangle$ is taken to be the same as above. In terms of these, the restricted fields, $a_{i\mathbf{p}}$ for example is given by

$$a_{1\mathbf{p}}^{(V)} = \frac{1}{V^{\frac{1}{2}}} \int_V e^{-i\mathbf{p}\mathbf{x}} \sum_n f_n(\mathbf{x})\, d\mathbf{x} \int_\infty f_n^{*}(\mathbf{y})\psi_1(\mathbf{y})\, d\mathbf{y}, \quad (4.13)$$

where the last integral is nothing but the $\psi_1(f_n)$. In view of the closure property of $\{f_n(\mathbf{x})\}$, we see that the right-hand side is actually equal to $a_{i\mathbf{p}}$ if acting upon any state confined in V, $\mathbf{y} \in V$, thus confirming the definition above as extension.[23] And all the operators used above can be extended to the operators in $\mathfrak{H}_a(\infty)$ by replacing $a_{i\mathbf{p}}$'s by such expressions as on the right-hand side of (4.13). The physical vacuum of the extention of $K'(V)$ is given by (4.8) with the operator $T_{\mathbf{p}}$ also extended:

$$|\Omega(V)\rangle = \prod_{\mathbf{p}}^{(V)} \exp[-\theta_{\mathbf{p}}^{(V)} T_{\mathbf{p}}^{(V)}]\,|0\rangle. \quad (4.14)$$

The operator $K'(V)$ is thus extended all over the Hilbert space $\mathfrak{H}_a(\infty)$ but is still restricted physically[24] to V. The dependence on V is then emphasized by the superscripts in (4.14). The Cauchy condition similar to (4.9) leads us to the condition (4.10), as expected, thus guaranteeing the existence of $|\Omega(V)\rangle$ in $\mathfrak{H}_a(\infty)$.

Now we ask the existence of $\lim_{V\to\infty} |\Omega(V)\rangle$. The Cauchy condition, $\| |\Omega(V)\rangle - |\Omega(V')\rangle \| \to 0$ as V, $V' \to \infty$, does not hold since, as seen in (4.11), $\sum_{\mathbf{p}}^{(V)} \log\cos\theta_{\mathbf{p}}^{(V)} \to -\infty$ as $V \to \infty$. While this shows that $|\Omega(V)\rangle$ does not converge strongly, we can also see that the weak limit of $|\Omega(V)\rangle$ is zero. Now, if $\lim_{V\to\infty} |\Omega(V)\rangle$ together with $\lim_{V\to\infty} K'(V)$ existed, the former could give the vacuum state of the latter. Since we have seen, however, that this is not the case, we now look for some representations suitable for the case of the infinite volume. One should still keep it in mind that, as far as $V < \infty$, the vacuum $|\Omega(V)\rangle$ of $\gamma_{i\mathbf{p}}$ is unitarily connected to the vacuum $|0\rangle$ of $a_{i\mathbf{p}}$'s and therefore the Fock representation of $\gamma_{i\mathbf{p}}$'s is unitarily equivalent to that of $a_{i\mathbf{p}}$'s.

C

The candidate we think of immediately is the Fock representation of $\gamma_{i\mathbf{p}}$'s defined in (4.2). Here-

[21] Since the set of momenta $\mathbf{p} = (n_1, n_2, n_3)\,\pi/l$ is countable, we can make the ordering of p's.

[22] See the latter half of the introduction.

[23] It is zero if acting on the states located outside of V.

[24] This means that the $K'(V)$ consists of such operators as $\psi(f)$ whose smearing function has the support only in V.

after, Eq. (4.2) should be regarded as defining a_{ip}'s in terms of γ_{ip}'s. Suppose that we can handle the convergence question of $K'(V)$ as $V \to \infty$ within the Fock representation of γ_{ip}'s, $\mathfrak{D}_F(\gamma)$, then the merit is that we can keep the representation space fixed to be the Fock space of γ_{ip}'s: As $V \to \infty$, the field operators ψ_i [written in terms of $a_{ip}^{(V)}$] do change and so do the apparent forms of the Wightman functions, yet the representation looked at as $\mathfrak{D}_F(\gamma)$ remains to be one and the same. It is the representation of a_{ip}'s and hence of the electron operators ψ_i that changes. Because the Fock representation of γ_{ip}'s, $\mathfrak{D}_F(\gamma)$, is unitary equivalent to that of a_{ip}'s, $\mathfrak{D}_F(a)$ so far as $V < \infty$, we can compute the Wightman functions describing the former by employing the latter; this is the major point of our argument. Summarizing, we can say that the Wightman functions computed with $\mathfrak{D}_F(a)$ for $V < \infty$ will lead us, as $V \to \infty$, to the right representation of field operators that suits to $K'(\infty)$ if $K'(\infty) = \lim_{V\to\infty} K'(V)$ exists within $\mathfrak{D}_F(\gamma)$. It is not superfluous to emphasize again that the whole limiting procedure is carried out in one and the same space, the Fock space of γ_{ip} [more precisely, the Fock space of $\Psi_i(f_n)$ that will be introduced below]; we could not say so if we used the Fock space of a_{ip}. It can be done in the same way as before by introducing new fields $\Psi_i(\mathbf{x})$ for γ_{ip}'s that correspond to the $\psi_i(\mathbf{x})$ for a_{ip}'s. The extension of γ_{ip} is defined by an equation similar to (4.13):

$$\gamma_{ip}^{(V)} = \frac{1}{V^{\frac{1}{2}}} \sum_n c_{np}\Psi_i(f_n), \quad c_{np} \equiv \int_V e^{-ipx} f_n(\mathbf{x})\, d\mathbf{x}. \tag{4.15}$$

This extension is practically unique as far as we want the Fock representation of $\Psi_i(\mathbf{x})$. The extended "Hamiltonian" is defined as

$$\begin{aligned} K'(V) &= \sum_p E(\mathbf{p})[\gamma_{1p}^{(V)*}\gamma_{1p}^{(V)} + \gamma_{2p}^{(V)*}\gamma_{2p}^{(V)}] \\ &= \frac{1}{V}\sum_p E(\mathbf{p}) \sum_{n,m} c_{np}^* c_{mp}[\Psi_1^*(f_n)\Psi_1(f_m) \\ &\quad + \Psi_2^*(f_n)\Psi_2(f_m)], \end{aligned} \tag{4.16}$$

and we use the Fock space of $\Psi_i(f_n)$ constructed from the common vacuum $|\Omega\rangle$ of $\gamma_{ip}^{(V)}$ and $\Psi_i(f_n)$. In this form we can ask the convergence of $K'(V)$.

Defining $K'(\infty)$ by

$$\begin{aligned} K'(\infty) = \frac{1}{(2\pi)^3}\int d\mathbf{p}E(\mathbf{p}) \sum_{n,m} c_n^*(\mathbf{p})c_m(\mathbf{p}) \\ \times [\Psi_1^*(f_n)\Psi_1(f_m) + \Psi_2^*(f_n)\Psi_2(f_m)] \end{aligned} \tag{4.17}$$

with

$$c_n(\mathbf{p}) \equiv \int_\infty e^{-i\mathbf{px}} f_n(\mathbf{x})\, d\mathbf{x}, \tag{4.18}$$

we shall show the strong convergence

$$\|\{K'(V) - K'(\infty)\}\,|\chi\rangle\|^2 \to 0 \quad \text{as} \quad V \to \infty \tag{4.19}$$

for any vector $|\chi\rangle$ from the dense set consisting of all the linear combinations of finite numbers of the basis $\Psi_1^*(f_n)\Psi_1^*(f_m)\cdots\Psi_2^*(f_k)\Psi_2^*(f_l)\cdots|\Omega\rangle$. For this purpose, we have only to compute $\{K'(V)-K'(\infty)\}^2$ as the sum of the following:

$$\begin{aligned} K'(V)^2 &= \frac{1}{V}\sum_p E^2(\mathbf{p}) \\ &\quad \times \sum_{n,m} c_{np}^* c_{mp}[\Psi_1^*(f_n)\Psi_1(f_m) + (1 \to 2)] \\ &\quad - \frac{1}{V^2}\sum_{p,q} E(\mathbf{p})E(\mathbf{q}) \sum_{n,m,k,l} c_{np}^* c_{mp} c_{kq}^* c_{lq} \\ &\quad \times [\Psi_1^*(f_n)\Psi_1^*(f_k)\Psi_1(f_m)\Psi_1(f_l) + (1 \to 2)], \end{aligned}$$

$K'(\infty)^2$ = expression obtained from the above by an obvious replacement,

$$\begin{aligned} -K'(V)K'(\infty) \\ = -\frac{1}{V}\sum_p \frac{1}{(2\pi)^3}\int d\mathbf{q}\ \delta_V(\mathbf{p}-\mathbf{q})E(\mathbf{p})E(\mathbf{q}) \\ \times \sum_{n,l} c_{np}^* c_l(\mathbf{q})[\Psi_1^*(f_n)\Psi_1(f_l) + (1 \to 2)] \\ + \frac{1}{V}\sum_p \frac{1}{(2\pi)^3}\int d\mathbf{q}E(\mathbf{p})E(\mathbf{q}) \sum_{n,m,k,l} c_{np}^* c_{mq} c_k^*(\mathbf{q}) c_l(\mathbf{q}) \\ \times [\Psi_1^*(f_n)\Psi_1^*(f_k)\Psi_1(f_m)\Psi_1(f_l) + (1 \to 2)], \end{aligned}$$

$K(\infty)K'(V)$ = expression similar to the above, where we have used such relations as

$$\begin{aligned} \sum_m c_{mp}c_{mq}^* = \int_V d\mathbf{x}\int_V d\mathbf{y} e^{-i(\mathbf{px}-\mathbf{qy})} \\ \times \sum_m f_m(\mathbf{x})f_m^*(\mathbf{y}) = V\delta_{\mathbf{pq}} \end{aligned}$$

for $\mathbf{p}$, $\mathbf{q}$ that are multiples of π/l, and

$$\begin{aligned} \sum_m c_{mp}c_m^*(\mathbf{q}) = \int_V d\mathbf{x}\int_\infty d\mathbf{y} e^{-i(\mathbf{px}-\mathbf{qy})} \sum_m f_m(\mathbf{x})f_m^*(\mathbf{y}) \\ = \int_V d\mathbf{x} e^{-i(\mathbf{p}-\mathbf{q})\mathbf{x}} \equiv (2\pi)^3\delta_V(\mathbf{p}-\mathbf{q}), \end{aligned}$$

the last equality being the definition of δ_V. Because $|\chi\rangle$ is the linear combination of the finite number of the basis, $\langle\chi|\ \{K'(V) - K'(\infty)\}^2\ |\chi\rangle$ picks up a finite number of terms from the $n, \cdots, l$ summations in the above. The convergence of the summations or the integrations over momenta can be secured by taking as $\{f_n(\mathbf{x})\}$ the energy eigenfunctions of a three-dimensional harmonic oscillator. Now, in the limit $V \to \infty$, $\delta_V(\mathbf{p})$ goes to $\delta(\mathbf{p})$ and the summation over $\mathbf{p}$ goes to integration thus confirming the convergence (4.19). The strong convergence (4.19)

can obviously be extended to the common domain of $K'(\infty)$ and $K'(V)$ (for all V). For the sake of the further extension, we have to invoke the unitary trick by considering $\exp[iK'(\infty)]$ and $\exp[iK'(V)]$.

It should be remarked here that, throughout the limiting procedure $V \to \infty$, the vacuum stays invariant; in other words, $|\Omega\rangle$ is the vacuum common to $K(V)$ of any V.

5. WIGHTMAN FUNCTIONS

In the last section, we have seen that $K'(V)$ of sufficiently large volume V can be an approximation to $K'(\infty)$ in a special representation. It is quite well-known that, in quantum field theory, Hamiltonian often requires a particular representation,[1-3] which is usually very difficult to find. The interest in our case is that the "Hamiltonian" is diagonalizable within the Fock representation of $\psi_i(\mathbf{x})$ if we restrict the volume V to be finite; the limit of eigenvectors as $V \to \infty$ does not remain in the Fock space. Then, remembering that the Wightman functions determine the representation completely,[9] we are led to a question naturally: Is it possible to obtain a right representation by a limiting process $V \to \infty$ from the Wightman functions for a finite volume V, which will, in principle, be calculable in some cases? In some other cases, the limiting procedures would be taken with respect to some parameters (e.g., cutoff momentum of the interaction in the case of the Van Hove model[1]) other than the space volume. Anyhow, if we could obtain the physical vacuum by restricting some parameters, say V, to finite values, we might be able to calculate the Wightman functions employing simply the Fock representation, then to examine their limit at $V \to \infty$ possibly arriving at a right representation.

There are at least two profitable aspects in this procedure: (i) The Wightman functions thus obtained as a limit satisfy the positivity condition automatically; they satisfy the condition for any values of the restricted parameters obviously, hence they do also in the limit. We come back to this point later. (ii) The expression of the Hamiltonian formally obtained by taking the limit of parameters can be legitimately used in the representation arrived at through the procedure above; all the matrix elements of the Hamiltonian are correctly obtained through the formal expression. The illustration in the following would be enough to show this generally. An undesirable feature is the possible reducibility of the limit representation. However, we see in the example in Sec. 6 that the way of reduction is suggested by the limiting procedure itself.

In our example of the BCS model, the physical vacuum is given by (4.8) for a finite volume V. Then, we can calculate the Wightman functions,

$$W^{(V)}(\cdots \mathbf{x}_2, \cdots, \mathbf{x}_1, \mathbf{y}_1, \cdots, \mathbf{y}_2, \cdots) = \langle \Omega(V)| \cdots \psi_2(\mathbf{x}_2) \cdots \psi_1(\mathbf{x}_1) \times \psi_1^*(\mathbf{y}_1) \cdots \psi_2^*(\mathbf{y}_2) \cdots |\Omega(V)\rangle, \tag{5.1}$$

by using the expansion of $\psi_i(\mathbf{x})$ in terms of $a_{i\mathfrak{p}}$. The point is that, when we restrict the testing functions to the chopped-off ones, (2.7), we can use (2.9) as the basis. For instance,

$$W^{(V)}(\mathbf{x}_2, \mathbf{x}_1) \equiv \langle \Omega(V)| \psi_2(\mathbf{x}_2)\psi_1(\mathbf{x}_1) |\Omega(V)\rangle = -\frac{1}{V}\sum_{\mathfrak{p}} e^{i\mathfrak{p}(\mathbf{x}_2-\mathbf{x}_1)} \sin\theta_{\mathfrak{p}} \cos\theta_{\mathfrak{p}}, \tag{5.2}$$

$$W^{(V)}(\mathbf{x}_i, \mathbf{y}_j) \equiv \langle \Omega(V)| \psi_i(\mathbf{x}_i)\psi_j^*(\mathbf{y}_j) |\Omega(V)\rangle = \frac{1}{V}\sum_{\mathfrak{p}} e^{i\mathfrak{p}(\mathbf{x}_i-\mathbf{y}_j)} \cos^2\theta_{\mathfrak{p}}, \tag{5.3}$$

$$W^{(V)}(\mathbf{x}_1, \mathbf{x}_1', \mathbf{y}_1, \mathbf{y}_1') \equiv \langle \Omega(V)| \psi_1(\mathbf{x}_1)\psi_1(\mathbf{x}_1')\psi_1^*(\mathbf{y}_1)\psi_1^*(\mathbf{y}_1') |\Omega(V)\rangle = W^{(V)}(\mathbf{x}_1, \mathbf{y}_1')W^{(V)}(\mathbf{x}_1', \mathbf{y}_1) - W^{(V)}(\mathbf{x}_1, \mathbf{y}_1)W^{(V)}(\mathbf{x}_1', \mathbf{y}_1'). \tag{5.4}$$

These functions coincide with the corresponding ones calculated in the Fock representation of $\gamma_{i\mathfrak{p}}$'s, because the Fock representations of $a_{i\mathfrak{p}}$'s and of $\gamma_{i\mathfrak{p}}$'s are unitary equivalent as far as $V < \infty$. As we have seen in the previous section, the latter goes to the representation proper to $\lim_{V\to\infty} K'(V) = K'(\infty)$, so the above functions go to the right ones at $V \to \infty$. In general, the representation proper to the "Hamiltonian" $\mathfrak{K} - \mu N$ in the infinite volume is realized by the system of Wightman functions

$$\langle \Omega(\infty)| \cdots \psi_2(\mathbf{x}_2) \cdots \psi_1(\mathbf{x}_1) \times \psi_1^*(\mathbf{y}_1) \cdots \psi_2^*(\mathbf{y}_2) \cdots |\Omega(\infty)\rangle \equiv \lim_{V\to\infty} \langle \Omega(V)| \cdots \psi_2(\mathbf{x}_2) \cdots \psi_1(\mathbf{x}_1) \times \psi_1^*(\mathbf{y}_1) \cdots \psi_2^*(\mathbf{y}_2) \cdots |\Omega(V)\rangle, \tag{5.5}$$

as defined by the volume limiting process.

The matrix element of the "Hamiltonian" can be represented in terms of the Wightman functions, in particular simply by using the expression (4.3) in the Fock representation of $\gamma_{i\mathfrak{p}}$'s. It is now obvious that the same matrix elements are obtained also by using (4.1) and (3.2), e.g.,

$$\langle \Omega(\infty)| \cdots \psi_2(\mathbf{x}_2) \cdots \psi_1(\mathbf{x}_1) \times \mathfrak{K}_I\psi_1^*(\mathbf{y}_1) \cdots \psi_2^*(\mathbf{y}_2) \cdots |\Omega(\infty)\rangle$$

$$\equiv \lim_{V\to\infty} \frac{1}{V} \int d\mathbf{X}_1\, d\mathbf{X}_2\, d\mathbf{Y}_1\, d\mathbf{Y}_2 v(\mathbf{Y}_2 - \mathbf{Y}_1, \mathbf{X}_2 - \mathbf{X}_1)$$
$$\times \langle \Omega(\infty)| \cdots \psi_2(\mathbf{x}_2) \cdots \psi_1(\mathbf{x}_1)\psi_1^*(\mathbf{Y}_1)\psi_2^*(\mathbf{Y}_2)\psi_2(\mathbf{X}_2)$$
$$\times \psi_2(\mathbf{X}_1)\psi_1^*(\mathbf{y}_1) \cdots \psi_2^*(\mathbf{y}_2) \cdots |\Omega(\infty)\rangle, \qquad (5.6)$$

where $\langle \Omega(\infty)| \cdots |\Omega(\infty)\rangle$ is used to represent the limit Wightman functions. We have to note here the following relation that can be proved by a straightforward computation[15]:

$$\lim_{V\to\infty} \langle \Omega(V)|\, Q \cdot \frac{1}{V} \int_V \psi_2(\mathbf{y} + \mathbf{n})\psi_1(\mathbf{y})\, d\mathbf{y} \cdot Q' \,|\Omega(V)\rangle$$
$$= \lim_{V\to\infty} \langle \Omega(V)|\, \frac{1}{V} \int_V \psi_2(\mathbf{y} + \mathbf{n})\psi_1(\mathbf{y})\, d\mathbf{y} \,|\Omega(V)\rangle$$
$$\times \langle \Omega(V)|\, QQ' \,|\Omega(V)\rangle, \qquad (5.7)$$

where Q and Q' are polynomials of field operators. This gives the basis in terms of the Wightman functions for the replacement of $\mathfrak{K}_I$ by the quadratic one.

In the sense described above we can conclude that the limit Wightman functions give the representation proper to our "Hamiltonian" $\mathfrak{K} - \mu N$ [see (3.1) and (4.1)]. The question raised at the beginning of this section is thus answered affirmatively in the case of the BCS model.

We have to notice here that it remains still to ask what representation of $\psi_i(x)$ is actually implied by the system of Wightman functions (5.5). We know that it implies the Fock representation for $\gamma_{i\mathbf{p}}^{(\infty)}$ (therefore the representation is irreducible), but what representation does it imply for $a_{i\mathbf{p}}^{(\infty)}$? This question entails asking what is meant by the relation (4.2) between $a_{i\mathbf{p}}$'s and $\gamma_{i\mathbf{p}}$'s. When we wrote down this relation, we were using the Fock representation of $a_{i\mathbf{p}}$'s, which, due to $V < \infty$, was unitary equivalent to the Fock representation of $\gamma_{i\mathbf{p}}$'s; $a_{i\mathbf{p}}$'s and $\gamma_{i\mathbf{p}}$'s could be regarded as operating in one and the same space. Since in the limit $V \to \infty$, the operator $a_{i\mathbf{p}}$ has no vacuum state in the Fock space of $\gamma_{i\mathbf{p}}$'s, we are compelled to ask the above questions. Of course, we can define $a_{i\mathbf{p}}$ through the Fock representation of $\gamma_{i\mathbf{p}}$'s—in other words, by the linear combination of $\gamma_{i\mathbf{p}}$'s and $\gamma_{i\mathbf{p}}^*$'s—the relation inverse to (4.2). But what we want to know is the implication of this definition to the Gårding–Wightman classification[8] of the representation. These questions are left for future investigation.

6. THE REPRESENTATION WITH SHARP PARTICLE NUMBER

The "Hamiltonian" $K(V) = \mathfrak{K}(V) - \mu N(V)$ we have started with is obviously invariant under the gauge transformation,

$$\psi_i(\mathbf{x}) \to \psi_i'(\mathbf{x}) = e^{-i\alpha/2}\psi_i(\mathbf{x}) \qquad (i = 1, 2), \qquad (6.1)$$

which can be effected by an unitary transformation within the Fock representation of $\psi_i(\mathbf{x})$,

$$\psi_i'(\mathbf{x}) = \mathfrak{U}\psi_i(\mathbf{x})\mathfrak{U}^*, \qquad \mathfrak{U} \equiv \exp\left[\tfrac{1}{2}i\alpha N(V)\right] \qquad (6.2)$$

as far as $V < \infty$.

If however, we replace $K(V)$ by an effective one $K'(V)$ on the assumption that an irreducible representation is taken for the field operators, we lose the invariance mentioned above; $K'(V)$ and its physical vacuum are no longer invariant as seen from (3.2) and (4.8). But, why? The answer was given by Nambu[14] and Haag.[17] Nambu was the first in suggesting that the physical vacuum is highly degenerate with respect to the particle number; the degeneracy accounts for the apparent gauge noninvariance. Then, Haag has shown that an important clue to the solvability of the diagonalization problem of $K(V)$ consists in the irreducibility of the representation, which turned out to be not fulfilled by the (gauge-invariant) representation with sharp particle number. He could suggest the Wightman functions that will describe such representation, on the one hand, and the way for its reduction, on the other.

In this section we want to show that the limiting procedure described in the last section leads us, in quite an intuitive way, to the Wightman functions suggested by Haag. Proceeding in this way, we are guaranteed that the resulting system of Wightman functions satisfies the positivity condition, on the one hand, and also describes the representation proper to $K(\infty)$.

As a preparatory step, we have to see the distribution of the (bare) particle number in the physical vacuum of $\gamma_{i\mathbf{p}}$'s given by (4.8): the average

$$\langle N_1(V)\rangle_{\mathrm{av}} \equiv \langle \Omega(V)|\, \int_V \psi_1^*(\mathbf{x})\psi_1(\mathbf{x})\, d\mathbf{x} \,|\Omega(V)\rangle$$
$$= \sum_{\mathbf{p}} \sin^2 \theta_{\mathbf{p}},$$

and the dispersion

$$\langle [N_1(V) - \langle N_1(V)\rangle_{\mathrm{av}}]^2\rangle_{\mathrm{av}} = \sum_{\mathbf{p}} \sin^2 \theta_{\mathbf{p}} \cos^2 \theta_{\mathbf{p}}, \qquad (6.3)$$

and the same results are obtained for the particles 2. In the limit $V \to \infty$, these quantities diverge linearly in V. Defining the particle density by $\rho_i(V) = N_i(V)/V$, we obtain

$$\lim_{V\to\infty} \rho_i(V) = \lim_{V\to\infty} \langle \Omega(V)|\, \rho_i(V) \,|\Omega(V)\rangle$$
$$= \frac{1}{(2\pi)^3} \int \sin^2 \theta_{\mathbf{p}}\, d\mathbf{p}, \qquad (6.4)$$

and the dispersion of its distribution

$$\lim_{V\to\infty} \langle [\rho_i(V) - \langle \rho_i(V)\rangle_{\mathrm{av}}]^2\rangle_{\mathrm{av}} = 0, \tag{6.5}$$

finding that the particle density is quite sharp at $V \to \infty$.

Now, we turn to the construction of Wightman functions; the first step is to obtain the physical vacuum with the sharp particle number. Expecting that $K'(V)$ with $V \to \infty$ gives a good approximation to $K(\infty)$ which is manifestly invariant under the gauge transformation (6.1), we introduce a series of transformed vacua ($0 \leq \alpha \leq 2\pi$),

$$|\Omega_\alpha(V)\rangle = \prod_{\mathbf{p}} (\cos\theta_{\mathbf{p}} - e^{i\alpha} a_{1\mathbf{p}}^* a_{2-\mathbf{p}}^* \sin\theta_{\mathbf{p}})\,|0\rangle, \tag{6.6}$$

which, in the limit $V \to \infty$, at least formally, gives an eigenstate of $K(\infty)$, and so does their linear combination,

$$|\Omega_N(V)\rangle = \frac{1}{2\pi}\int_0^{2\pi} e^{-i\alpha N}\,|\Omega_\alpha(V)\rangle \qquad (N = \text{integer}). \tag{6.7}$$

This is an eigenstate of $N(V) = N_1(V) + N_2(V)$ with the eigenvalue $2N$. We are interested in the limit $V \to \infty$. The physical situation requires us to keep the particle density $\rho = \rho_1 + \rho_2$ fixed when we take the limit. From the discussion in Sec. 4, we already know that $\lim_{V\to\infty} |\Omega(V)\rangle$ does not exist in the sense of strong convergence or it is zero in the sense of weak convergence. Thus the argument here is quite formal but still it is sufficient to find the Wightman functions in question.

Let us ask the normalization of $|\Omega_N(V)\rangle$ in (6.7): for the case of large V, we have

$$\langle \Omega_N(V) \mid \Omega_N(V)\rangle \sim \frac{1}{(2\pi)^2}\int_0^{2\pi} d\beta \int_0^{2\pi} d\alpha \exp[NG_\rho(\alpha,\beta)], \tag{6.8}$$

where

$$G_\rho(\alpha,\beta) = i(\beta-\alpha) + \frac{1}{(2\pi)^3\rho}\int \log(\cos^2\theta_{\mathbf{p}} + e^{-i(\beta-\alpha)}\sin^2\theta_{\mathbf{p}})\,d\mathbf{p} \tag{6.9}$$

is independent of V. Since $N = \rho V$ is also very large, the saddle-point method can be applied successfully to the α integration. The saddle point is determined by

$$\frac{1}{i}\frac{\partial G_\rho}{\partial\alpha} = -1 + \frac{1}{(2\pi)^3\rho} \times \int \frac{\sin^2\theta_{\mathbf{p}} e^{-i(\beta-\alpha)}}{\cos^2\theta_{\mathbf{p}} + e^{-i(\beta-\alpha)}\sin^2\theta_{\mathbf{p}}}\,d\mathbf{p} = 0 \tag{6.10}$$

to be

$$\alpha = \beta, \tag{6.11}$$

where we have taken into account that the reality condition requires $e^{-i(\beta-\alpha)} = +1$ or -1 and that the latter gives $\partial^2 G_\rho/\partial\alpha^2 > 0$, thus being rejected. That (6.11) actually leads us to (6.10) can be seen from the fact that the particle density is quite sharp and is given by (6.4) also for any $|\Omega_\alpha(V)\rangle$ ($V \to \infty$). The function $G_\rho(\alpha,\beta)$ is zero at the saddle point, and so

$$\langle \Omega_N(V) \mid \Omega_N(V)\rangle \sim \frac{1}{(2\pi)^2}\int_0^{2\pi} d\beta \int_{-\infty}^{\infty} d\alpha \times \exp[-(\alpha-\beta)^2 N\sigma] = \frac{1}{2\pi}\left(\frac{\pi}{N\sigma}\right)^{\frac{1}{2}}, \tag{6.12}$$

where

$$\sigma = -\frac{1}{2}\left(\frac{\partial^2 G_\rho}{\partial\alpha^2}\right)_{\alpha=\beta} = \frac{1}{2}\frac{1}{(2\pi)^3\rho}\int \sin^4\theta_{\mathbf{p}}\,d\mathbf{p} > 0, \tag{6.13}$$

and the α integration is extended over an infinite range. In addition, we note the orthogonality; for $N' \neq N$

$$\langle \Omega_{N'}(V) \mid \Omega_N(V)\rangle \sim \frac{1}{(2\pi)^2}\int_0^{2\pi} d\beta e^{i(N'-N)\beta} \times \int_{-\infty}^{\infty} d\alpha \exp[-(\alpha-\beta)^2 N\sigma] = 0. \tag{6.14}$$

Now, it is natural to define the gauge-invariant Wightman functions by

$$W_{\mathrm{g}}^{(\infty)} = \lim_{\substack{V\to\infty \\ N/V=\rho}} \frac{\langle \Omega_N(V)|\cdots\psi_2(\mathbf{x}_2)\cdots\psi_1(\mathbf{x}_1)\psi_1^*(\mathbf{y}_1)\cdots\psi_2^*(\mathbf{y}_2)\cdots|\Omega_N(V)\rangle}{\langle \Omega_N(V) \mid \Omega_N(V)\rangle}, \tag{6.15}$$

which will symbolically be denoted by[25]

[25] Here, g stands for gauge invariance. Since the unitary transformation of Haag[17],

$$U^* = \lim_{V\to\infty}(1/V)\int \psi_2(\mathbf{y}+\mathbf{n})\,\psi_1(\mathbf{y})\,d\mathbf{y}/\langle\Omega(V)\,|\psi_2(\mathbf{y}+\mathbf{n})\,\psi_1(\mathbf{y})|\,\Omega(V)\rangle,$$

transforms $|\Omega_0^{(\infty)}\rangle$ into the other gauge-invariant vacuum by decreasing the total number N (actually ∞) of electrons by two, we are tempted to preserve the subscript N to the vacuum. The operator U^* commutes with $\psi_s(f)$ in the irreducible representations [see (5.5)].

$$W_g^{(\infty)} = \langle \Omega_g(\infty)| \cdots \psi_2(\mathbf{x}_2) \cdots \psi_1(\mathbf{x}_1) \times \psi_1^*(\mathbf{y}_1) \cdots \psi_2^*(\mathbf{y}_2) \cdots |\Omega_g(\infty)\rangle. \quad (6.16)$$

In order to compute these functions we use the saddle-point method again. For example, in computing

$$W_g^{(\infty)}(\mathbf{x}_1, \mathbf{y}_1) = \langle \Omega_g^{(\infty)}| \psi_1(\mathbf{x}_1)\psi_1^*(\mathbf{y}_1) |\Omega_g^{(\infty)}\rangle = \lim_{N\to\infty} \left(\frac{1}{(2\pi)^2}\frac{\pi}{N\sigma}\right)^{-\frac{1}{2}} \frac{1}{(2\pi)^2} \int_0^{2\pi} d\beta \times \int_0^{2\pi} d\alpha F(\mathbf{x}_1, \mathbf{y}_1; \alpha, \beta) \exp [NG_\rho(\alpha, \beta)],$$

we note

$$F(\mathbf{x}_1, \mathbf{y}_1; \alpha, \beta) \equiv \frac{1}{(2\pi)^3} \int \frac{e^{i\mathbf{k}(\mathbf{x}_1 - \mathbf{y}_1)} \cos^2 \theta_\mathbf{k}}{\cos^2 \theta_\mathbf{k} + e^{-i(\beta-\alpha)} \sin^2 \theta_\mathbf{k}} d_\mathbf{k}$$

is independent of N so that the saddle point is determined by $G_\rho(\alpha, \beta)$ only. Hence

$$W_g^{(\infty)}(\mathbf{x}_1, \mathbf{y}_1) = F(\mathbf{x}_1, \mathbf{y}_1; \alpha = \beta) = \lim_{V\to\infty} W^{(V)}(\mathbf{x}_1, \mathbf{y}_1), \quad (6.17)$$

where the function in the extreme right has been given by (5.2). In general, $\exp [NG_\rho(\alpha, \beta)]$ behaves like a δ function in the integral for the n-point function also ($n < \infty$). For the sake of later reference, we add one more example: in the same way as above we obtain

$$W_g^{(\infty)} = \langle \Omega_g(\infty)| \psi_2(\mathbf{x}_2)\psi_1(\mathbf{x}_1) |\Omega_g(\infty)\rangle = \frac{1}{2\pi} \int_0^{2\pi} d\beta F'(\mathbf{x}_2, \mathbf{x}_1; \alpha = \beta) = 0, \quad (6.18)$$

since the integral vanishes due to the exponential $e^{i\beta}$ in

$$F'(\mathbf{x}_2, \mathbf{x}_1; \alpha = \beta) = -e^{i\beta} \frac{1}{(2\pi)^3} \int e^{i\mathbf{k}(\mathbf{x}_2 - \mathbf{x}_1)} \sin \theta_\mathbf{k} \cos \theta_\mathbf{k} \, d\mathbf{k}.$$

We notice here that this function can be written as

$$F'(\mathbf{x}_2, \mathbf{x}_1; \alpha = \beta) = \langle \Omega_\beta(\infty)| \psi_2(\mathbf{x}_2)\psi_1(\mathbf{x}_1) |\Omega_\beta(\infty)\rangle. \quad (6.19)$$

In this way, we find generally

$$\langle \Omega_g(\infty)| \cdots \psi_2(\mathbf{x}_2) \cdots \psi_1(\mathbf{x}_1)\psi_1^*(\mathbf{y}_1) \cdots \psi_2^*(\mathbf{y}_2) \cdots |\Omega_g(\infty)\rangle$$

$$= \begin{cases} \langle \Omega(\infty)| \cdots \psi_2(\mathbf{x}_2) \cdots \psi_1(\mathbf{x}_1)\psi_1^*(\mathbf{y}_1) \cdots \psi_2^*(\mathbf{y}_2) \cdots |\Omega(\infty)\rangle, & \text{for } n(1) = n(1^*), n(2) = n(2^*), \quad (6.20a) \\ 0, \quad \text{otherwise}, & (6.20b) \end{cases}$$

where $n(1^*)$ stands for the number of ψ_1^* operator, etc., and $\langle \Omega(\infty)| \cdots |\Omega(\infty)\rangle$ is given in the previous section. These results are in accord with Haag's Ansatz.[17]

Now we prove that the positivity condition[26] is satisfied by the system (6.20). We start our argument from the case of $V < \infty$ where we are quite sure that any vector obtained from the gauge-invariant vacuum

$$|\Psi(V)\rangle = \sum c_{ff'\cdots gg'\cdots} \psi_1^*(f)\psi_1^*(f') \cdots \times \psi_2^*(g)\psi_2^*(g') \cdots |\Omega_N(V)\rangle / \langle \Omega_N(V) | \Omega_N(V)\rangle \quad (6.21)$$

has a positive norm, $\langle \Psi(V) \mid \Psi(V)\rangle \geq 0$. Because this is true for any V, on the one hand, and the limit $V \to \infty$ of $\langle \Psi(V) \mid \Psi(V)\rangle$ is well-defined in terms of the gauge-invariant Wightman functions (6.20) on the other, we can conclude that the system (6.20) satisfies the positivity condition.

Rigorously speaking, we have to notice that two limiting procedures are involved in the above argument: one is the limit $V \to \infty$ and the other is the possibly infinite number $n \to \infty$ of terms in the series (6.21). The existence of the vector (6.21) implies the existence of

$$\lim_{n\to\infty} || \, |\Psi^{(n)}(V)\rangle \, ||^2,$$

where $|\Psi^{(n)}(V)\rangle$ is obtained by truncating the series (6.21) somehow at the nth term. While we may assume further that $\lim_{V\to\infty} [\lim_{n\to\infty} || \, |\Psi^{(n)}(V)\rangle \, ||^2]$ exists, the requirement of the positivity condition is that $\lim_{n\to\infty} [\lim_{V\to\infty} || \, |\Psi^{(n)}(V)\rangle \, ||^2] \geq 0$ with the order of limiting procedures interchanged. Here, we do not enter the interesting question whether the existence of the former limit implies the exist-

[26] The positivity condition can be proved in a way different from that given in the text. Because of the orthogonality implied by (6.20b), the positivity of the norm square of (6.21) can be proved by showing

$$||\textstyle\sum^{(n)} c_{f\ldots g\ldots} \psi_1^*(f) \cdots \psi_2^*(g) \cdots |\Omega_g(\infty)\rangle||^2 \geq 0$$

for the partial sum $\sum^{(n)}$ of the terms with n creation operators applied on the vacuum. Now, the left-hand side can be rewritten by the help of (6.20a) as

$$\textstyle\sum^{(n)} c^*_{f'\ldots g'\ldots} c_{f\ldots g\ldots} \langle \Omega_g(\infty)| \cdots \psi_2(g') \cdots \psi_1(f')\psi_1^*(f) \times \cdots \psi_2^*(g) \cdots |\Omega_g(\infty)\rangle$$

$$= \textstyle\sum^{(n)} c_{f'\ldots g'\ldots} c_{f\ldots g\ldots} \langle \Omega(\infty)| \cdots \psi_2(g') \cdots \psi_1(f')\psi_1^*(f) \cdots \psi_2^*(g) \cdots |\Omega(\infty)\rangle$$

$$= ||\textstyle\sum^{(n)} c_{f\ldots g\ldots} \psi_1^*(f) \cdots \psi_2^*(g) \cdots |\Omega(\infty)\rangle||^2.$$

The last expression is by nature positive.

ence of the latter. However, we have to understand the positivity condition as implying that $\lim_{n\to\infty}(\lim_{V\to\infty} \| |\Psi^{(n)}(V)\rangle \|^2) \geq 0$ so far as the required limits exist. In this sense, the positivity condition is the necessary consequence of our limiting procedure.

The reducibility of the representation described by (6.20) was shown by Haag using the c-number property of the operator $B(\xi)$ explained before [cf. Sec. 3 and (5.7)[25]]. In our language, the reducibility can be expressed by rewriting (6.20) as

$$\begin{aligned}&\Omega_g(\infty)| \cdots \psi_2(\mathbf{x}_2) \cdots \psi_1(\mathbf{x}_1) \cdots \\ &\quad \times \psi_1^*(\mathbf{y}_1) \cdots \psi_2^*(\mathbf{y}_2) \cdots |\Omega_g(\infty)\rangle \\ &= \frac{1}{2\pi}\int_0^{2\pi} \langle\Omega_\beta(\infty)| \cdots \psi_2(\mathbf{x}_2) \cdots \psi_1(\mathbf{x}_1) \cdots \\ &\quad \times \psi_1^*(\mathbf{y}_1) \cdots \psi_2^*(\mathbf{y}_2) \cdots |\Omega_\beta(\infty)\rangle\, d\beta, \qquad (6.22)\end{aligned}$$

where use has been made of the relations such as (6.18) and (6.19). This equation shows that the space generated from $|\Omega_g(\infty)\rangle$ is decomposed into the direct sum of spaces generated by the repeated application of the creation operators $\psi_1^*(\mathbf{x})$ on $|\Omega_\beta(\infty)\rangle$'s, or more explicitly that

$$|\Omega_g(\infty)\rangle = \frac{1}{(2\pi)^{\frac{1}{2}}}\int_0^{2\pi} c(\beta)\,|\Omega_\beta(\infty)\rangle\, d\beta, \qquad |c(\beta)| = 1, \qquad (6.23)$$

and $|\Omega_\beta(\infty)\rangle$'s cannot be connected with each other by the finite number of field operators. It might be worthwhile to note here that in contrast to the normalization $\langle\Omega_\beta(V) \mid \Omega_\beta(V)\rangle = 1$ for any V we have $\lim_{V\to\infty}\langle\Omega_\alpha(V) \mid \Omega_\beta(V)\rangle = 0$ $(\alpha \neq \beta)$, and moreover that, as seen from (6.8) *et seq.*,

$$\lim_{\substack{V\to\infty \\ N/V=\rho}} e^{iN(\alpha-\beta)}\left(\frac{1}{(2\pi)^2}\frac{\pi}{N\sigma}\right)^{-\frac{1}{2}}\langle\Omega_\alpha(V) \mid \Omega_\beta(V)\rangle = \delta(\alpha-\beta), \qquad (6.24)$$

although the implication of these relations for (6.23) is not precise mathematically.

Finally, that the gauge-invariant Wightman functions (6.20) are suitable for the BCS Hamiltonian is seen after reduction by the use of the argument in the last section.

7. DISCUSSION

As a possible method of obtaining the representation suitable for a given Hamiltonian in quantum field theory, we have examined the limiting procedure proposed by Araki and others for a slightly different purpose.[4,5] The tool is the system of Wightman functions,[9] which can be computed when we specify the cyclic vector (vacuum) to make the expectation values. Taking the examples of free-particle gases, Araki and others specified the vacuum in terms of the thermodynamical quantities, particle density and temperature. If the volume of the gas is finite, the total number of particles is also finite and the Fock representation can be used to compute the Wightman functions. Then, in the limit of the infinite volume, the limit of Wightman functions describe certain non-Fock representation, the analysis of which is the main theme of their investigation.

In the quantum theory of interacting fields, it has been known that a particular Hamiltonian often requires a particular representation.[1-3] On the contrary, in the quantum mechanics of particles, where the degrees of freedom are finite, each representation of the canonical commutation relation is known to be unitary-equivalent to every other[27,28] we can handle the problem with any one of them. Then, the usual method of diagonalizing a given Hamiltonian, for example, is as follows: Starting with any representation one likes, one computes the matrix of the Hamiltonian operator then diagonalizing the matrix by the principal axis transformation. The same method cannot be used in the quantum theory of fields since such a principal-axis transformation does not necessarily exist in an arbitrarily chosen representation.[1] We have to choose *ab initio* the representation suitable for the Hamiltonian. But, in general, this task is difficult.

In some cases, however, if we restrict some parameters—e.g., the cutoff momentum in the Van Hove model and the space volume in the BCS model—to be finite, then the degrees of freedom become practically finite and we can diagonalize the Hamiltonians within the simple Fock representation which is traditional and easy to handle. Thus, the idea of this paper is: Compute the Wightman functions for the finite parameters, say V, using the physical vacuum as the cyclic vector. Then, the limit $V \to \infty$ of the Wightman functions will give you the representation required by the Hamiltonian. The point is, of course, that the physical vacuum for the re-

[27] J. von Neumann, Math. Ann. **104**, 570 (1931); F. Rellich, Nachr. Göttingen **107**, (1946).

[28] Even in the (one-degree-of-freedom)-particle quantum mechanics, we meet the examples of the inequivalence, but they are more or less exceptional physically. An example: the pairs of canonical variables defined in the usual fashion for the particles—each one in an infinitely deep square-well potential equal in size but different in the location of their center—are not equivalent to each other. The author owes this remark to Dr. E. J. Woods.

stricted parameters, e.g., $V < \infty$, would be obtained within a simple representation. If this is the case, the quantum field theory can be regarded as the limiting case of the quantum mechanics of particles. And, in fact, we could carry out this program for the BCS model of superconductivity. This method has been found useful to analyze the gauge-invariance problem of the model.

We have to admit that, if the limit of the Wightman function is obtained, the analysis of them constitute a difficult task because not much has been known about the Wightman functions. Yet, the transition from the finite to the infinite degrees of freedom is interesting and worth studying.

ACKNOWLEDGMENTS

In conclusion, the author would like to express his sincere thanks to Professor J. S. Toll and other members of the Department of Physics and Astronomy, University of Maryland for their helpful discussions. It is a pleasure to thank Professor O. W. Greenberg for his careful reading of the manuscript and valuable advice.

The author is grateful to the National Science Foundation and the United States Air Force Office of Scientific Research for the financial support. Thanks are also due to the Fulbright Committee for the Travel Grant which made it possible for the author to visit the University of Maryland.

Note added in proof. The construction of the gauge-invariant Wightman functions through the same procedure as ours is also suggested by L. Leplae and H. Umezawa, preprint (1964). The author is grateful to Professor Umezawa for sending him the preprint.

Linear Heisenberg Model of Ferro- and Antiferromagnetism*

SHIGETOSHI KATSURA AND SAKARI INAWASHIRO
Department of Applied Science, Tohoku University, Sendai, Japan
(Received 3 December 1963; revised manuscript received 18 February 1964)

The partition function of the one-dimensional Heisenberg model is considered. Hamiltonian of the system

$$H = -\tfrac{1}{2}\sum [J_\perp(\sigma_l^x\sigma_{l+1}^x + \sigma_l^y\sigma_{l+1}^y) + J_\parallel\sigma_l^z\sigma_{l+1}^z] - m\mathcal{H}\sum \sigma_l^z$$

is expressed in terms of Fermi operators. The term which contains $J_\parallel$, the quartic term and a part of quadratic term in Fermi operators, have been regarded as perturbation, keeping the symmetry with respect to the magnetic field. Linked-cluster expansion in an appropriate form for this case has been developed and the partition function has been obtained up to the third order in $J_\parallel$. Numerical values of energy, specific heat, and susceptibility up to second order in $J_\parallel$ are shown. The ground-state energy is

$$\frac{E}{N|J_\perp|} = -\frac{2}{\pi} - \frac{2}{\pi^2}\left(\frac{J_\parallel}{|J_\perp|}\right) - \frac{16}{\pi^3}\left(\frac{1}{6} - \frac{\pi^2}{144}\right)\left(\frac{J_\parallel}{J_\perp}\right)^2 + O\left[\left(\frac{J_\parallel}{J_\perp}\right)^3\right].$$

$E/N|J_\perp|$ for the antiferromagnetic case $J_\parallel = -|J_\perp| = J$ is -0.8899. Agreement with the exact value, -0.8863, is quite satisfactory.

1. INTRODUCTION

THE one-dimensional Heisenberg model of magnetism has been investigated by many authors. The interest in it arises from several aspects.

First, the theory of one-dimensional Heisenberg model is a touchstone of that of two- and three-dimensional models. High-temperature expansion of the free energy and the susceptibility were obtained by Kubo, Obata, and Ohno,[1] Brown and Luttinger,[2] and by Rushbrooke and Wood.[3] Low-temperature properties of this model by various investigations, especially on $\langle S^z_{\text{sublattice}}\rangle$ and $\chi(0)$, are sometimes contradictory to each other (see Oguchi[4]). Usually the Heisenberg ferromagnet is said to have no spontaneous magnetization in one and two dimensions since the Bloch's integral expressing the magnetization diverges.[5] This divergence is caused by using Bose statistics and the situation is similar also to that in Dyson's theory.[6] On the other hand, theories by Frank[7] and Mannari[8] predicted the existence of the spontaneous magnetization by using Fermi statistics. Real nonexistence of the spontaneous magnetization should be proved not by the divergence but by the vanishing of the integral expressing the magnetization. The exact proof of the existence or nonexistence of the phase change in the Heisenberg ferro- and antiferromagnet has not yet been presented in any dimension.

Second, the one-dimensional Heisenberg model has its own interest. Bethe[9] solved the ground state of the ferromagnetic linear chain. The exact value of the ground-state energy of antiferromagnetic linear chain was obtained by Bethe[10] and by Hulthén.[11] Orbach[12] and Walker[13] obtained the ground-state energy of the generalized model of Kasteleijn.[14] The exact partition function or exact ground-state wavefunction is not yet known, even in one-dimension.

Third, problems of linear chain have become interesting because real substances which can be regarded as practically linear in their magnetic structures have been found [$Cu(NH_3)_4SO_4H_2O$, Haseda and Miedema[15]; $CuCl_2$, and $CrCl_2$, Stout *et al*[16]; $Cu(NO_3)_2$

* This work has been supported in part by a research grant from the National Science Foundation and in part by that from the Ministry of Education in Japan.

[1] R. Kubo, Y. Obata and A. Ohno, Busseiron Kenkyu No. 43, 22 (1951); No. 47, 35 (1952); No. 57, 45 (1952).
[2] H. A. Brown and J. M. Luttinger, Phys. Rev. **100**, 685 (1955).
[3] G. S. Rushbrooke and P. J. Wood, Proc. Phys. Soc. (London) **68**, 1161 (1955); see C. Domb, Advan. Phys. 9, 329 (1960).
[4] T. Oguchi, J. Phys. Chem. Solids, **24**, 1049 (1963).
[5] F. Bloch, Z. Physik. **61**, 206 (1930).
[6] F. J. Dyson, Phys. Rev. **102**, 1217, 1230 (1956).
[7] D. Frank, Z. Physik. **146**, 615 (1956).
[8] I. Mannari, Progr. Theoret. Phys. (Kyoto) **19**, 201 (1958).
[9] H. A. Bethe, Z. Physik. **71**, 205 (1931).
[10] A. Sommerfeld and H. A. Bethe, *Handbuch der Physik*, edited by H. Geiger and K. Scheel (Springer-Verlag, Berlin, 1933), Vol. 24, Part 2, p. 618.
[11] L. Hulthén, Arkiv Mat. Astron. Fysik, **26A**, No. 11, 1 (1938).
[12] R. Orbach, Phys. Rev. **112**, 309 (1958).
[13] L. R. Walker, Phys. Rev. **116**, 1089 (1959).
[14] P. W. Kasteleijn, Physica **18**, 104 (1952).
[15] T. Haseda and A. R. Miedema, Physica **27**, 1102 (1961); A. R. Miedema, H. van Kempen, H. Haseda, and W. J. Huiskamp, Physica **28**, 119 (1962); H. Hasaeda and H. Kobayashi, J. Phys. Soc. Japan, to be published.
[16] R. C. Chisholm and J. W. Stout, J. Chem. Phys. **36**, 972 (1962); J. W. Stout and R. C. Chisholm, *ibid.* **36**, 979 (1962).

$2.5H_2O$, Berger *et al.*[17]] Haseda's data have been interpreted as a Heisenberg linear chain by Griffiths.[18] A general review of the linear chain from the experimental side was given by Haseda.[19]

In a former paper[20,21] by one of the authors, the partition function under finite magnetic field, the ground-state wavefunction, and the ground-state energy of the one-dimensional Heisenberg model with $S = 1/2$, of which the Hamiltonian is given by

$$H = -\tfrac{1}{2}\sum_{l=1}^{N}(J_x\sigma_l^x\sigma_{l+1}^x + J_y\sigma_l^y\sigma_{l+1}^y + J_z\sigma_l^z\sigma_{l+1}^z) - m\mathfrak{H}\sum_{l=1}^{N}\sigma_l^z \qquad (1.1)$$

(a generalization of Kasteleijn's model), has been examined for the special cases (i) $J_x=J_y=J$, $J_z=0$; (ii) $J_x = J$, $J_y = J_z = 0$; (iii) $J_x = J_y = 0$; (iv) $J_x = J_y = J_z$. Here J_x, J_y, and J_z are Cartesian components of the exchange integral, σ^x, σ^y, σ^z are Pauli spin operators, $m = \frac{1}{2}g\mu_B$, g is the g factor, μ_B is the Bohr magneton, and $\mathfrak{H}$ is the external magnetic field. Cases (i) and (ii) were treated exactly with finite magnetic field and perpendicular susceptibilities were obtained. These results do not depend on the sign of $J_\perp$. Correction of numerical calculation in susceptibility in the case (i) shows that both (i) and (ii) have perpendicular susceptibilities. (Ref. 21, and also see Fig. 12 in this paper). Long- and short-range order of Cases (i) and (ii) was discussed by Lieb, Schultz, and Mattis.[22]

In this paper, the case $J_x=J_y=J_\perp$, $J_z=J_\parallel$ has been treated by a perturbation method. The terms containing $J_\perp$ and $J_\parallel$ are regarded as an unperturbed and perturbed parts, respectively. The reason is that the usual procedure where the terms which are quadratic and quartic in Fermi operators are regarded as an unperturbed and perturbed parts, respectively, destroys the symmetry of the Hamiltonian with respect to the magnetic field. Furthermore, in the antiferromagnetic case, the ground state of the case of $J_x = J_y$, $J_z = 0$ and that of $J_x = J_y = J_z$, has been found to be very close in the finite system. [See (A1) and (A2) in I.]

Analytical expression of the partition function up to third order and numerical calculation up to second order has been obtained. The agreement of the second-order antiferromagnetic ground-state energy in the case $J_\parallel = -|J_\perp|$ with Bethe[10] and Hulthen's[11] value is quite satisfactory.

2. GENERAL FORMULATION

Equation (1.1), under the cyclic condition $N+l \equiv l$, can be written rigorously in terms of Fermion creation and annihilation operators and has been given by (2.9), (2.16), and (2.17) in I (see also errata[21]). In this paper, however, we are concerned only with properties of sufficiently large systems, and can start with the Hamiltonian

$$H = H_0 + H_1, \qquad (2.1)$$

$$H_1 = H_A + H_B, \qquad (2.2)$$

$$H_0 = -\frac{J_\parallel N}{2} + \sum_{k=1}^{N}(\epsilon_k a_k^+ a_k - m\mathfrak{H}), \qquad (2.3)$$

$$\epsilon_k = -2J_\perp\cos\omega + 2m\mathfrak{H}, \qquad (2.4)$$

$$H_A = 2J_\parallel\sum_{k=1}^{N} a_k^+ a_k, \qquad (2.5)$$

$$H_B = -\frac{2J_\parallel}{N}\sum_{k_1}\sum_{k_2}\sum_{k_3}\sum_{k_4}\delta_K(k_1+k_2-k_3-k_4)\times\exp[i(\omega_1-\omega_4)]a_{k_1}^+a_{k_2}^+a_{k_3}a_{k_4}, \qquad (2.6)$$

where $\omega = 2\pi k/N$, and the modified Kronecker's symbol $\delta_K(k)$ is equal to unity if the argument is zero (mod N) and vanishes otherwise. $J_\perp$ and $J_\parallel$ are perpendicular and parallel components of the exchange integrals. The first term in the right-hand side of (2.3) is a part of the $J_\parallel$ term in the original Hamiltonian (1.1), but is included in our unperturbed term since it gives only a constant energy shift of the total system.

It is convenient to use a more symmetrical expression for H_B than (2.6). One can write

$$H_B = -\frac{J_\parallel}{N}\sum_{k_1}\sum_{k_2}\sum_{k_3}\sum_{k_4}V(k_1,k_2,k_3,k_4)\times a_{k_1}^+a_{k_2}^+a_{k_3}a_{k_4}, \qquad (2.7)$$

where

$$V(k_1,k_2,k_3,k_4) = \delta_K(k_1+k_2-k_3-k_4)\times[\cos(\omega_1-\omega_4)-\cos(\omega_1-\omega_3)], \qquad (2.8)$$

with the properties

$$V(k_1,k_2,k_3,k_4) = V(k_3,k_4,k_1,k_2) = -V(k_2,k_1,k_3,k_4). \qquad (2.9)$$

Usually, H_A which is quadratic in Fermion operators is included in the unperturbed Hamiltonian.

[17] L. Berger, S. A. Friedberg, and J. T. Schriempf, Phys. Rev. 132, 1057 (1963).
[18] R. B. Griffiths, Phys. Rev. 135, A 659 (1964).
[19] T. Haseda, Metal Physics (in Japanese) 9, 23 (1963).
[20] S. Katsura, Phys. Rev. 127, 1508 (1962). This paper is referred to as I.
[21] S. Katsura, Phys. Rev. 129, 2835 (1963).
[22] E. Lieb, T. Schulz, D. Mattis, Ann. Phys. 16, 407 (1961).

Frank[7] and Mannari[8] obtained spontaneous magnetization for the one-dimensional Heisenberg ferromagnet in this way. Such separation of the Hamiltonian, however, destroys the symmetry of the Hamiltonian with respect to the magnetic field. A curve of the magnetization vs magnetic field in their unperturbed system is a parallel shift of that in Case (i) in the preceding paper. That is

$$M_{F-M}(m\mathfrak{H}/kT) = M_{J_\parallel=0}(m\mathfrak{H}/kT \pm |J_\perp|/kT).$$

The upper and lower signs correspond to cases of $J_\parallel > 0$ and $J_\parallel < 0$, respectively. Thus, in this paper, H_A is included in the perturbation to keep the symmetry with respect to the magnetic field.

The partition function Z is given by[23]

$$Z = Z_0 \sum_{n=0}^{\infty} \langle S_n(\beta) \rangle, \tag{2.10}$$

where Z_0 is the partition function of the unperturbed system and $\langle\ \rangle$ means the average over the canonical ensemble

$$\langle A \rangle = \text{tr}\,[e^{-\beta H_0} A]/Z_0, \tag{2.11}$$

and

$$S_n(\beta) = (-1)^n \int_0^\beta ds_1 \int_0^{s_1} ds_2 \cdots \times \int_0^{s_{n-1}} ds_n H_1(s_1)H_1(s_2) \cdots H_1(s_n), \tag{2.12}$$

where

$$H_1(s) = H_A(s) + H_B(s), \tag{2.13}$$

$$H_A(s) = \exp(sH_0)H_A \exp(-sH_0), \tag{2.14}$$

$$H_B(s) = \exp(sH_0)H_B \exp(-sH_0).$$

$H_A(s)$ and $H_B(s)$ are obtained by replacing the creation and annihilation operators in (2.5) and (2.7) by

$$a_k^+(s) = a_k^+ e^{s\epsilon_k} \quad \text{and} \quad a_k(s) = a_k e^{-s\epsilon_k}. \tag{2.15}$$

The partition function of the unperturbed system is

$$Z_0 = \text{tr} \exp(-\beta H_0) = \exp(\tfrac{1}{2}\beta J_\parallel N)\ \text{Tr}\,[\prod_k \exp(-\beta\epsilon_k a_k^+ a_k + \beta m\mathfrak{H})] = \exp(K_\parallel N) \prod_k 2 \exp(2K \cos\omega) \times \cosh(2K \cos\omega - C), \tag{2.16}$$

where $K = \frac{1}{2}\beta J_\perp$, $K_\parallel = \frac{1}{2}\beta J_\parallel$ and $C = \beta m\mathfrak{H}$. (2.17)

Hence

$$\frac{1}{N}\log Z_0 = K_\parallel + \frac{1}{N}\sum_k \log[2 \cosh(2K \cos\omega - C)].$$

[23] T. Matsubara, Progr. Theoret. Phys. (Kyoto) 14, 351 (1955).

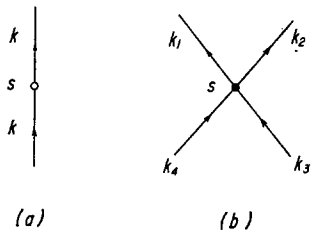


FIG. 1. (a) A white vertex; (b) A black vertex.

In the limit $N \to \infty$, $(1/N)\sum_k$ is to be replaced by $(1/2\pi)\int_0^{2\pi} d\omega$. On account of the symmetry of the integrand with respect to $k = \pi$, we have

$$\lim_{N\to\infty} \frac{1}{N}\log Z_0 = K_\parallel + \frac{1}{\pi}\int_0^\pi \log[2 \cosh(2K \cos\omega - C)]\, d\omega. \tag{2.18}$$

The second term in (2.18) agrees with (3.1) in I, although (2.1)–(2.5) are different from (2.31) in I. The coincidence justifies the replacement of the original Hamiltonian (2.16) and (2.17) in I by (2.1)–(2.5) in this paper.

3. GRAPH REPRESENTATION

$S_n(\beta)$ contains products of the creation and annihilation operators. Now we introduce two notations of the contraction for two kinds of pairs of the creation and annihilation operators,

$$f_k^+ = \langle a_k a_k^+ \rangle = 1/[1 + \exp(-\beta\epsilon_k)], \tag{3.1}$$

and

$$f_k^- = \langle a_k^+ a_k \rangle = \exp(-\beta\epsilon_k)/[1 + \exp(-\beta\epsilon_k)]. \tag{3.2}$$

The average of $S_n(\beta)$ is then expanded[23–26] in terms of products of contractions and additional factors according to Wick's theorem. Each term of the expansion is represented by a graph which is constructed by vertices and lines with arrows. A line in a graph represents a contraction, (3.1) or (3.2). Corresponding to the interaction $H_A(s)$ and $H_B(s)$, there appear two kinds of vertices—a white vertex with an incoming and an outgoing lines [Fig. 1(a)], and a black vertex with two incoming and two outgoing lines [Fig. 1(b)]. The sequence of interactions in a graph is arranged according to the ascending order of the parameters in (2.12) from the bottom to the top of the graph.

We have only to consider the "vacuum–vacuum" graphs without any external lines. The contribution

[24] N. Hugenholtz, Physica 23, 481 (1957).
[25] C. Bloch, Nucl. Phys. 7, 459 (1958); C. Bloch and C. DeDominicis, Nucl. Phys. 10, 181 (1959).
[26] A. W. Glassgold, W. Heckrotte, and K. M. Watson, Phys. Rev. 115, 1374 (1959).

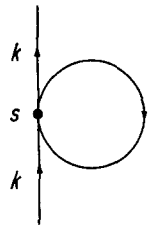


FIG. 2. A self-closed line at a black vertex.

to the average of $S_n(\beta)$ from each graph can be written according to the following rules in our case.

(1) We associate a factor

$$-(J/N)V(k_1, k_2, k_3, k_4) \times \exp[s(\epsilon_{k_1} + \epsilon_{k_2} - \epsilon_{k_3} - \epsilon_{k_4})] \qquad (3.3)$$

to each black vertex, where the momenta k_1 and k_2 are attributed to two outgoing lines, and k_3 and k_4 to two incoming lines.

(2) We associate a factor $2J_\parallel$ to each white vertex.

(3) We associate f_k^+ to a line directed upwards; we associate f_k^- to a line directed downwards and to a "self-closed line," i.e., a loop consisting of only one line and one vertex (Figs. 2 and 3).

(4) We associate $(-1)^{n+h+l}$ to the graph where n is the number of interactions, h the number of lines associated with f_k^-, and l is defined as follows. If we, at each black vertex, separate the lines corresponding to k_1 and k_4 in (3.3) from the other lines corresponding to k_2 and k_3, the original graph is decomposed into a set of loops and the number of the loops is denoted by l.

A graph, every part of which is connected to every other part by lines, is called a connected graph. The contributions to the free energy arise from the connected graphs, i.e.,

$$\log Z = \log Z_0 + \sum_{n=1}^{\infty} \langle S_n(\beta)\rangle_0. \qquad (3.4)$$

Before going into details of perturbation calculation, it is convenient to prove a theorem on an effect of the interaction H_A.

Theorem. When a graph has any self-closed lines as a part of it, another graph is obtained by replacing one of the black vertices associated with the self-closed lines by a white vertex. The sum of the contributions of these two graphs is obtained by replacing f_k^- of the self-closed line in the original graph by $(f_k^- - \frac{1}{2})$. [The graph with two self-closed lines at a black vertex shown in Fig. 3(b) is excluded as an exception.]

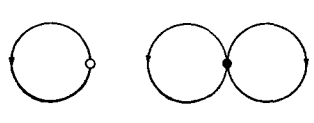


FIG. 3. First-order graphs: (a) first order in H_A; (b) first order in H_B.

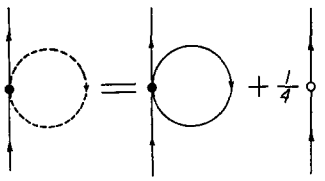


FIG. 4. A broken self-closed line.

Proof: We consider a part of the graph which consists of a black vertex with a self-closed line and two external lines (Fig. 2). Because of the momentum conservation, the outgoing momentum is the same as the incoming one and the exponential factor in (3.3) reduces to unity giving no dependence on the parameter s. For a fixed external momentum, there arise four contributions from the interaction $H_B(s)$. On account of the relation (2.9), these four contributions are the same:

$$\left[-\frac{J_\parallel}{N}\sum_{k'} V(k, k', k', k) f_{k'}^-\right] a_k^+ a_k. \qquad (3.5)$$

When the black vertex in Fig. 2 is replaced by a white vertex in Fig. 1(a) with the other parts of the graph unchanged, another graph is obtained and the contribution of the white vertex part is given by

$$2J_\parallel a_k^+ a_k = \left[\frac{2J_\parallel}{N}\sum_{k'} V(k, k', k', k)\right] a_k^+ a_k, \qquad (3.6)$$

where we used the identity

$$\frac{1}{N}\sum_{k'} V(k, k', k', k) = 1.$$

Now we are able to include a quarter of the contribution of the white vertex in the contribution of the self-closed line at the black vertex, replacing $f_{k'}^-$ in (3.5) by $(f_{k'}^- - \frac{1}{2})$. In the exceptional case in Fig. 3(b), the replacement should be executed by taking $(f_{k'}^- - 1)$ instead of $(f_{k'}^- - \frac{1}{2})$ because of the symmetric nature of the graph.

We introduce a self-closed line made of a broken line as shown in Fig. 4, representing the sum of (3.5) and a quarter of (3.6). Thus the following rule is obtained:

(5) We replace any self-closed lines at black vertices [except Fig. 3(b)] by the broken self-closed line, and associate to it a factor

$$(f_k^- - \tfrac{1}{2}), \qquad (3.7)$$

and neglect all the graphs that contain the interaction H_A except Fig. 3(a).

At a final step, the integration over the parameter s and the summation over the momentum are to be carried out. In the perturbation expansion beyond the first order in $J_\parallel$, the effect of H_A is thus entirely taken into account by the above Rule (5).

In the expansion of $\langle S_n(\beta)\rangle$ into products of contractions, there are several terms that correspond to the same graph. The number of such terms is called the multiplicity of the graph and denoted by N_m, which is enumerated according to the formula

$$N_m = 4^B/2^S, \tag{3.8}$$

where B is the number of black vertices and S the number of pairs of equivalent lines. [When two incoming (or outgoing) lines at a black vertex are connected together to a black vertex, they are called "equivalent lines."]

4. THE FIRST-ORDER CALCULATION

The first-order graphs are shown in Fig. 3(a) and (b), and their contributions are given by

$$\langle S_{1a}(\beta)\rangle = -2J_\parallel\beta \sum_k f_k^-, \tag{4.1}$$

and

$$\langle S_{1b}(\beta)\rangle = \frac{2J_\parallel}{N}\beta \sum_{k_1}\sum_{k_2} V(k_1, k_2, k_2, k_1) f_{k_1}^- f_{k_2}^-, \tag{4.2}$$

respectively. It is convenient to introduce a function

$$g_k = \tanh(2K\cos\omega - C), \tag{4.3}$$

and express f_k^+ and f_k^- in terms of it, i.e.,

$$f_k^+ = \tfrac{1}{2}(1 - g_k), \tag{4.4}$$

$$f_k^- = \tfrac{1}{2}(1 + g_k). \tag{4.5}$$

Then we have

$$\lim_{N\to\infty}\frac{1}{N}\langle S_1(\beta)\rangle_c$$

$$= \lim_{N\to\infty}\frac{1}{N}\{\langle S_{1a}(\beta)\rangle + \langle S_{1b}(\beta)\rangle\} = -K_\parallel$$

$$+ \frac{K_\parallel}{\pi^2}\int_0^\pi\int_0^\pi (1 - \cos\omega_1\cos\omega_2) g_{k_1} g_{k_2}\, d\omega_1\, d\omega_2. \tag{4.6}$$

5. THE SECOND-ORDER CALCULATION

The second order contribution to the partition function comes from the two graphs shown in Fig. 5(a) and (b). The former, so-called "anomalous graph", includes the effect of H_A and has two closed loops of broken lines, and the latter, "ground state graph", includes only the second interaction H_B and leads, as shown in Sec. 7, to a nonvanishing amount

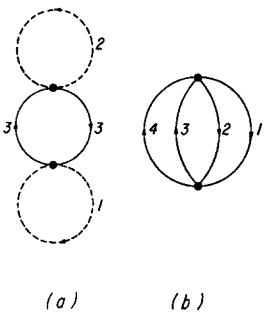


FIG. 5. Second-order graphs: (a) anomalous graph; (b) ground-state graph. Numbers 1, 2, 3, 4 are abbreviations of k_1, k_2, k_3, k_4. Upper and lower black points are labeled as s_1, and s_2, respectively.

of the ground-state energy in the second-order perturbation. The contribution of the anomalous graph has no s-dependent factor and we have

$$\langle S_{2a}(\beta)\rangle = 16\frac{J_\parallel^2}{N^2}\frac{\beta^2}{2}\sum_{k_1}\sum_{k_2}\sum_{k_3} V(k_3, k_1, k_1, k_3)$$

$$\times V(k_3, k_2, k_2, k_3)(f_{k_1}^- - \tfrac{1}{2})(f_{k_3}^- - \tfrac{1}{2}) f_{k_2}^+ f_{k_3}^-, \tag{5.1}$$

$$\lim_{N\to\infty}\frac{1}{N}\langle S_{2a}(\beta)\rangle = 2K_\parallel^2$$

$$\times \frac{1}{\pi^3}\int_0^\pi\int_0^\pi\int_0^\pi (1 - \cos\omega_1\cos\omega_3)(1 - \cos\omega_2\cos\omega_3)$$

$$\times g_{k_1} g_{k_2}(1 - g_{k_3}^2)\, d\omega_1\, d\omega_2\, d\omega_3. \tag{5.2}$$

The contribution of the ground-state graph is given by

$$\langle S_{2b}(\beta)\rangle = 4\frac{J_\parallel^2}{N^2}\sum_{k_1}\sum_{k_2}\sum_{k_3}\sum_{k_4}[V(k_1, k_2, k_3, k_4)]^2$$

$$\times f_{k_1}^- f_{k_2}^- f_{k_3}^+ f_{k_4}^+ \int_0^\beta ds_1\int_0^{s_1} ds_2$$

$$\times \exp[(s_1 - s_2)(\epsilon_{k_1} + \epsilon_{k_2} - \epsilon_{k_3} - \epsilon_{k_4})]. \tag{5.3}$$

Changing the variables $k_1 \leftrightarrows k_4$, $k_2 \leftrightarrows k_3$ in (5.3), and making use of the identity

$$f_{k_1}^+ f_{k_2}^+ f_{k_3}^- f_{k_4}^- = \exp[\beta(\epsilon_{k_1} + \epsilon_{k_2} - \epsilon_{k_3} - \epsilon_{k_4})]$$

$$\times f_{k_1}^- f_{k_2}^- f_{k_3}^+ f_{k_4}^+, \tag{5.4}$$

we have another expression for $\langle S_{2b}(\beta)\rangle$. Then the average of the two expressions and further use of the relation

$$\left\{\int_0^\beta ds_1\int_0^{s_1} ds_2 \exp[(s_1 - s_2)\epsilon] + e^{\beta\epsilon}\int_0^\beta ds_1\int_0^{s_1} ds_2\right.$$

$$\left.\times \exp[-(s_1 - s_2)\epsilon]\right\} = (\beta/\epsilon)(e^{\beta\epsilon} - 1)$$

gives

$$\langle S_{2b}(\beta)\rangle = 2\left(\frac{J_\parallel}{N}\right)^2\sum_{k_1}\sum_{k_2}\sum_{k_3}\sum_{k_4}[V(k_1, k_2, k_3, k_4)]^2$$

$$\times f_{k_1}^- f_{k_2}^- f_{k_3}^+ f_{k_4}^+ \frac{\beta}{(\epsilon_{k_1} + \epsilon_{k_2} - \epsilon_{k_3} - \epsilon_{k_4})}$$

$$\times \{\exp[\beta(\epsilon_{k_1} + \epsilon_{k_2} - \epsilon_{k_3} - \epsilon_{k_4})] - 1\}. \tag{5.5}$$

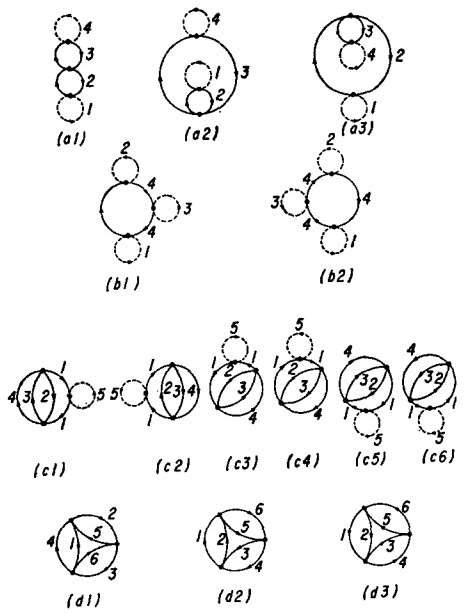


FIG. 6. Third-order graphs. Numbers 1, 2, $\cdots$ 6 are abbreviations of k_1, k_2, $\cdots$ k_6, respectively. Black points are labeled as s_1, s_2, s_3 from up to down. An upward or downward line is defined according to relative positions of the starting and end points.

Now

$$\frac{e^{\beta\epsilon}-1}{\epsilon}=\frac{1}{\epsilon}\frac{e^{\beta\epsilon}-1}{e^{\beta\epsilon}+1}e^{\beta\epsilon}+\frac{1}{\epsilon}\frac{e^{\beta\epsilon}-1}{e^{\beta\epsilon}+1}. \tag{5.6}$$

Again, the symmetric properties of the summand in (5.5) with regard to the transformation $k_1 \leftrightarrows k_4$, $k_2 \leftrightarrows k_3$ and the identity (5.4) applied to the first term of (5.6) permit the replacement of $e^{\beta\epsilon}-1$ by $2\tanh(\frac{1}{2}\beta\epsilon)$, where $\epsilon=\epsilon_{k_1}+\epsilon_{k_2}-\epsilon_{k_3}-\epsilon_{k_4}$. Thus $\langle S_{2b}(\beta)\rangle$ can be expressed in a form where zeros of the energy denominator do not yield singularities, i.e.,

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\langle S_{2b}(\beta)\rangle &= \frac{K_\parallel^2}{2}\frac{1}{(2\pi)^3}\int_0^{2\pi}\int_0^{2\pi}\int_0^{2\pi}\int_0^{2\pi}\delta(\omega_1+\omega_2-\omega_3-\omega_4)\\ &\times[1-\cos(\omega_1-\omega_2)][1-\cos(\omega_3-\omega_4)]\\ &\times(1+g_{k_1})(1+g_{k_2})(1-g_{k_3})(1-g_{k_4})\\ &\times\frac{\tanh 2K(\cos\omega_1+\cos\omega_2-\cos\omega_3-\cos\omega_4)}{2K(\cos\omega_1+\cos\omega_2-\cos\omega_3-\cos\omega_4)}\\ &\qquad d\omega_1\,d\omega_2\,d\omega_3\,d\omega_4, \end{aligned}\tag{5.7}$$

where delta function in the right-hand side means that only such combinations of the variables that

$$\omega_1+\omega_2-\omega_3-\omega_4=0 \pmod{2\pi}$$

contribute to the integral. This convention is also used in Sec. 6.

6. THIRD-ORDER CALCULATION

In the third-order calculation, we must consider four groups of graphs as shown in Fig. 6. We have

$$\begin{aligned}\langle S_3(\beta)\rangle_c &= \langle S_{3a}(\beta)\rangle+\langle S_{3b}(\beta)\rangle\\ &\quad+\langle S_{3c}(\beta)\rangle+\langle S_{3d}(\beta)\rangle,\end{aligned}\tag{6.1}$$

and each contribution is given as follows:

(a group) Three a graphs belong to this group. Their contributions do not include the s-dependent factor and they give one and the same expression, i.e.,

$$\begin{aligned}&\langle S_{3a1}(\beta)\rangle=\langle S_{3a2}(\beta)\rangle=\langle S_{3a3}(\beta)\rangle\\ &=64\left(\frac{J_\parallel}{N}\right)^3(\tfrac{1}{6}\beta^3)\sum_{k_1}\sum_{k_2}\sum_{k_3}\sum_{k_4}V(k_1,k_2,k_2,k_1)\\ &\times V(k_2,k_3,k_3,k_2)V(k_3,k_4,k_4,k_3)\\ &\times(f_{k_1}^-\!-\tfrac{1}{2})(f_{k_4}^-\!-\tfrac{1}{2})f_{k_2}^+f_{k_2}^-f_{k_3}^+f_{k_3}^-.\end{aligned}\tag{6.2}$$

Then we have

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\langle S_{3a}(\beta)\rangle &= \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{3}\langle S_{3ai}(\beta)\rangle\\ &=4K_\parallel^3\frac{1}{\pi^4}\int_0^\pi\int_0^\pi\int_0^\pi\int_0^\pi(1-\cos\omega_1\cos\omega_2)\\ &\times(1-\cos\omega_2\cos\omega_3)(1-\cos\omega_3\cos\omega_4)\\ &\times g_{k_1}g_{k_4}(1-g_{k_2}^2)(1-g_{k_3}^2)\,d\omega_1\,d\omega_2\,d\omega_3\,d\omega_4.\end{aligned}\tag{6.3}$$

(b group) There are two graphs without s-dependence. One of the graphs, say $b1$ graph, gives

$$\begin{aligned}&\langle S_{3b1}(\beta)\rangle\\ &=-64\left(\frac{J_\parallel}{N}\right)^3(\tfrac{1}{6}\beta^3)\sum_{k_1}\sum_{k_2}\sum_{k_3}\sum_{k_4}V(k_1,k_4,k_4,k_1)\\ &\times V(k_2,k_4,k_4,k_2)V(k_3,k_4,k_4,k_3)\\ &\times(f_{k_1}^-\!-\tfrac{1}{2})(f_{k_2}^-\!-\tfrac{1}{2})(f_{k_4}^-\!-\tfrac{1}{2})f_{k_4}^+f_{k_4}^-f_{k_3}^-.\end{aligned}\tag{6.4}$$

The other graph, i.e., $b2$ graph, can be obtained by reversing the directions of the arrows in $b1$ graph. In general, when the directions of all the arrows in a graph are reversed, another graph is obtained and called a "reversed" graph. The contribution of the "reversed" graph is obtained from the original one by altering the sign of the energy in the s-dependent factor and replacing f^+ or f^- of the full line by $-f^-$ or $-f^+$, respectively. Now $b2$ graph is the "reversed" graph of $b1$ graph, and $b2$ contribution is the same as the right-hand side of (6.4), except one of $f_{k_4}^-$ is

replaced by $-f_{k_4}^+$. Then we have

$$\frac{1}{N}\lim_{N\to\infty}\langle S_{3b}(\beta)\rangle = \lim_{N\to\infty}\frac{1}{N}[\langle S_{3b1}(\beta)\rangle + \langle S_{2b2}(\beta)\rangle]$$
$$= -\tfrac{8}{3}K_\parallel^3\int_0^\pi\int_0^\pi\int_0^\pi\int_0^\pi (1-\cos\omega_1\cos\omega_4)$$
$$\times(1-\cos\omega_2\cos\omega_4)(1-\cos\omega_3\cos\omega_4)$$
$$\times g_{k_1}g_{k_2}g_{k_3}g_{k_4}(1-g_{k_4}^2)\,d\omega_1\,d\omega_2\,d\omega_3\,d\omega_4. \tag{6.5}$$

(c group) There are six graphs, three of which $c2$, $c4$, and $c6$ are the "reversed" graph of $c1$, $c3$, and $c5$, respectively. Their contributions can be written in the form

$$\langle S_{3ci}(\beta)\rangle$$
$$= 32(J_\parallel/N)^3\sum_{k_1}\sum_{k_2}\sum_{k_3}\sum_{k_4}\sum_{k_5}V(k_1,k_2,k_3,k_4)$$
$$\times V(k_1,k_5,k_5,k_1)V(k_4,k_3,k_2,k_1)$$
$$\times(f_{k_4}-\tfrac{1}{2})F_{ci}(k_1,k_2,k_3,k_4)I_{ci}(\epsilon_c), \tag{6.6}$$

and

$$\begin{aligned}
F_{c1} &= f_{k_1}^- f_{k_1}^- f_{k_2}^- f_{k_3}^+ f_{k_4}^+,\\
F_{c2} &= -f_{k_1}^+ f_{k_1}^+ f_{k_2}^+ f_{k_3}^- f_{k_4}^-,\\
F_{c3} &= -f_{k_1}^+ f_{k_1}^- f_{k_2}^- f_{k_3}^+ f_{k_4}^+,\\
F_{c4} &= f_{k_1}^+ f_{k_1}^- f_{k_2}^+ f_{k_3}^- f_{k_4}^-,\\
F_{c5} &= -f_{k_1}^+ f_{k_1}^- f_{k_2}^- f_{k_3}^+ f_{k_4}^+,\\
F_{c6} &= f_{k_1}^+ f_{k_1}^- f_{k_2}^+ f_{k_3}^- f_{k_4}^-,
\end{aligned} \tag{6.7}$$

and

$$\begin{aligned}
I_{c1}(x) &= \int_0^\beta ds_1\int_0^{s_1} ds_2\int_0^{s_2} ds_3\,\exp\{(s_1-s_3)x\}\\
&= \int_0^\beta s(\beta-s)e^{sx}\,ds,\\
I_{c2}(x) &= I_{c1}(-x) = e^{-\beta x}I_{c1}(x),\\
I_{c3}(x) &= \int_0^\beta ds_1\int_0^{s_1} ds_1\int_0^{s_2} ds_3\,\exp\{(s_2-s_3)x\}\\
&= \tfrac{1}{2}\int_0^\beta(\beta-s)^2e^{sx}\,ds,\\
I_{c4}(x) &= I_{c3}(-x)\\
&= \tfrac{1}{2}e^{-\beta x}\int_0^\beta s^2e^{sx}\,ds,\\
I_{c5}(x) &= \int_0^\beta ds_1\int_0^{s_1} ds_2\int_0^{s_2} ds_3\,\exp\{(s_1-s_2)x\}\\
&= I_{c3}(x),\\
I_{c6}(x) &= I_{c5}(-x) = I_{c4}(x),
\end{aligned} \tag{6.8}$$

and

$$\epsilon_c = \epsilon_{k_1}+\epsilon_{k_2}-\epsilon_{k_3}-\epsilon_{k_4}. \tag{6.9}$$

By use of the relation (5.4) and

$$\begin{aligned}
I_{c1}(x)+2I_{c3}(x) &= \beta^3I_1(x)-\tfrac{1}{2}\beta^3I_2(x),\\
I_{c1}(x)+2I_{c4}(x) &= \tfrac{1}{2}\beta^3I_2(x),
\end{aligned} \tag{6.10}$$

where

$$\begin{aligned}
I_1(x) &= \frac{1}{\beta}\int_0^\beta e^{sx}\,ds = \frac{e^{\beta x}-1}{\beta x},\\
I_2(x) &= \frac{1}{\frac{1}{2}\beta^2}\int_0^\beta se^{sx}\,ds = \frac{e^{\beta x}-\beta x-1}{\frac{1}{2}\beta^2x^2},
\end{aligned} \tag{6.11}$$

we have

$$\lim_{N\to\infty}\frac{1}{N}\langle S_{3c}(\beta)\rangle = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{6}\langle S_{3ci}(\beta)\rangle$$
$$= -4\left(\frac{K_\parallel}{N}\right)^3\frac{1}{(2\pi)^4}\int_0^{2\pi}\int_0^{2\pi}\int_0^{2\pi}\int_0^{2\pi}\int_0^{2\pi}\delta(\omega_1+\omega_2$$
$$-\omega_3-\omega_4)\{1-\cos\omega_1\cos\omega_5\}\{1-\cos(\omega_1-\omega_2)\}$$
$$\times\{1-\cos(\omega_3-\omega_4)\}(1+g_{k_1})(1+g_{k_2})$$
$$\times(1-g_{k_3})(1-g_{k_4})g_{k_5}[I_2(\epsilon_c)-(1+g_{k_1})$$
$$\times I_1(\epsilon_c)]\,d\omega_1\,d\omega_2\,d\omega_3\,d\omega_4\,d\omega_5. \tag{6.12}$$

(d group) There are three graphs $d1$, $d2$, and $d3$, the latter two being the "reversed" graph of each other.

$$\langle S_{3d1}(\beta)\rangle$$
$$= 64\left(\frac{J_\parallel}{N}\right)^3\sum_{k_1}\sum_{k_2}\sum_{k_3}\sum_{k_4}\sum_{k_5}\sum_{k_6}V(k_1,k_2,k_5,k_4)$$
$$\times V(k_3,k_5,k_2,k_6)V(k_4,k_6,k_3,k_1)$$
$$\times f_{k_1}^-f_{k_3}^-f_{k_5}^-f_{k_4}^+f_{k_2}^+f_{k_6}^+(\tfrac{1}{6}\beta^3)I_d(\epsilon_{d11},\epsilon_{d12}), \tag{6.13}$$

$$\langle S_{3d2}(\beta)\rangle$$
$$= 8\left(\frac{J_\parallel}{N}\right)^3\sum_{k_1}\sum_{k_2}\sum_{k_3}\sum_{k_4}\sum_{k_5}\sum_{k_6}V(k_5,k_6,k_1,k_2)$$
$$\times V(k_3,k_4,k_6,k_5)V(k_1,k_2,k_3,k_4)$$
$$\times f_{k_1}^+f_{k_3}^+f_{k_5}^-f_{k_4}^-f_{k_2}^-f_{k_6}^-(\tfrac{1}{6}\beta^3)I_d(\epsilon_{d21},\epsilon_{d22}), \tag{6.14}$$

where

$$I_d(x,y) = \frac{1}{\frac{1}{6}\beta^3}\int_0^\beta ds_1\int_0^{s_1} ds_2\int_0^{s_3} ds_3$$
$$\times\exp\{(s_1-s_3)x+(s_2-s_3)y\}$$
$$= \frac{1}{\frac{1}{6}\beta^3}\left[\frac{\beta}{x(x+y)}-\frac{1}{y}\left\{\frac{e^{\beta x}-1}{x^2}-\frac{e^{\beta(x+y)}-1}{(x+y)^2}\right\}\right], \tag{6.15}$$

and

$$\begin{aligned}\epsilon_{d11} &= \epsilon_{k_1} + \epsilon_{k_2} - \epsilon_{k_4} - \epsilon_{k_5},\\ \epsilon_{d12} &= \epsilon_{k_3} + \epsilon_{k_5} - \epsilon_{k_2} - \epsilon_{k_6},\\ \epsilon_{d21} &= \epsilon_{k_5} + \epsilon_{k_6} - \epsilon_{k_1} - \epsilon_{k_3},\\ \epsilon_{d22} &= \epsilon_{k_3} + \epsilon_{k_4} - \epsilon_{k_6} - \epsilon_{k_5}.\end{aligned} \tag{6.16}$$

The contribution of $d3$ graph is easily obtained from (6.14) according to the reversion rule mentioned above. We have

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\langle S_{3d}(\beta)\rangle &= \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{3}\langle S_{3d_i}(\beta)\rangle\\ &= \tfrac{4}{3}K_{\parallel}^3\frac{1}{(2\pi)^4}\int_0^{2\pi}\int_0^{2\pi}\int_0^{2\pi}\int_0^{2\pi}\int_0^{2\pi}\int_0^{2\pi}\delta(\omega_1+\omega_2\\ &\quad-\omega_4-\omega_5)\delta(\omega_3+\omega_5-\omega_2-\omega_6)\\ &\quad\times\{\cos(\omega_1-\omega_4)-\cos(\omega_1-\omega_5)\}\\ &\quad\times\{\cos(\omega_3-\omega_6)-\cos(\omega_3-\omega_2)\}\\ &\quad\times\{\cos(\omega_4-\omega_1)-\cos(\omega_4-\omega_3)\}I_d(\epsilon_{d11},\epsilon_{d12})\\ &\quad\times(1+g_{k_1})(1+g_{k_2})(1+g_{k_3})(1-g_{k_4})(1-g_{k_5})\\ &\quad\times(1-g_{k_6})\,d\omega_1\,d\omega_2\,d\omega_3\,d\omega_4\,d\omega_5\,d\omega_6\\ &\quad+\tfrac{1}{6}K_{\parallel}^3\frac{1}{(2\pi)^4}\int_0^{2\pi}\int_0^{2\pi}\int_0^{2\pi}\int_0^{2\pi}\int_0^{2\pi}\int_0^{2\pi}\delta(\omega_1+\omega_2\\ &\quad-\omega_3-\omega_4)\delta(\omega_3+\omega_4-\omega_5-\omega_6)\\ &\quad\times\{\cos(\omega_5-\omega_2)-\cos(\omega_5-\omega_1)\}\\ &\quad\times\{\cos(\omega_3-\omega_5)-\cos(\omega_3-\omega_6)\}\\ &\quad\times\{\cos(\omega_1-\omega_4)-\cos(\omega_1-\omega_3)\}\\ &\quad\times\{(1-g_{k_1})(1-g_{k_2})(1+g_{k_3})(1+g_{k_4})\\ &\quad\times(1+g_{k_5})(1+g_{k_6})I_d(\epsilon_{d21},\epsilon_{d22})\\ &\quad+(1+g_{k_1})(1+g_{k_2})(1-g_{k_3})(1-g_{k_4})\\ &\quad\times(1-g_{k_5})(1-g_{k_6})I_d(-\epsilon_{d21},-\epsilon_{d22})\}\\ &\quad\times d\omega_1\,d\omega_2\,d\omega_3\,d\omega_4\,d\omega_5\,d\omega_6.\end{aligned} \tag{6.17}$$

7. THERMODYNAMIC QUANTITIES

Here we list expressions of energy and susceptibility of zero field up to second order. The abbreviations

$$A_{mn}=\frac{1}{\pi}\int_0^{\pi}\cos^m\omega\tanh^n(2K\cos\omega)\,d\omega \tag{7.1}$$

are used, and quantities of the zeroth order, of the first order, of the second order of the anomalous-state graph, of the second order of the ground-state graph, are denoted by the subscripts, 0, 1, $2a$, $2b$, respectively.

Energy:

$$\frac{E_0}{N|J_\perp|}=-\frac{J_\perp}{|J_\perp|}A_{11}-\frac{1}{2}\frac{J_\parallel}{|J_\perp|}, \tag{7.2}$$

$$\begin{aligned}\frac{E_1}{N|J_\perp|}&=\frac{J_\perp}{|J_\perp|}K_\parallel A_{11}(1-2A_{22})\\ &\quad+\frac{1}{2}\frac{J_\parallel}{|J_\perp|}A_{11}^2+\frac{1}{2}\frac{J_\parallel}{|J_\perp|},\end{aligned} \tag{7.3}$$

$$\begin{aligned}\frac{E_{2a}}{N|J_\perp|}&=-\frac{J_\parallel}{|J_\perp|}K_\parallel(A_{11}^2-2A_{22}A_{11}^2)\\ &\quad-\frac{J_\perp}{|J_\perp|}K_\parallel^2(A_{11}-4A_{11}A_{22}+4A_{11}A_{22}^2\\ &\quad-4A_{11}^2A_{31}+4A_{11}^2A_{33}),\end{aligned} \tag{7.4}$$

$$\begin{aligned}\frac{E_{2b}}{N|J_\perp|}&=-\frac{K_\parallel}{2}\frac{2}{\pi^3}\int_0^{\pi}dx\int_0^{\pi}dz\int_0^{\pi/2}du\,4\sin^2x\sin^2z\\ &\quad\times[1+\tanh(2K\cos(u+x))]\\ &\quad\times[1+\tanh(2K\cos(u-x))]\\ &\quad\times[1+\tanh(2K\cos(u+z))]\\ &\quad\times[1+\tanh(2K\cos(u-z))]\\ &\quad\times\frac{1}{2}\Big\{\frac{J_\parallel}{|J_\perp|}\frac{\tanh 2KW}{2KW}+\frac{J_\perp}{|J_\perp|}\frac{1}{\cosh^2 2KW}\\ &\quad+\frac{J_\perp}{|J_\perp|}\tanh 2KW-\frac{J_\perp}{|J_\perp|}\frac{\tanh 2KW}{2KW}\\ &\quad\times[2K\cos(u+x)\tanh(2K\cos(u+x))\\ &\quad+2K\cos(u-x)\tanh(2K\cos(u-x))\\ &\quad+2K\cos(u+z)\tanh(2K\cos(u+z))\\ &\quad+2K\cos(u-z)\tanh(2K\cos(u-z))]\Big\},\end{aligned} \tag{7.5}$$

where

$$\begin{aligned}&x=\tfrac{1}{2}(\omega_1-\omega_2),\ z=\pi-\tfrac{1}{2}(\omega_3-\omega_4),\ u=\tfrac{1}{2}(\omega_1+\omega_2),\\ &W=\cos(u+x)+\cos(u-x)\\ &\qquad+\cos(u+z)+\cos(u-z).\end{aligned} \tag{7.6}$$

Susceptibility:

$$\frac{|J_\perp|\chi_0}{Nm^2}=2|K|(1-A_{02}), \tag{7.7}$$

$$\begin{aligned}\frac{|J_\perp|\chi_1}{Nm^2}&=4|K|K_\parallel(1-2A_{02}+A_{02}^2\\ &\quad+2A_{11}^2-2A_{11}A_{13}),\end{aligned} \tag{7.8}$$

$$\begin{aligned}\frac{|J_\perp|\chi_{2a}}{Nm^2}&=8|K|K_\parallel^2[(1-A_{02})^3\\ &\quad+4(A_{11}-A_{13})A_{11}(1-A_{02})\\ &\quad-2A_{11}(A_{11}-A_{13})(\tfrac{1}{2}-A_{22})\\ &\quad+A_{11}^2(-A_{20}+4A_{22}-3A_{24})],\end{aligned} \tag{7.9}$$

$$\frac{|J_\perp|\chi_{2b}}{Nm^2} = 2|K|\frac{K_\parallel^2}{2}\frac{2}{\pi^3}\int_0^\pi dx\int_0^\pi dz\int_0^{\pi/2} du\, 4\sin^2 x\sin^2 z$$
$$\times [1+\tanh(2K\cos(u+x))][1+\tanh(2K\cos(u-x))]$$
$$\times [1+\tanh(2K\cos(u+z))][1+\tanh(2K\cos(u-z))]$$
$$\times \frac{\tanh(2K[\cos(u+x)+\cos(u-x)+\cos(u+z)+\cos(u-z)])}{2K[\cos(u+x)+\cos(u-x)+\cos(u+z)+\cos(u-z)]}$$
$$\times \{[\tanh(2K\cos(u+x))+\tanh(2K\cos(u-z))-\tanh(2K\cos(u+z))-\tanh(2K\cos(u-z))]^2$$
$$-4+\tanh^2(2K\cos(u+x))+\tanh^2(2K\cos(u-x))+\tanh^2(2K\cos(u+z))+\tanh^2(2K\cos(u-z))\}. \quad (7.10)$$

We can obtain the ground-state energy by taking the limit $T \to 0$ in (7.6). The zeroth-order value is $E_0/N|J_\perp| = -2/\pi - \frac{1}{2}(J_\parallel/|J_\perp|)$. The first-order contribution to the ground-state energy is

$$\frac{E_1}{N|J_\perp|} = \left(\frac{1}{2}+\frac{2}{\pi^2}\right)\frac{J_\parallel}{|J_\perp|} = (0.5+0.20264236)\frac{J_\parallel}{|J_\perp|},$$

since $\lim_{K\to\infty} A_{11} = 2/\pi$, and $1-2A_{22}$ tends to zero more rapidly than $O(1/K)$. The same value up to first order in $J_\parallel$, $-2/\pi - 2/\pi^2$, has been obtained by Meyer,[27] Rodriguez,[28] and Bulaevskiĭ.[29] In the same way $E_{2a}/N|J_\perp| = 0$, and the second-order ground-state energy is due to only the ground-state graphs. As $T \to 0$, $\tanh x$ can be replaced by $\mathrm{sign}(x)$, sign (x) $[=1, x>0; =-1, x<0]$. The third and the fourth terms in $\{\ \}$ in (7.5) cancel in the region where $\mathrm{sign}(x) = 1$, and the second term can be neglected compared to the first, and

$$\frac{E_{2b}}{N|J_\perp|} = -\left(\frac{J_\parallel}{J_\perp}\right)^2\frac{16}{\pi^3}\int_0^{\frac{1}{2}\pi} du\int_0^{\frac{1}{2}\pi-u} dz$$
$$\times\int_0^\pi dx\,\frac{\sin^2 x\sin^2 z}{\cos u(\cos x+\cos z)}$$
$$= -\frac{16}{\pi^3}\left(\frac{1}{6}-\frac{\pi^2}{144}\right)\left(\frac{J_\parallel}{J_\perp}\right)^2$$
$$= -0.05063633\left(\frac{J_\parallel}{J_\perp}\right)^2. \quad (7.11)$$

Hence the ground-state energy up to second order is given by

$$\frac{E_G}{N|J_\perp|} = -\frac{2}{\pi}+\frac{2}{\pi^2}\frac{J_\parallel}{|J_\perp|}$$
$$-\frac{16}{\pi^3}\left(\frac{1}{6}-\frac{\pi^2}{144}\right)\left(\frac{J_\parallel}{J_\perp}\right)^2+O\left[\left(\frac{J_\parallel}{J_\perp}\right)^3\right], \quad (7.12)$$

and, in particular for the isotropic interaction $(J_\perp = J_\parallel = J)$,

$$E_G/N|J| = -0.88989845 \quad \text{(antiferromagnetic case)}, \quad (7.13)$$

and

$$E_G/N|J| = -0.48461373 \quad \text{(ferromagnetic case)}, \quad (7.14)$$

respectively. The comparison with the exact values $(-0.8863$ and $-0.5000)$ is satisfactory, though $J_\parallel/J_\perp = -1$ might be the branch point of (7.12).[22]

8. CONCLUSIONS AND DISCUSSIONS

In this paper, the partition function of one-dimensional Heisenberg model has been obtained. $J_\parallel$ term has been regarded as a perturbation to keep the symmetry with respect to the magnetic field. It includes quadratic term as well as a quartic term of fermion operators. Linked-cluster expansion where the quadratic term is treated as a part of perturbation has been developed, and perturbation up to third order has been carried out. Equations (6.2), (6.4), (6.6), (6.13), and (6.14) can be used as general formulas of the third-order perburbation at finite temperatures for calculations of similar problems with a slight modification (replacing back $f_k^- - \frac{1}{2}$ by f_k^-).

Our results are expressed in (2.18), (4.6), (5.2), (5.7), (6.3), (6.5), (6.12), and (6.17), which are symmetric with respect to the magnetic field. High-temperature (and low-field) Taylor expansion can be obtained directly from these equations. The result is

$$\lim_{N\to\infty}\frac{1}{N}\log\frac{Z}{2^N}$$
$$= (K_\parallel + K^2 + \tfrac{1}{2}C^2 - \tfrac{1}{2}K^4 - K^2C^2 + \tfrac{1}{12}C^4$$
$$+ \tfrac{4}{9}K^6 + 2K^4C^2 + \tfrac{2}{3}K^2C^4 + \tfrac{1}{45}C^6 + \cdots) \quad (8.1)$$
$$+ [-K_\parallel + K_\parallel(-K^2 + C^2 + 2K^4$$

[27] K. Meyer, Z. Naturforsch. 11A, 865 (1956).
[28] S. Rodriguez, Phys. Rev. 116, 1474 (1959).
[29] L. N. Bulaevskiĭ, Zh. Eksperim. i Teor. Fiz. 43, 968 (1962) [English transl.: Soviet Phys—JETP 16, 685 (1963)].

$$-2K^2C^2-\tfrac{2}{3}C^4+\cdots)] \quad (8.2)$$

$$+K_\parallel^2(K^2+2C^2-5K^4-7K^2C^2 \quad (8.3)$$
$$-\tfrac{10}{3}C^4+\cdots)$$

$$+K_\parallel^2(\tfrac{1}{2}-\tfrac{5}{3}K^2-C^2+\tfrac{43}{10}K^4 \quad (8.4)$$
$$+9K^2C^2+\tfrac{7}{6}C^4+\cdots)$$

$$+K_\parallel^3(-K^2+4C^2+\cdots) \quad (8.5)$$

$$+K_\parallel^3[O(K^4)+O(K^2C^2)] \quad (8.6)$$

$$+K_\parallel^3(\tfrac{10}{3}K^2-4C^2+\cdots) \quad (8.7)$$

$$+K_\parallel^3(-\tfrac{4}{3}K^2+\tfrac{2}{3}C^2+\cdots) \quad (8.8)$$

$$+O(K_\parallel^4).$$

Expressions (8.1), (8.2), (8.3), (8.4), (8.5), (8.6), (8.7), and (8.8) are contributions [$\lim_{N\to\infty}(1/N)$] (per lattice site in the limit of an infinite ring) of $\log Z_0$, $S_1(\beta)$, $S_{2a}(\beta)$, $S_{2b}(\beta)$, $S_{3a}(\beta)$, $S_{3b}(\beta)$, $S_{3c}(\beta)$, and $S_{3d}(\beta)$, respectively. The first term in (8.2) cancels the first term in (8.1), since a constant term $-\frac{1}{2}J_\parallel N$ has been included in the unperturbed system. The sum of (8.1)–(8.8) has been compared with direct high-temperature series expansion in powers of J_z/kT, J_y/kT, J_z/kT, $m\mathfrak{H}/kT$ [Appendix I, Eq. (I5)] up to the 4th power of these variables, and also with high-temperature expansion of the isotropic case by Rushbrooke and Wood[3] up to

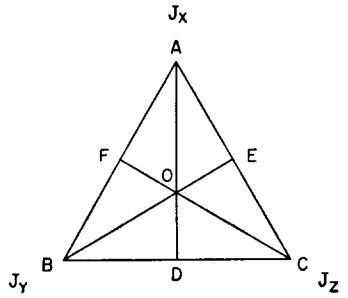


FIG. 7. Ternary diagram of the anisotropic antiferromagnetic Heisenberg model.

TABLE I. Ground-state energy of the linear antiferromagnet according to several calculation.[a]

Neél[b]	1932	0.5000	1/2
Kubo (variation)[c]	1953	0.698	$(1/2)(1+0.198\times 2)$
Syozi[d]	1951	0.7160	$-1/2+(2\sqrt{2}/\pi)\,\mathbf{E}((1/2)^{\frac{1}{2}})$
Oguchi (zeroth approx.)[e,f]	1963		
Anderson (singlet pair)[g]	1951	0.7500	
Hulthén (1st approx.)[h]	1938		
Kasteleijn[i]	1952	0.8156	$\dfrac{7+7(7^{\frac{1}{2}})}{54}$
Marshall[j]	1955		
Taketa-Nakamura[k]	1956		
Meyer[l]	1956		
Rodriguez[m]	1959	0.8393	$2/\pi+2/\pi^3$
Bulaevski[n]	1962		
Oguchi[o] (Correction to Fisher[p])	1963	0.846	
Hulthén (2nd approx.)[q]	1938	0.847	
Anderson (spin wave)[r]	1952	0.8634	$(1/2)+(1-2/\pi)$
Ruijgrok-Rodriguez[s]	1960	0.8646	$-\frac{1}{2}+(2/\pi^2)\,\mathbf{K}^2(k)$,
Oguchi (1st approx.)[e,f]	1963		where $\mathbf{K}(k)-\mathbf{E}(k)=k\pi/2$
Davis[t]	1960	0.8682	
Kubo (spin wave)[u]	1952	0.9294	$(1/2)+(1-2/\pi)$
Oguchi (spin wave)[v]	1960		$+\frac{1}{2}(1-2/\pi)^2$
Katsura and Inawashiro	1963	0.8899	$(2/\pi)+(2/\pi^2)$ $+(16/\pi^3)(1/6-\pi^2/144)$
Bethe[w]	1933	0.8863	$2\log 2-1/2$
Hulthén[x]	1938		

[a] $\mathbf{K}(k)$ and $\mathbf{E}(k)$ are the complete elliptic integrals of the first and the second kind, respectively.
[b] L. Neél, Ann. Phys. (Paris) 17, 64 (1932).
[c] R. Kubo, Rev. Mod. Phys. 25, 344 (1953).
[d] I. Syozi, Busseiron Kenkyu No. 39, 55 (1951); Review Kobe Univ. Merchantile Marine, Sci. Tech. Sec. No. 1, 1 (1954).
[e] T. Oguchi, Phys. Rev. Letters 11, 266 (1963).
[f] Oguchi's zeroth approximation[e] to the ground-state energy is equivalent to Syozi's work.[d] Syozi's value and expression, however, is to be corrected as written here. Oguchi's first approximation is equivalent to the result of Ruijgrok and Rodriguez[s] (the authors are indebted to the referee in the latter point). Oguchi's value is to be corrected as written here.
[g] P. W. Anderson, Phys. Rev. 83, 1260 (1951).
[h] See Ref. 11.
[i] See Ref. 14.
[j] W. Marshall, Proc. Roy. Soc. (London) A232, 48 (1955).
[k] H. Taketa and T. Nakamura, J. Phys. Soc. Japan 11, 919 (1956).
[l] See Ref. 27.
[m] See Ref. 28.
[n] See Ref. 29.
[o] See Ref. 4.
[p] J. C. Fisher, J. Phys. Chem. Solids 10, 44 (1959).
[q] See Ref. 11.
[r] P. W. Anderson, Phys. Rev. 86, 694 (1952).
[s] T. W. Ruijgrok and S. Rodriguez, Phys. Rev. 119, 596 (1960).
[t] H. L. Davis, Phys. Rev. 120, 789 (1960).
[u] R. Kubo, Phys. Rev. 87, 568 (1952).
[v] T. Oguchi, Phys. Rev. 117, 117 (1960).
[w] See Ref. 10.
[x] See Ref. 11.

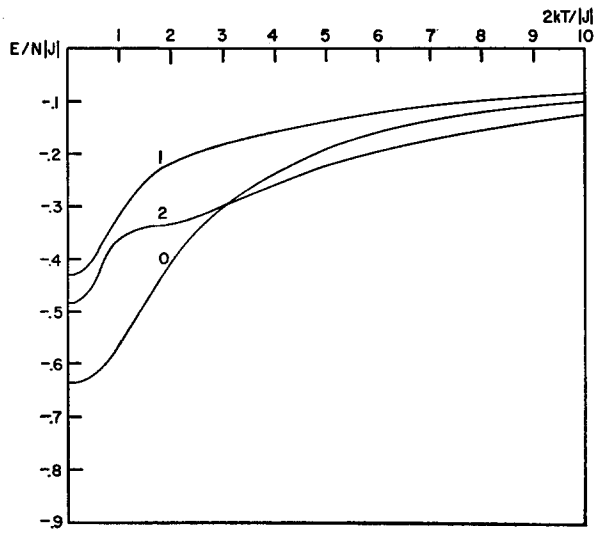


FIG. 8. Energy of ferromagnetic case. 0, 1, and 2 indicate the zeroth- , the first- , and the second-order results, respectively.

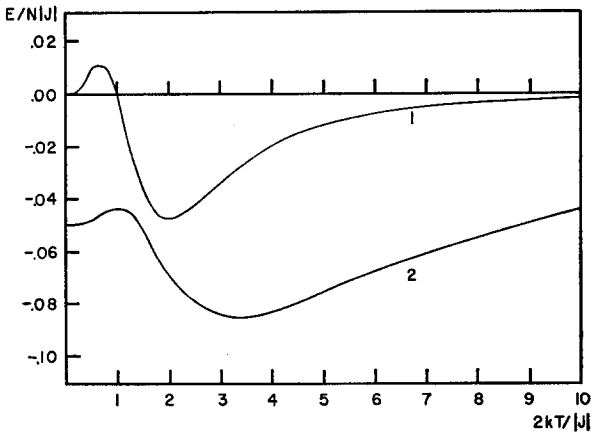


FIG. 10. Anomalous graph component (denoted by 1) and ground-state graph component (denoted by 2) of the second-order correction to the energy.

$(J/kT)^5$ and $(J/kT)^5(m\mathfrak{C}/kT)^2$ [Eq. (17)].[2] The agreement is complete.

On the other hand, from the asymptotic expansion of $A_{nm}(K)$ (see Appendix II) in powers of $1/K$, we can get the low-temperature asymptotic expansion. First-order specific heat is shown to be linear in T in both ferro- and antiferromagnetic cases at low temperatures using this asymptotic expansion. To the problem "Is a linked-cluster expansion asymptotic or convergent?", it seems probable that a linked cluster expansion itself at finite temperatures is convergent but shows an asymptotic nature when each term is replaced by its asymptotic expansion. Accordingly, zero-temperature linked cluster expansions seem to be always of asymptotic nature.

Table I shows values of the ground-state energy of linear Heisenberg antiferromagnet according to several calculations. The agreement of our results with the exact value is quite satisfactory.

Expressing the "ternary" (J_x, J_y, J_z) antiferromagnetic Heisenberg model in trilinear coordinate (Fig. 7), the ground-state energy on a line OF is given by (7.12) of the present paper, that on OC is given by Walker[13] and by Orbach[12], and that on AB is given by one of the authors[20] and by Lieb, Schulz, and Mattis.[22]

In Figs. 8–13, energy, specific heat, and suscepti-

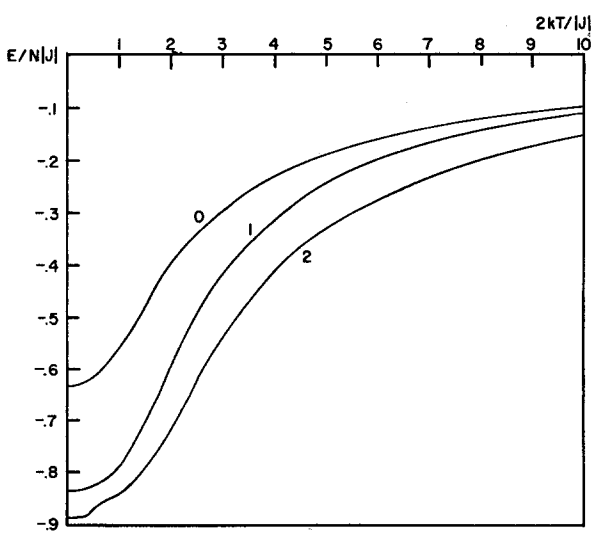


FIG. 9. Energy of antiferromagnetic case. 0, 1, and 2 indicate the zeroth- , the first- , and the second-order results, respectively.

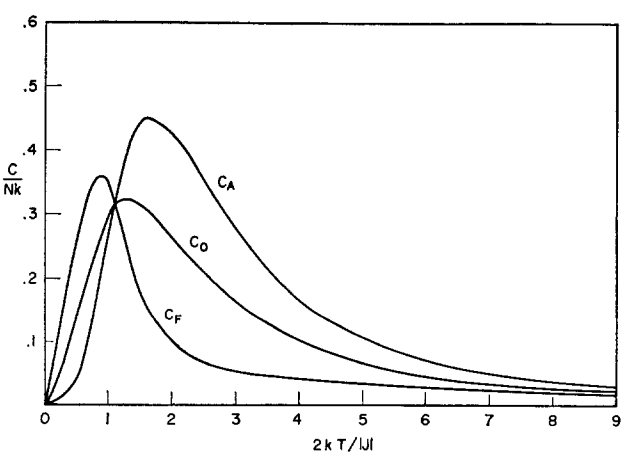


FIG. 11. Specific heat: C_0, C_F, and C_A denote the zeroth-order result, the first-order ferromagnetic result, and the first-order antiferromagnetic result, respectively.

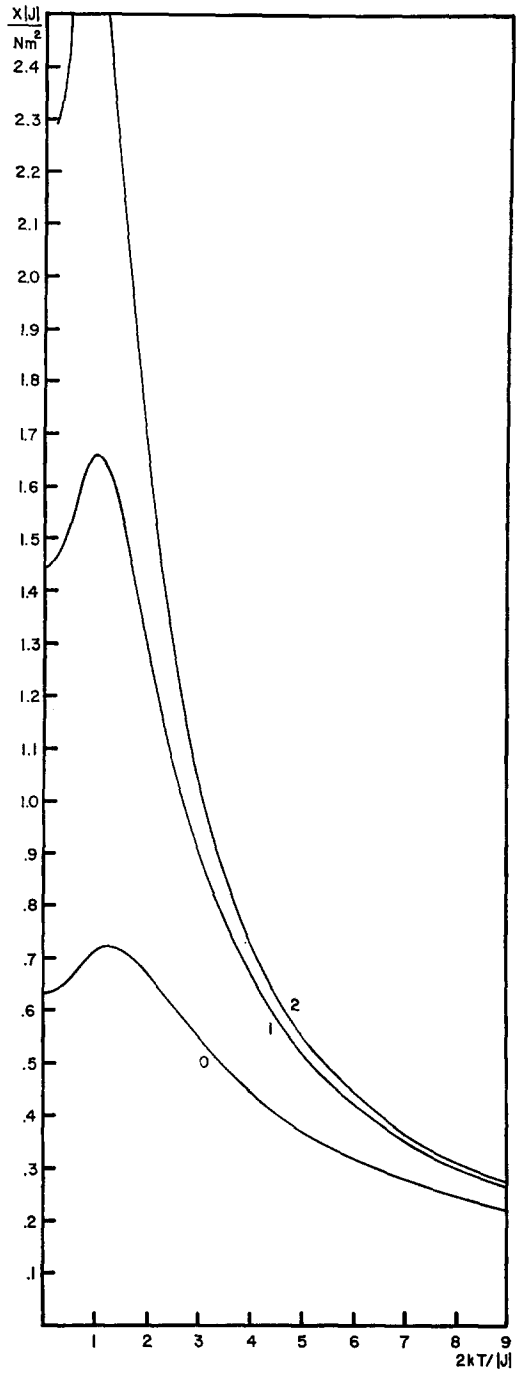


FIG. 12. Susceptibility of the ferromagnetic case. 0, 1, and 2 indicate the zeroth, the first- and the second-order results, respectively.

bility in zero magnetic field in the case of $J_{\parallel} = \pm|J_{\perp}|$ are listed. Figures 8 and 9 show energy of the ferro- and antiferromagnetic cases, respectively. 0, 1, 2 indicates the zeroth, first, and second-order results, respectively. Figure 10, shows the anomalous graph component (denoted by 1) and the ground-state graph component (denoted by 2) of the second-order correction to the energy. As $T \to 0$ and $T \to \infty$, contribution of anomalous graph tends to zero, but there is a temperature range where both contributions of the anomalous and ground-state graphs are of comparable order. Figure 11 shows the specific heat, and C_0, C_F, and C_A, denote the zeroth-order results, the first-order ferromagnetic results, and the first-order antiferromagnetic results, respectively. According to the bump of the energy in the second-order results, the specific heat in the ferromagnetic case has a broad second maximum near $2kT/|J| \sim 2$, and that of antiferromagnetic case a small second maximum near $2kT/|J| \sim 0.5$. These second maxima in the second-order calculation, figures of which are not shown, may be due to the insufficiency of the perturbation.

Figures 12 and 13 show the susceptibility of the ferro- and antiferromagnetic cases, respectively. Figure 14 shows the anomalous graph component and the ground-state graph component of the second-order corrections to the susceptibility. (Notations are the same as Figs. 8, 9, and 10. A broken line in Fig. 13 is the susceptibility of a finite system[20] $N = 6$.) Contribution of the anomalous graph component is

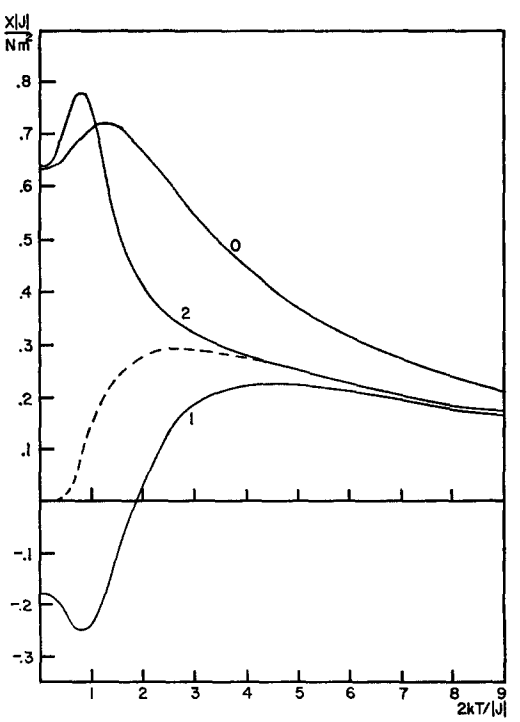


FIG. 13. Susceptibility of the antiferromagnetic case. 0, 1, and 2 indicate the zeroth, the first- and second-order results, respectively. A broken line indicates the susceptibility of a finite system $N = 6$.

larger than the ground-state graph component at low temperatures.

Numerical results in the present paper, especially satisfactory agreement of the ground-state energy and slow convergence of the low-temperature susceptibility, seem to show that the perturbation $H_A + H_B$ is much better than to regard only H_B as a perturbation, but is still not small when $J_\parallel$ becomes of the comparable order of $J_\perp$.

A method to improve the approximation is to be mentioned; that is, to determine the perturbed part of the Hamiltonian so as to eliminate all graphs with self-closed lines. The Hartree–Fock approximation[29] can be reproduced as a result of the first-order perturbation in this method. Ferromagnetic spontaneous magnetization has been obtained by this method. Detailed discussions and calculation up to second order is to be reported in another paper. The value of the antiferromagnetic susceptibility at absolute-zero temperature has been evaluated to be 0.2800 (first approximation) and 0.2590 (second approximation). The values in the present paper are to be compared with these values, and also with 0.2372 (Hulthén, Ref. 11), 0.2222 [V. M. Fain, referred to in Soviet Phys.—JETP 15, 131 (1962)], and 0.2024 [R. Griffiths, Phys. Rev. 133, A768 (1964)].

ACKNOWLEDGMENTS

The authors wish to thank Professor T. L. Hill, Professor G. S. Rushbrooke, Professor R. Mazo, and Professor T. Oguchi for their valuable discussions.

Appreciation is also expressed to the Statistical Laboratory and Computing Center, University of Oregon, for the use of their IBM 1620, and to the Western Data Processing Center, University of California, for the use of their IBM 7090. One of the authors (S. K.) is also grateful for the hospitality of the University of Oregon, where this work was begun.

APPENDIX I. DIRECT MULTIPLE-POWER-SERIES EXPANSION

Direct power-series expansion of the partition function in terms of K_x, K_y, K_z, C has been carried out to check the proceding results. In this section we consider a system

$$Z = \mathrm{tr}\exp\Big[\sum (K_x\sigma_l^x\sigma_{l+1}^x + K_y\sigma_l^y\sigma_{l+1}^y + K_z\sigma_l^z\sigma_{l+1}^z + C\sigma_l^z)\Big]. \tag{I1}$$

It is to be noticed that the exchange term and Zeeman term do not commute. We expand (I1),

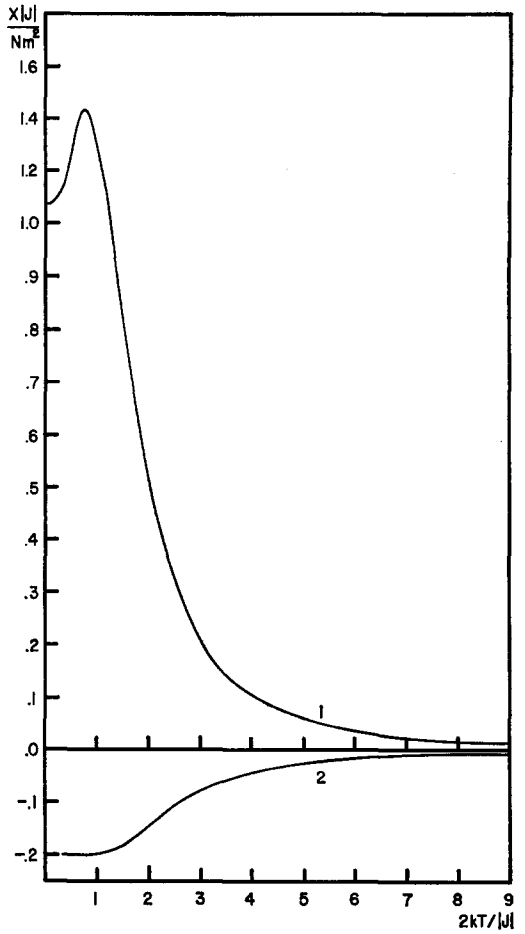


FIG. 14. Anomalous graph component (denoted by 1) and ground-state graph component (denoted by 2) of the second-order corrections to the susceptibility.

$$\begin{aligned} Z &= \mathrm{tr}\,1 + \mathrm{tr}\sum K\sigma_l\sigma_{l+1} + \mathrm{tr}\sum C\sigma_l^z \\ &\quad + \frac{1}{2!}\mathrm{tr}\Big[\Big(\sum K\sigma_l\sigma_{l+1}\Big)^2 + \sum K\sigma_l\sigma_{l+1}\sum C\sigma_l^z \\ &\quad + \sum C\sigma_l^z\sum K\sigma_l\sigma_{l+1} + \Big(\sum C\sigma_l^z\Big)^2\Big] \\ &\quad + \frac{1}{3!}\mathrm{tr}[\cdots] + \cdots \qquad \text{(I2)} \\ &= 2^N + \tfrac{1}{2}\mathrm{tr}\{\tfrac{1}{4}N(K\sigma_1\sigma_2)^2 + \tfrac{1}{8}N(K\sigma_1\sigma_2)(K\sigma_2\sigma_3) \\ &\quad + \tfrac{1}{16}(N^2-2N)(K\sigma_1\sigma_2)(K\sigma_3\sigma_4) \\ &\quad + \tfrac{1}{4}N(K\sigma_1\sigma_2)(C\sigma_1^z + C\sigma_2^z) \\ &\quad + \tfrac{1}{8}(N^2-2N)(K\sigma_1\sigma_2)C\sigma_3^z \qquad \text{(I3)} \\ &\quad + \tfrac{1}{4}N(C\sigma_1^z + C\sigma_2^z)(K\sigma_1\sigma_2) \\ &\quad + \tfrac{1}{8}(N^2-2N)C\sigma_3^z(K\sigma_1\sigma_2) + \tfrac{1}{2}N(C\sigma_1^z)^2 \\ &\quad + \tfrac{1}{4}(N^2-2N)C\sigma_1^z\sigma_2^z\} + \cdots . \qquad \text{(I3)} \end{aligned}$$

In (1–2) space,

$$K\sigma_1\sigma_2 = \begin{Bmatrix} K_x & & & K_x - K_y \\ & -K_z & K_x + K_y & \\ & K_x + K_y & -K_z & \\ K_x - K_y & & & K_z \end{Bmatrix},$$

$$C\sigma_1^z = \begin{Bmatrix} C & & & \\ & C & & \\ & & -C & \\ & & & -C \end{Bmatrix},$$

$$C\sigma_2^z = \begin{Bmatrix} C & & & \\ & -C & & \\ & & C & \\ & & & -C \end{Bmatrix}.$$

Hence

$$\begin{aligned} \mathrm{tr}\,(K\sigma_1\sigma_2)^2 &= 4(K_x^2 + K_y^2 + K_z^2), \\ \mathrm{tr}\, C\sigma_1^z K\sigma_1\sigma_2 C\sigma_2^z &= 4C^2K_z, \\ \mathrm{tr}\,(K\sigma_1\sigma_2)^3 &= -24K_xK_yK_z, \qquad \text{(I4)} \\ &\cdots \end{aligned}$$

In this way we can get the high-temperature series expansion in terms of K_x, K_y, K_z, and C. The result up to fourth order is

$$\begin{aligned} &\lim_{N\to\infty} \frac{1}{N} \log \frac{Z}{2^N} \\ &= \tfrac{1}{2}(K_x^2 + K_y^2 + K_z^2) + \tfrac{1}{2}C^2 - K_xK_yK_z + C^2K_z \\ &\quad - \tfrac{1}{12}[K_x^4 + K_y^4 + K_z^4 + 4K_x^2K_y^2 + 4K_y^2K_z^2 \\ &\quad + 4K_z^2K_x^2 + 4C^2(K_x^2 + K_y^2 + K_xK_y - 3K_z^2) + C^4] \\ &\quad + O(K^5, K^3C^2, KC^4). \qquad \text{(I5)} \end{aligned}$$

Equation (I.5) with $K_x = K_y$ agrees with the expansion (8.1)–(8.8). When $K_x = K_y = K_z$, more terms are available:

$$\begin{aligned} &\lim_{N\to\infty} \frac{1}{N} \log \frac{Z}{2^N} = \log\cosh C \\ &+ (\tfrac{3}{2}K^2 - K^3 - \tfrac{5}{4}K^4 + 3K^5 + \tfrac{47}{30}K^6 + \cdots) \\ &+ \tanh^2 C(K - \tfrac{4}{3}K^3 + \tfrac{5}{3}K^4 + \tfrac{14}{5}K^5 + \tfrac{133}{15}K^6 + \cdots) \\ &+ \tanh^4 C(-\tfrac{3}{2}K^2 - K^3 + \tfrac{23}{6}K^4 + \cdots) \\ &+ \tanh^6 C(\tfrac{10}{3}K^3 + 4K^4 + \cdots) \\ &+ \tanh^8 C(-\tfrac{33}{4}K^4 + \cdots) + \cdots \qquad \text{(I6)} \\ &= \tfrac{2}{3}K^2 - K^3 - \tfrac{5}{4}K^4 + 3K^5 + \tfrac{47}{30}K^6 + \cdots \\ &\quad + \tfrac{1}{2}C^2(1 + 2K - \tfrac{5}{3}K^3 + \tfrac{10}{3}K^4 + \tfrac{28}{5}K^5 \\ &\quad + \tfrac{266}{15}K^6 + \cdots) + C^4(-\tfrac{2}{3}K - \tfrac{3}{2}K^2 + \cdots) \\ &\quad + \cdots . \qquad \text{(I7)} \end{aligned}$$

Equation (I7) is a Kirkwood-type expansion obtained by Rushbrooke and Wood, and (I6) is Opechowski-type expansion.

APPENDIX II. ASYMPTOTIC EXPANSION OF $A_{m,n}$

From the definition (7.1), it is easily seen that

$$\begin{aligned} A_{m,n} &= \frac{2}{\pi}\int_0^{\frac{1}{2}\pi} \cos^m\omega \tanh^n(2K\cos\omega)\, d\omega \\ &\qquad (m+n \quad \text{even}) \qquad \text{(II1)} \\ &= 0 \qquad (m+n \quad \text{odd}). \qquad \text{(II2)} \end{aligned}$$

$A_{m,n}$ is an even function of K for an even n and an odd function for an odd n. Except the case $n = 0$, it has an essential singularity at $K = \infty$ and the asymptotic expression for large values of K will be derived in this appendix.

Differentiation with respect to K leads immediately to a recurrence relation

$$\begin{aligned} A_{m+1,n+2} &= A_{m+1,n} \\ &\quad - [1/2(n+1)](d/dK)A_{m,n+1}. \qquad \text{(II3)} \end{aligned}$$

Particularly, for $m = 0$ and $m = 1$, (II3) becomes

$$A_{1,2n+1} = A_{1,2n-1} - (1/4n)(d/dK)A_{0,2n}, \qquad \text{(II4)}$$

and

$$\begin{aligned} A_{2,2n+2} &= A_{2,2n} \\ &\quad - [1/2(2n+1)](d/dK)A_{1,2n+1}. \qquad \text{(II5)} \end{aligned}$$

Noticing the following transformation of an integral,

$$\begin{aligned} &\int_0^{\frac{1}{2}\pi} (1 - \cos^2\omega)\tanh^{2n+2}(2K\cos\omega)\, d\omega \\ &= \int_0^{\frac{1}{2}\pi} (1 - \cos^2\omega)\tanh^{2n}(2K\cos\omega)\, d\omega \\ &\quad + \frac{1}{2K(2n+1)}\int_0^{\frac{1}{2}\pi} \sin\omega \frac{\partial}{\partial\omega}\tanh^{2n+1}(2K\cos\omega)\, d\omega, \end{aligned}$$

and performing partial integration in the last term of the above expression, we obtain another recurrence relation,

$$A_{0,2n+2} = A_{0,2n} - A_{2,2n} + A_{2,2n+2} - [1/2K(2n+1)]A_{1,2n+1}. \quad \text{(II6)}$$

Equations (II5) and (II6) lead to an important recurrence relation

$$A_{0,2n+2} = A_{0,2n} - \frac{1}{2(2n+1)}\frac{1}{K}\frac{d}{dK}(KA_{1,2n+1}). \quad \text{(II7)}$$

Now

$$A_{2m,0} = (1/\pi^{\frac{1}{2}})[\Gamma(m+\tfrac{1}{2})/\Gamma(m+1)], \quad \text{(II8)}$$

where $\Gamma(x)$ denotes the gamma function.

On the other hand, a derivation of the asymptotic expression of $A_{2m+1,1}$ is a complicated procedure. In the integrand in the rhs of (II1) with $n = 1$, $\tanh(2K\cos\omega)$ is expanded in power of an exponential function. Changing the order of the integration and the summation, we have

$$A_{2m+1,1} = I(0) + 2\sum_{p=1}^{\infty} I(4Kp) \quad (K > 0), \quad \text{(II9)}$$

where

$$I(x) = \frac{2}{\pi}\int_0^{\frac{1}{2}\pi} \cos^{2m+1}\omega \exp(-x\cos\omega)\, d\omega, \quad \text{(II10)}$$

and in particular

$$I(0) = (1/\pi^{\frac{1}{2}})[\Gamma(m+1)/\Gamma(m+\tfrac{3}{2})]. \quad \text{(II11)}$$

Making use of Mellin integral representation

$$\exp(-z) = \frac{1}{2\pi i}\int_{-\frac{1}{2}-i\infty}^{-\frac{1}{2}+i\infty} \Gamma(-s)z^s\, ds, \quad \text{(II12)}$$

and changing the order of integration in rhs of (II10), we have

$$I(x) = \frac{1}{2\pi i}\frac{2}{\pi}\int_{-\frac{1}{2}-i\infty}^{-\frac{1}{2}+i\infty} \Gamma(-s)x^s\, ds \int_0^{\frac{1}{2}\pi} \cos^{s+2m+1}\omega\, d\omega.$$

The integral with regard to ω is expressed in terms of gamma functions, and the introduction of a new variable $t = \frac{1}{2}s$ leads to

$$I(x) = \frac{1}{2\pi i}\frac{1}{\pi^{\frac{1}{2}}}\int_{-\frac{1}{2}-i\infty}^{-\frac{1}{2}+i\infty} \frac{\Gamma(-2t)\Gamma(t+m+1)}{\Gamma(t+m+\frac{3}{2})} x^{2t}\, dt.$$

In the left half of the complex t plane, the integrand has simple poles at $t = -(m+1)-q (q=0, 1, 2, 3, \cdots)$ on the real axis. Then the path of integration is shifted to the left-hand side and decomposed to a straight line from $-(m+\frac{3}{2}+N) - i\infty$ to $-(m+\frac{3}{2}+N)+i\infty$ and a contour that encircles counter-clockwise around the $(N+1)$ poles on the real axis between $t = -(m+\frac{3}{2}+N)$ and $t = -(m+\frac{1}{2})$, as shown in Fig. 15. Then we obtain

$$I(x) = 2\frac{1}{\pi^{\frac{1}{2}}}\sum_{q=0}^{N}\frac{\Gamma(q+\frac{1}{2})\Gamma(2q+2m+2)}{\Gamma(q+1)} \times x^{-2q-2m-2} + \Delta I_N, \quad \text{(II13)}$$

where

$$\Delta I_N = \frac{2}{\pi^{\frac{1}{2}}}\frac{1}{2\pi i} \times \int_{-(m+N+\frac{3}{2})-i\infty}^{-(m+N+\frac{3}{2})+i\infty} \frac{\Gamma(-2t)\Gamma(t+m+1)}{\Gamma(t+m+\frac{3}{2})} x^{2t}\, dt. \quad \text{(II14)}$$

It is shown that

$$|\Delta I_N| \leq \frac{1}{x^{2N+2m+3}\pi^{\frac{1}{2}}} \times \left|\int_{-\infty}^{\infty} \frac{\Gamma(3+2m+2N-2iT)}{\Gamma(-N+iT)} \times \Gamma(-N-\tfrac{1}{2}+iT)\, dT\right|, \quad (x > 0)$$

and the integral in the right-hand side converges to a constant which is independent of x.[30]

Now the asymptotic expression of $A_{2m+1,1}$ is given by

$$\begin{aligned} A_{2m+1,1} &= m!/\pi^{\frac{1}{2}}\Gamma(m+\tfrac{3}{2}) \\ &+ \frac{4}{\pi^{\frac{1}{2}}}\sum_{q=0}^{N}\frac{\Gamma(q+\frac{1}{2})\Gamma(2q+2m+2)}{q!} \\ &\times \sum_{p=1}^{\infty}(-1)^p(4Kp)^{-2m-2q-2} + O(K^{-2m-2q-3}) \\ &= \frac{1}{\pi^{\frac{1}{2}}}\frac{m!}{\Gamma(m+\frac{3}{2})} - \frac{2}{\pi^{\frac{1}{2}}}\sum_{q=0}^{N} B_{m+q+1} \\ &\times \frac{(1-2^{-2q-2m-1})\Gamma(q+\frac{1}{2})}{(2q+2m+2)q!}\left(\frac{\pi}{2K}\right)^{2q+2m+2} \\ &+ O(K^{-2N-2m-3}). \end{aligned} \quad \text{(II15)}$$

In the derivation above, we have used the formula

$$\sum_{p=1}^{\infty}\frac{(-1)^p}{p^{2n}} = (2^{2n-1}-1)\pi^{2n}B_n/(2n)! \quad (n \geq 2),$$

where B_n denotes a Bernoulli's number.[31]

If $A_{0,0}$ and $A_{1,1}$ are once known, the asymptotic expression of $A_{0,2n}$ and $A_{1,2n+1}$ are obtained by successive use of (II4) and (II7). Further, it is easily shown that the knowledge about $A_{2m+1,1}$ (or $A_{2m+2,0}$), together with the knowledge about $A_{2m,2n}$ (or

[30] A. L. Dixon and W. L. Ferrar, Quart. J. Math. 7, 81 (1936).

[31] The definition of Bernoulli's number is the same as that of E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge University Press, Cambridge, England, 1935), 4th ed.

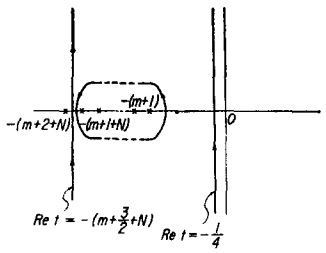


FIG. 15. Paths of the integration in the t plane.

$A_{2m+1,2n+1}$) leads, through the recurrence relation (II3), to the asymptotic expression of $A_{2m+1,2n+1}$ (or $A_{2m+2,2n}$).

The asymptotic expansion of $A_{m,n}$ has been obtained by Mazo using another method (private communication).

APPENDIX III.
Errata to "Statistical Mechanics of the Anisotropic Linear Heisenberg Model"

SHIGETOSHI KATSURA

[Phys. Rev. 127, 1508 (1962) and Phys. Rev. 129, 2835 (1962)]

Page 1509, line 3, left column, "(not the Bethe–Hulthén cyclic condition)" should be deleted.

Page 1509, line 15, left column, "$m = g\mu_B$" should read "$m = g\mu_B/2$".

Page 1509, one line from the bottom, and page 1510, line 3, "J_z" should read "$J_z/2$".

Page 1510, six lines from the bottom, "$\omega_k = 2\pi k/N$" should read "$\omega_k = \pi k/N$".

Page 1515, Fig. 2, Curve (iv$_f$) contains a numerical error and should be deleted.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 8 AUGUST 1964

On the Theory of Randomly Dilute Ising and Heisenberg Ferromagnetics

G. S. RUSHBROOKE
Physics Department, University of Newcastle upon Tyne, England
(Received 11 February 1964)

The Brout expansion for the free energy of the Ising or Heisenberg model is formally summed over all interaction graphs with not more than m vertices: the result is expressed in terms of the partition functions of isolated physical clusters, again having not more than m vertices. These partition functions are multiplied by occurrence factors closely related to the occurrence factors for the corresponding isolated physical clusters in a randomly dilute ferromagnet at low concentrations of the magnetic elements. Comparison is made with earlier work on the randomly dilute Ising and Heisenberg models.

1. INTRODUCTION

IN a series of recent papers,[1] Morgan and the author have developed a theory of the randomly dilute Ising and Heisenberg ferromagnetic models. These models are defined by the Hamiltonians

$$\mathcal{H} = -2J \sum_{(ij)} S_3^{(i)} S_3^{(j)} - g\beta H \sum_i S_3^{(i)}, \qquad (1)$$

and

$$\mathcal{H} = -2J \sum_{(ij)} \mathbf{S}^{(i)} \cdot \mathbf{S}^{(j)} - g\beta H \sum_i S_3^{(i)}, \qquad (2)$$

respectively, where J is a positive exchange energy, $\mathbf{S}^{(i)}$ is a spin vector, with components $(S_1^{(i)}, S_2^{(i)}, S_3^{(i)})$, located at site i of a physical lattice, g a gryomagnetic ratio, β the Bohr magneton, H an external magnetic field (in the 3-direction), and $\langle ij \rangle$ means that sites i and j are nearest neighbors on the lattice. For the randomly dilute problems we impose the further restriction that the sites of the physical lattice are occupied at random by magnetic and nonmagnetic elements, of which a proportion p are magnetic. There is no spin variable associated with a nonmagnetic site.

The theory has been based on the two equations

$$\tilde{\chi}\theta = \tfrac{1}{3}S(S+1)p\left[1 + \sum_{n=1}^{\infty} \frac{a_n(p)}{\theta^n}\right], \qquad (3)$$

and

$$\tilde{\chi}\theta = \tfrac{1}{3}S(S+1)p\left[1 + \sum_{n=1}^{\infty} \alpha_n(\theta)p^n\right], \qquad (4)$$

in which θ is the reduced temperature, kT/J, $\tilde{\chi}$ the reduced susceptibility, $J\chi/Ng^2\beta^2$, where N is the total number of lattice sites, and S is the maximum value of $S_3^{(i)}$. Equation (3) is essentially an expansion of susceptibility in powers of inverse tem-

[1] G. S. Rushbrooke and D. J. Morgan, Mol. Phys. 4, 1 (1961); D. J. Morgan and G. S. Rushbrooke, *ibid.*, 4, 291 (1961); 6, 477 (1963).

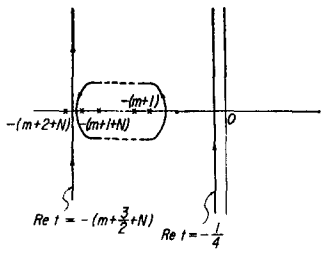


FIG. 15. Paths of the integration in the t plane.

$A_{2m+1,2n+1}$) leads, through the recurrence relation (II3), to the asymptotic expression of $A_{2m+1,2n+1}$ (or $A_{2m+2,2n}$).

The asymptotic expansion of $A_{m,n}$ has been obtained by Mazo using another method (private communication).

APPENDIX III.
Errata to "Statistical Mechanics of the Anisotropic Linear Heisenberg Model"

SHIGETOSHI KATSURA

[Phys. Rev. 127, 1508 (1962) and Phys. Rev. 129, 2835 (1962)]

Page 1509, line 3, left column, "(not the Bethe–Hulthén cyclic condition)" should be deleted.

Page 1509, line 15, left column, "$m = g\mu_B$" should read "$m = g\mu_B/2$".

Page 1509, one line from the bottom, and page 1510, line 3, "J_z" should read "$J_z/2$".

Page 1510, six lines from the bottom, "$\omega_k = 2\pi k/N$" should read "$\omega_k = \pi k/N$".

Page 1515, Fig. 2, Curve (iv$_f$) contains a numerical error and should be deleted.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 8 AUGUST 1964

On the Theory of Randomly Dilute Ising and Heisenberg Ferromagnetics

G. S. RUSHBROOKE
Physics Department, University of Newcastle upon Tyne, England
(Received 11 February 1964)

The Brout expansion for the free energy of the Ising or Heisenberg model is formally summed over all interaction graphs with not more than m vertices: the result is expressed in terms of the partition functions of isolated physical clusters, again having not more than m vertices. These partition functions are multiplied by occurrence factors closely related to the occurrence factors for the corresponding isolated physical clusters in a randomly dilute ferromagnet at low concentrations of the magnetic elements. Comparison is made with earlier work on the randomly dilute Ising and Heisenberg models.

1. INTRODUCTION

IN a series of recent papers,[1] Morgan and the author have developed a theory of the randomly dilute Ising and Heisenberg ferromagnetic models. These models are defined by the Hamiltonians

$$\mathcal{H} = -2J \sum_{(ij)} S_3^{(i)} S_3^{(j)} - g\beta H \sum_i S_3^{(i)}, \qquad (1)$$

and

$$\mathcal{H} = -2J \sum_{(ij)} \mathbf{S}^{(i)} \cdot \mathbf{S}^{(j)} - g\beta H \sum_i S_3^{(i)}, \qquad (2)$$

respectively, where J is a positive exchange energy, $\mathbf{S}^{(i)}$ is a spin vector, with components $(S_1^{(i)}, S_2^{(i)}, S_3^{(i)})$, located at site i of a physical lattice, g a gryomagnetic ratio, β the Bohr magneton, H an external magnetic field (in the 3-direction), and $\langle ij \rangle$ means that sites i and j are nearest neighbors on the lattice. For the randomly dilute problems we impose the further restriction that the sites of the physical lattice are occupied at random by magnetic and nonmagnetic elements, of which a proportion p are magnetic. There is no spin variable associated with a nonmagnetic site.

The theory has been based on the two equations

$$\tilde{\chi}\theta = \tfrac{1}{3}S(S+1)p\left[1 + \sum_{n=1}^{\infty} \frac{a_n(p)}{\theta^n}\right], \qquad (3)$$

and

$$\tilde{\chi}\theta = \tfrac{1}{3}S(S+1)p\left[1 + \sum_{n=1}^{\infty} \alpha_n(\theta)p^n\right], \qquad (4)$$

in which θ is the reduced temperature, kT/J, $\tilde{\chi}$ the reduced susceptibility, $J\chi/Ng^2\beta^2$, where N is the total number of lattice sites, and S is the maximum value of $S_3^{(i)}$. Equation (3) is essentially an expansion of susceptibility in powers of inverse tem-

[1] G. S. Rushbrooke and D. J. Morgan, Mol. Phys. 4, 1 (1961); D. J. Morgan and G. S. Rushbrooke, *ibid.*, 4, 291 (1961); 6, 477 (1963).

perature, which should be useful at sufficiently high temperatures. Equation (4) is similarly an expansion of the susceptibility in powers of the concentration p, which we might expect to be useful for small concentrations of magnetic elements. The leading term in both these expansions is the purely paramagnetic term which must obtain both at sufficiently high temperatures when the coupling between neighboring spins is insignificant, and also at sufficiently low concentrations when the magnetic elements will be essentially isolated. We must stress, of course, that these magnetic elements are assumed to be distributed *at random* on the lattice, and not in accordance with energetic considerations. A similar theory, but based on Eq. (4) only, has been developed independently by Elliott and Heap.[2]

Now Rushbrooke and Morgan were led to Eq. (4) by first evaluating the coefficients $a_1(p) \cdots a_6(p)$ of Eq. (3) and then noticing that it was possible to pick out the terms linear in p not only from these first six coefficients but from the whole series $n = 1 \cdots$. This derivation $\alpha_1(\theta)$ has never been published, because having found the answer it was at once clear that the same result could be obtained much more simply by classifying the physical clusters of neighboring magnetic elements, themselves isolated by nonmagnetic elements, which must occur on the lattice at sufficiently low concentrations; and then summing the products of the susceptibility of each type of cluster and the number of such clusters expected to occur.

It seems desirable, however, to discuss more precisely in what sense, if any, Eq. (4) is a rearrangement of Eq. (3). One reason for this is that Eq. (4), the coefficients in which have been derived by considering isolated physical clusters, does in fact prove useful for values of p which are not small. Indeed, if we expand the coefficients $\alpha_n(\theta)$ in inverse powers of θ and collect together terms in successive powers of $1/\theta$ we recover the coefficients $a_n(p)$ of Eq. (3), at least as far as comparison is possible; and these coefficients $a_n(p)$ make sense even when $p = 1$, i.e., when dealing with a purely magnetic substance. On the other hand there are certainly no isolated physical clusters to serve as a basis for discussion when $p = 1$.

The main motivation for the present work has been the desire to clarify this matter. We shall start by considering the purely magnetic case ($p = 1$), and then pass to the randomly dilute case.

[2] R. J. Elliott and B. R. Heap, Proc. Roy. Soc. (London) **A265**, 264 (1962); B. R. Heap, Proc. Phys. Soc. (London) **82**, 252 (1963).

2. GENERAL THEORY, WHEN $p = 1$, BASED ON THE BROUT EXPANSION

We shall work in terms of the Heisenberg Hamiltonian (1), and write

$$\mathcal{H} = \mathcal{H}_1 + \mathcal{H}_0, \tag{5}$$

where $\mathcal{H}_1$ denotes the pair interaction term and $\mathcal{H}_0$ the Zeeman term.

The free energy of the system at temperature T is thus given by

$$F = -kT \ln Z, \tag{6}$$

where

$$Z = \text{tr} \exp(-\mathcal{H}/kT). \tag{7}$$

Since $\mathcal{H}_0$ and $\mathcal{H}_1$ commute, we can write $Z = Z_0 Z_1$ where

$$Z_0 = \text{tr} \exp(-\mathcal{H}_0/kT) = [Q(1)]^N,$$

with

$$Q(1) = \text{tr} \exp[(g\beta H/kT)S_3] \tag{8}$$

and

$$Z_1 = \left\langle \exp\left[\frac{2J}{kT} \sum_{\langle ij \rangle} \mathbf{S}^{(i)} \cdot \mathbf{S}^{(j)}\right] \right\rangle_{(0)}, \tag{9}$$

where, for any operator 0,

$$\langle 0 \rangle_{(0)} \equiv \text{tr}(0e^{-\mathcal{H}_0/kT})/\text{tr}(e^{-\mathcal{H}_0/kT}).$$

If we expand the exponential in (9) in powers of J/kT, we obtain a high-temperature expansion of the partition function Z_1. For the Heisenberg model with which we are here concerned, this high-temperature expansion was first introduced by Opechowski[3]; the most complete study to date is that of Rushbrooke and Wood.[4] In the details of this expansion, when calculating the term pertaining to $(J/kT)^n$ we are concerned with n-line interaction graphs, not necessarily connected, with each of which is associated (i) a trace, or moment, and (ii) an occurrence factor. Of these (i) is associated with the graph itself, as an abstract entity and (ii) introduces the structure of the lattice to which the Heisenberg model is related.

For calculating, say, the leading coefficients in the high-temperature expansion of the zero-field susceptibility, this Opechowski method is quite convenient. But for general theoretical work it is easier, as well as more elegant, to write

$$F = -NkT \ln Q(1) + F_1, \tag{10}$$

[3] W. Opechowski, Physica **4**, 181 (1937).

[4] G. S. Rushbrooke and P. J. Wood, Mol. Phys. **1**, 257 (1958).

where $F_1 = -kT \ln Z_1$, and to consider not the above expansion of Z_1 but the corresponding expansion of F_1. This latter expansion is due to Brout; a particularly lucid exposition of the theory has been given by Horwitz and Callen.[5]

In this Brout expansion of F_1, the term pertaining to $(J/kT)^n$ in the expansion of the exponential in (9) is again associated with n-line interaction graphs: but there are two differences between the present expansion and that of the partition function: first, we are now concerned only with connected (one-part) graphs, and, secondly, associated with any such graph is not the moment (trace) of the operators concerned but the corresponding cumulant.

There are, of course, many topologically different connected n-line interaction graphs, and we shall label these with the indices (n, t), where t runs over the various types of n-line graph. For the nearest-neighbor-interaction problem with which we are here concerned, the Brout expansion of F_1 then reads

$$F_1 = -kT \sum_{(n,t)} \left(\frac{2J}{kT}\right)^n w_{(n,t)} K_{(n,t)} N_{(n,t)}. \qquad (11)$$

In this equation, $K_{(n,t)}$ denotes the cumulant function associated with the (n, t) interaction graph; it is an intrinsic property of this graph (and for our present purposes we have no need to be more precise).

The symbols $w_{(n,t)}$ and $N_{(n,t)}$ denote pure numbers. For defining them, it is convenient to introduce the following language to describe the abstract (nonlocalized) interaction graphs. Any such graph consists of points (vertices) joined by lines; each line links two, and only two, distinct vertices. In an (n, t) graph there are n-lines, whereby all the vertices are mutually connected. If two vertices are joined by a line we shall call them bonded. If they are joined by k lines we shall speak of a bond of multiplicity k. Then

$$w_{(n,t)} = 1/\prod_s k_s!,$$

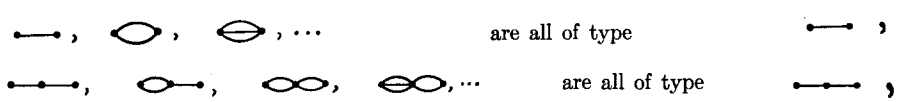
where s runs over the bonds of the (n, t) graph, and k_s is the corresponding multiplicity. Thus $w_{(n,t)}$, like $K_{(n,t)}$, is an intrinsic property of the abstract (n, t) graph. On the other hand, $N_{(n,t)}$ introduces the structure of the physical lattice to which the Heisenberg model refers. And this occurrence factor, as we shall call it, is most unambiguously defined for a given graph by regarding (for this one purpose only) the vertices of the graph as labeled. Let $X_{(n,t)}$ be the number of ways these labeled vertices can be identified with the distinguishable (localized) sites of the underlying physical lattice to which the Heisenberg Hamiltonian refers, such that (a) no two vertices are identified with the same lattice site, and (b) each bond of the graph implies that the corresponding vertices are identified with nearest-neighbor sites on the lattice. Then

$$N_{(n,t)} = X_{(n,t)}/S_{(n,t)}, \qquad (12)$$

where $S_{(n,t)}$ is the symmetry number of the (n, t) graph, i.e., the number of ways such a graph with labeled vertices can be superimposed on another such graph (identical apart from the vertex labelings) so that only bonds of equal multiplicity are in coincidence. Since $X_{(n,t)}$ introduces the physical lattice structure, $N_{(n,t)}$, unlike the other terms in (11), introduces this physical lattice structure into the expression for the free energy F_1.

3. BASIC GRAPHS

At this point it is convenient to introduce the concept of graphs of a given *basic type*. Two graphs will be said to be of the same basic type if they differ only in the multiplicities of their bonds. Thus the graphs



and so on. We shall denote these basic graphs by the indices (m, τ), where m specifies the number of *vertices* and τ distinguishes between different types of basic graph with the same number of vertices. Then for all graphs (n, t) corresponding to the same basic type (m, τ), $X_{(n,t)}$ is the same, and may conveniently be denoted $X_{(m,\tau)}$. Associated with a basic graph there is an occurrence factor $N_{(m,\tau)}$ defined by

$$N_{(m,\tau)} = X_{(m,\tau)}/S_{(m,\tau)},$$

where $S_{(m,\tau)}$ is the symmetry factor of the basic graph; and thus (11) can be written

$$F_1 = -kT \sum_{(m,\tau)} f_{m,\tau} N_{(m,\tau)}, \qquad (13)$$

[5] G. Horwitz and H. B. Callen, Phys. Rev. **124**, 1757 (1961).

where

$$f_{m,\tau} = \sum \left(\frac{2J}{kT}\right)^n w_{(n,t)} K_{(n,t)} S_{(m,\tau)}/S_{(n,t)} \qquad (14)$$

[all (n, t) for fixed (m, τ)].

Equation (13) is the basic formula of the present study. It expresses the interaction free energy as a linear combination of contributions from basic connected graphs (i.e., connected graphs without multiple bonds). The factors $f_{m,\tau}$ are complicated functions of temperature, to be discussed more fully below; but specifying a basic graph specifies the function $f_{m,\tau}$. On the other hand, to specify the number $N_{(m,\tau)}$ we need not only the specification (m, τ) of the basic graph but also to know the structure of the physical lattice to which the Heisenberg model refers. It is these occurrence factors $N_{(m,\tau)}$ which serve to differentiate one physical lattice from another. Table I expresses these occurrence factors, for basic graphs with $m = 2$, 3, and 4, in terms of appropriate lattice parameters. The notation is taken from Rushbrooke and Morgan.[1] z is the lattice coordination number; $Nzr_n/2n$ is the number of (unlabeled) closed, noncrossing circuits of n points on the lattice (passing always from a lattice site to one of its nearest neighbors); and $\frac{1}{24}Nzq_1$ is the number of tetrahedra of nearest neighbors occurring on the lattice.

TABLE I. Occurrence factors.

(m, τ)	Basic Graph	$N_{(m,\tau)}/N$
1		1
2		$\frac{1}{2}z$
3a		$\frac{1}{2}z(z-1)$
3b		$\frac{1}{6}zr_3$
4a		$\frac{1}{2}z[(z-1)^2 - r_3]$
4b		$\frac{1}{6}z(z-1)(z-2)$
4c		$\frac{1}{2}z(z-2)r_3$
4d		$\frac{1}{8}zr_4$
4e		$\frac{1}{4}zr_3(r_3-1)$
4f		$\frac{1}{24}zq_1$

We have included an isolated point among the basic graphs, although it does not occur in the expansion (13), for reasons which will be apparent shortly.

Now the crux of the whole matter is that Eq. (13) applies not only to an essentially infinite regular physical lattice (such as the body-centered cubic lattice), but to *any arbitrary set of points among which "nearest neighbors" are defined*, i.e., to any arbitrary set of points joined by bonds. (Some notes on the derivation of the Brout formula are relegated to an Appendix.) We shall call such bonded point sets "finite lattices"—though the points concerned are by no means necessarily all equivalent. In particular, Eq. (13) applies to the finite lattices illustrated by the basic graphs of Table I. And it is this which enables us to determine the functions $f_{m,\tau}$.

4. SUCCESSIVE APPROXIMATIONS TO THE FREE ENERGY F

For any finite lattice we shall denote the partition function, relative to the Hamiltonian (1) in which now the labels i, j refer to the sites of this finite lattice, by Q. For such a finite lattice, however irregular, the two parts of $\mathcal{H}$, $\mathcal{H}_1$ and $\mathcal{H}_0$, still commute and the general formalism continues to apply. Then, from (10) and (13),

$$\ln Q = \sum_{(m,\tau)} f_{m,\tau} n_{(m,\tau)}, \qquad (15)$$

where in (15) the set (m, τ) includes the isolated point, $f_1 = \ln Q(1)$, and we use $n_{(m,\tau)}$ rather than $N_{(m,\tau)}$ since we are dealing with a finite lattice.

Let us denote by $F^{(m)}$ that approximation to the free energy of a macroscopic physical lattice which results from summing over all graphs, in (11) or (13), having not more than m vertices. Then

$$F^{(m)} = -kT \sum_{(m',\tau')} f_{m',\tau'} N_{(m',\tau')}, \qquad (m \geq m' \geq 1), \qquad (16)$$

and to find the functions $f_{m',\tau'}$ in (16) we need only apply (15) to a set of finite lattices which have the structures of the basic graphs with which we are concerned.

To illustrate the general case, we shall write down the equations (15) for the ten finite lattices illustrated (as basic graphs) in Table I. By inspection, these are

$$\ln Q(1) = f_1,$$

$$\ln Q(2) = f_2 + 2f_1,$$

$$\ln Q(3a) = f_{3a} + 2f_2 + 3f_1,$$

$$\ln Q(3b) = f_{3b} + 3f_{3a} + 3f_2 + 3f_1,$$
$$\ln Q(4a) = f_{4a} + 2f_{3a} + 3f_2 + 4f_1,$$
$$\ln Q(4b) = f_{4b} + 3f_{3a} + 3f_2 + 4f_1,$$
$$\ln Q(4c) = f_{4c} + f_{4b} + 2f_{4a} + f_{3b} + 5f_{3a} + 4f_2 + 4f_1,$$
$$\ln Q(4d) = f_{4d} + 4f_{4a} + 4f_{3a} + 4f_2 + 4f_1,$$
$$\ln Q(4e) = f_{4e} + f_{4d} + 4f_{4c} + 2f_{4b} + 6f_{4a} + 2f_{3b} + 8f_{3a} + 5f_2 + 4f_1,$$
$$\ln Q(4f) = f_{4f} + 6f_{4e} + 3f_{4d} + 12f_{4c} + 4f_{4b} + 12f_{4a} + 4f_{3b} + 12f_{3a} + 6f_2 + 4f_1.$$

These equations can be solved successively for $f_1 \cdots f_{4f}$, in terms of $Q(1) \cdots Q(4f)$, and the results then substituted into equation (16), in which we take the coefficients $N_{(m',\tau')}$ from Table I. We thus find

$$-F^{(4)}/kT = \sum A^{(4)}_{m,\tau} \ln Q(m, \tau),$$

where

$$A^{(4)}_1/N = 1 - z + \tfrac{1}{2}z(z-1) - \tfrac{1}{6}z(z-1)(z-2),$$
$$A^{(4)}_2/N = \tfrac{1}{2}z - \tfrac{1}{2}z(2z - 2 - r_3) + \tfrac{1}{2}z[(z-1)^2 - r_3] + \tfrac{1}{2}z(z-1)(z-2) - z(z-2)r_3 + \tfrac{1}{4}zr_3(r_3-1),$$
$$A^{(4)}_{3a}/N = \tfrac{1}{2}z(z - 1 - r_3) - z[(z-1)^2 - r_3] - \tfrac{1}{2}z(z-1)(z-2) + \tfrac{5}{2}z(z-2)r_3 + \tfrac{1}{2}zr_4 - 2zr_3(r_3-1) + \tfrac{1}{2}zq_1,$$
$$A^{(4)}_{3b}/N = \tfrac{1}{6}zr_3 - \tfrac{1}{2}z(z-2)r_3 + \tfrac{1}{2}zr_3(r_3-1) - \tfrac{1}{6}zq_1,$$
$$A^{(4)}_{4a}/N = \tfrac{1}{2}z[(z-1)^2 - r_3] - z(z-2)r_3 - \tfrac{1}{2}zr_4 + \tfrac{3}{2}zr_3(r_3-1) - \tfrac{1}{2}zq_1,$$
$$A^{(4)}_{4b}/N = \tfrac{1}{6}z(z-1)(z-2) - \tfrac{1}{2}z(z-2)r_3 + \tfrac{1}{2}zr_3(r_3-1) - \tfrac{1}{6}zq_1,$$
$$A^{(4)}_{4c}/N = \tfrac{1}{2}z(z-2)r_3 - zr_3(r_3-1) + \tfrac{1}{2}zq_1,$$
$$A^{(4)}_{4d}/N = \tfrac{1}{8}zr_4 - \tfrac{1}{4}zr_3(r_3-1) + \tfrac{1}{8}zq_1,$$
$$A^{(4)}_{4e}/N = \tfrac{1}{4}zr_3(r_3-1) - \tfrac{1}{4}zq_1,$$
$$A^{(4)}_{4f}/N = \tfrac{1}{24}zq_1.$$

Although a little lengthy, these expressions have been given in full in order to stress their relationship with similar expressions already published elsewhere by Rushbrooke and Morgan.[1] These latter expressions, which we shall here denote by $A^{(4)}_{m,\tau}(p)$, expressed, as polynomials in p through terms in p^4, the numbers of physical clusters of magnetic elements having structures of types (m, τ) to be found in a randomly dilute lattice in which a fraction p of sites is occupied by magnetic elements. They are

$$A^{(4)}_1(p)/N = p - zp^2 + \tfrac{1}{2}z(z-1)p^3 - \tfrac{1}{6}z(z-1)(z-2)p^4,$$
$$A^{(4)}_2(p)/N = \tfrac{1}{2}zp^2 - \tfrac{1}{2}z(2z - 2 - r_3)p^3 + \{\tfrac{1}{2}z[(z-1)^2 - r_3] + \tfrac{1}{2}z(z-1)(z-2) - z(z-2)r_3 + \tfrac{1}{4}zr_3(r_3-1)\}p^4,$$
$$A^{(4)}_{3a}(p)/N = \tfrac{1}{2}z(z - 1 - r_3)p^3 - \{z[(z-1)^2 - r_3] + \tfrac{1}{2}z(z-1)(z-2) - \tfrac{5}{2}z(z-2)r_3 - \tfrac{1}{2}zr_4 + 2zr_3(r_3-1) - \tfrac{1}{2}zq_1\}p^4$$
$$A^{(4)}_{3b}(p)/N = \tfrac{1}{6}zr_3p^3 - \{\tfrac{1}{2}z(z-2)r_3 - \tfrac{1}{2}zr_3(r_3-1) + \tfrac{1}{6}zq_1\}p^4,$$

and

$$A^{(4)}_{4,\tau}(p) = A^{(4)}_{4,\tau}p^4 \qquad (\tau = a \cdots f).$$

In all cases, $A^{(4)}_{m,\tau} = A^{(4)}_{m,\tau}(p)$ with $p = 1$. Indeed we can say more. If we had multiplied the occurrence numbers $N_{(m,\tau)}$ of Table I by p^m before computing the coefficients $A^{(4)}_{m,\tau}$ we should simply have obtained the coefficients $A^{(4)}_{m,\tau}(p)$ already listed by Rushbrooke and Morgan. And this, of course, is just what we should have to do if using the Brout expansion to calculate the free energy of a randomly dilute lattice.

Before proving this result quite generally, i.e., for all orders of approximation, we shall state it formally. In order to do so, it is convenient to introduce the following notation:

> A basic graph of m vertices will be denoted, as above, by the suffices m, τ in curved brackets, i.e., (m, τ); the corresponding physical cluster will be denoted by the same suffices in square brackets, i.e., $[m, \tau]$.

As isolated entities, these finite sets of bonded points are, of course, identical. But if we ask how frequently these occur on a lattice, finite or infinite, the answers are different. Thus the basic graph (3a) occurs, on an effectively infinite lattice of N points, $\frac{1}{2}Nz(z-1)$ times; whereas the physical cluster $[3a]$ occurs $\frac{1}{2}Nz\,(z - 1 - r_3)$ times. Two points of a physical cluster are not allowed to be nearest neighbors unless already joined by a bond, in the definition of the cluster. To further emphasize this distinction, we shall denote the number of occurrences of the basic graph (m, τ) on an effectively infinite physical lattice by $N_{(m,\tau)}$ as above; and denote the corresponding number of occurrences of

the physical cluster $[m, \tau]$ by $N_{[m,\tau]}$.[6] In the case of randomly dilute infinite physical lattices, having a concentration p of magnetic elements, for sufficiently small p these elements will necessarily be grouped in finite physical clusters. And the number of such clusters of type $[m, \tau]$ denoted by $P_{[m,\tau]}(p)$ will be given by

$$P_{[m,\tau]}(p) = \sum_s N_{[m,\tau]s} p^m (1-p)^s, \qquad (17)$$

where $N_{[m,\tau]s}$ is the number of ways we can choose, on an effectively infinite physical lattice of N points, a physical cluster of type $[m, \tau]$ having s nearest neighbors on the lattice (not themselves included in the cluster). Here we require the factors p^m in order that each site of the cluster shall itself be occupied by a magnetic element, and the factors $(1-p)^s$ in order that the cluster shall be an isolated group, surrounded by nonmagnetic elements. We denote the terms through p^m in $P_{[m',\tau']}(p)$ by $P^{(m)}_{[m',\tau']}(p)$. Then our theorem is

$$P^{(m)}_{[m',\tau']}(p) = A^{(m)}_{m',\tau'}(p), \qquad (18)$$

where $A^{(m)}_{m',\tau'}(p)$ is defined by

$$-\frac{1}{kT} F^{(m)} = \sum_{(m',\tau')} f_{m',\tau'} N_{(m',\tau')} p^{m'} \qquad (19)$$

$$= \sum_{[m',\tau']} A^{(m)}_{m',\tau'}(p) \ln Q(m', \tau'), \qquad (20)$$

where we sum over all m', τ' with $m' \leq m$, and $m \geq 1$. In passing from (19) to (20) we use the connection between the f's and ln Q's specified in Eq. (15). In other words $A^{(m)}_{m',\tau'}(p)$ is the coefficient associated with the physical cluster $[m', \tau']$ when that approximation to the free energy which results from summing the Brout expansion over graphs with not more than m vertices is rearranged in terms of the partition functions of finite physical clusters.

5. PROOF OF BASIC THEOREM

We now prove the theorem stated in Eq. (18).

Let $[m', \tau'; k]$ denote the set of physical clusters of $m' + k$ points, of which m' have, among themselves, the structure $[m', \tau']$ and of which the remaining k points are all bonded to points of the set m'. Let $N_{[m',\tau';k]}$ be the corresponding occurrence number on a physical lattice. We have in mind an effectively infinite physical lattice, but this is not essential. $N_{[m',\tau';k]}$ is the number of ways we can select $m' + k$ lattice points in accordance with the prescription $[m', \tau'; k]$: specifying not only the $m' + k$ points but also which comprise the subset $[m', \tau']$. Then Eq. (17) reads

$$P_{[m',\tau']}(p) = p^{m'} \sum_k (-1)^k p^k N_{[m',\tau';k]}. \qquad (21)$$

Let $\mathbf{T}$ denote the matrix of coefficients relating the f's to the ln Q's, i.e.,

$$\ln Q(m', \tau') = \sum_{(m'',\tau'')} t^{[m',\tau']}_{(m'',\tau'')} f_{m'',\tau''},$$

where $t^{[m',\tau']}_{(m'',\tau'')}$ is the number of ways the basic graph (m'', τ'') can be found on the physical cluster $[m', \tau']$. $\mathbf{T}$ is nonsingular; in fact, det $\mathbf{T} = 1$. Thus, in matrix shorthand, Eq. (16) reads

$$-(1/kT)F^{(m)} = (\mathbf{N}, \mathbf{f}) = (\mathbf{N}, \mathbf{T}^{-1} \ln \mathbf{Q})$$

$$= (\tilde{\mathbf{T}}^{-1}\mathbf{N}, \ln \mathbf{Q}),$$

where $\mathbf{N}$, $\mathbf{f}$ and $\ln \mathbf{Q}$ are the vectors with components $N_{(m',\tau')}$, $f_{m',\tau'}$ and $\ln Q(m', \tau')$, brackets denote inner products, and $\tilde{\mathbf{T}}$ is the transpose of $\mathbf{T}$. For $F^{(m)}$ we work only in the space $m' \leq m$. Thus, from its definition,

$$\mathbf{A} = \tilde{\mathbf{T}}^{-1} \cdot \mathbf{N},$$

and, more generally,

$$\mathbf{A}(p) = \tilde{\mathbf{T}}^{-1} \cdot \mathbf{N}(p),$$

where $\mathbf{N}(p)$ has components $N_{(m',\tau')} p^{m'}$. Therefore (18) will be established if we can prove

$$\mathbf{P}(p) = \tilde{\mathbf{T}}^{-1} \cdot \mathbf{N}(p),$$

$\mathbf{P}(p)$ having components $P_{[m',\tau']}(p)$, i.e., if we can prove

$$\tilde{\mathbf{T}} \cdot \mathbf{P}(p) = \mathbf{N}(p),$$

i.e.,

$$\sum_{[m',\tau']} t^{[m',\tau']}_{(m'',\tau'')} P_{[m',\tau']}(p) = N_{(m'',\tau'')} p^{m''}, \qquad (22)$$

where $P_{[m',\tau']}(p)$ is given by (21). And (22) will be established, as an identity in p, if we can prove

$$\sum_{[m',\tau']} N_{[m',\tau';m-m']} (-1)^{m-m'} t^{[m',\tau']}_{(m'',\tau'')}$$

$$= N_{(m'',\tau'')} \delta_{m,m''}, \qquad (23)$$

where the summation is for $m' \leq m$.

Equation (23) is self-evident when $m = m''$, for

[6] These constants, $N_{(m,\tau)}$ and $N_{[m,\tau]}$, arise in a wide variety of physical problems and, in their various applications, have an extensive bibliography. A useful source of numerical values for specific lattices is provided by Appendices III and IV, respectively, of Domb's review article on cooperative phenomena in crystals.[7] They are there called high- and low-temperature lattice constants, but the particular application to the Ising problem which gives rise to the latter designation is not one with which we are here concerned.

[7] C. Comb, Phil. Mag. Suppl. 9, Nos. 34, 35 (1960).

it then reads

$$N_{(m,\tau)} = \sum_{[m,\tau']} N_{[m,\tau']} t_{(m,\tau)}^{[m,\tau']},$$

and simply expresses the fact that every time the basic graph (m, τ) occurs on a physical lattice its m vertices occupy the sites of a cluster $[m, \tau']$, which itself lies on the physical lattice. It tells us that the coefficient of $\ln Q(m, \tau)$ in $-F^{(m)}(p)/kT$ is simply $N_{[m,\tau]}p^m$, where $N_{[m,\tau]}$ is the number of times the physical cluster to which Q refers can be found on the lattice.

To prove (23) quite generally, it suffices to prove

$$\sum_{\substack{[m',\tau'] \\ (m' \leq m)}} N_{[m',\tau';m-m']}^{[m,\tau]} (-1)^{m-m'} t_{(m'',\tau'')}^{[m',\tau']} = t_{(m'',\tau'')}^{[m,\tau]} \delta_{m,m''}, \qquad (24)$$

where $N_{[m',\tau';m-m']}^{[m,\tau]}$ is the occurrence number for $[m', \tau'; m - m']$ on $[m, \tau]$, for the m vertices involved in $[m', \tau'; m - m']$ must themselves occupy the sites of a physical cluster on the lattice. And (24), of course, is also self-evident when $m = m''$.

The left-hand side of (24) is a function of two point sets, $[m, \tau]$ and (m'', τ''), one representing a physical cluster and the other a basic graph. To form this left-hand side, we select from $[m, \tau]$ a second physical cluster $[m', \tau']$, take $(-1)^{m-m'}$, multiply by the number of times (m'', τ'') occurs on $[m', \tau']$, and sum over all permissible choices of the cluster $[m', \tau']$. Now we can express this differently: let us first choose a position of (m'', τ'') on $[m, \tau]$, then choose $[m', \tau']$ containing (m'', τ'') and contained in $[m, \tau]$, and sum the quantity $(-1)^{m-m'}$ over both these choices. We again obtain the left-hand side of (24), and we wish to prove that the sum vanishes if $m'' < m$. To prove this, we prove the stronger theorem that the sum vanishes when we consider a fixed position of (m'', τ'') on $[m, \tau]$ and sum only over possible choices of $[m', \tau']$ consistent with this. More precisely, on a physical cluster $[m, \tau]$ we first specify a basic graph (m'', τ''), i.e., we pick out m'' points themselves forming a connected set: let us label these points A. We then extend this fixed set A by adding to it $(m' - m'')$ further points B, so that the A's and B's together form a connected (i.e., bonded) set on $[m, \tau]$. If there are any points of $[m, \tau]$ left over, we shall call these points C: let the number of them be k. Then we require to sum the quantity $(-1)^k$ over all choices of the points B (including the null set) subject to the one condition that all C points are bonded on $[m, \tau]$ to either A's or B's. And we wish to prove that this sum vanishes if $m'' < m$.

To prove this theorem we shall introduce the following language. The points A on $[m, \tau]$ will be called its center: the neighbors of A's, not themselves central points, will be called the first shell: neighbors of first shell points, not themselves in the center or first shell, will be called the second shell, and so on. The last set of sites of $[m, \tau]$ reached in this way will be called its edge; the previous shell will be called the subedge.

We now observe, first, that if all the subedge sites are B's, any edge site can be either B or C. Thus for those labelings of the points for which all the subedge sites are B's, $\sum (-1)^k$ certainly vanishes, contributions canceling in pairs. Secondly, the subedge sites, or indeed, the sites of any inner shell, cannot all be C's, for this would cut off the outer sites, preventing them from being occupied either by B's forming a connected set with the A's, or by C's bonded to A's or B's. In proving that the permissible allocations of B's and C's to the noncentral sites of $[m, \tau]$ give contributions to $\sum (-1)^k$ which cancel in pairs, we shall make use, essentially, of these two elementary observations.

Consider the center (A's) and the first shell. Allocate B's and C's to the first shell. Not all allocations are permissible (e.g., not all C's if there are further sites), but this does not matter; if we choose an impermissible allocation, it will simply prove impossible to carry out the scheme. As far as outer points are concerned, these C's on first shell sites are "dead" points; they do not serve to connect B's to the A's nor to provide A or B neighbors for further C's. We therefore *remove* these points (and the bonds attached to them). If the allocation of C's to the first shell sites was permissible, we are left with a connected graph, for a detached part would consist of sites blocked by these C's from the A's, and could therefore hold neither B's nor C's. This new connected graph has as center the original A's and the B's in the original first shell. We now do the same thing again. Of course not all sites in the second shell of the original graph are in the first shell of this new graph, but this does not matter. Allocate B's and C's to the first shell of the new graph and remove the C sites. If the allocation was permissible, we are left with a connected graph, with an enlarged center of A's and B's, and so on. Eventually we reach an edge for which the subedge is all B's. The edge sites can be B or C, and the contributions cancel in pairs.

Finally, we observe that any permissible allocation of B's and C's to the sites of $[m, \tau]$ is described in this way, and uniquely so. Each labeled diagram

is counted once, and once only. Their contributions to $\sum (-1)^k$ cancel in pairs. The condition $m > m''$ simply enables us to get started on this scheme.

6. SUMMARY AND FINAL COMMENTS

Having thus established the equalities (24), (23), (22), and (18), it is useful to summarize precisely what has been proved.

We start with the Brout high-temperature expansion of the free energy and sum this over all interaction graphs having not more than m vertices. The result may be written, for a pure ferromagnetic,

$$-F^{(m)}/kT = \sum_{[m',\tau']} A^{(m)}_{m',\tau'} \ln Q(m', \tau'), \qquad (25)$$

and for a randomly dilute ferromagnetic, with a concentration p of magnetic elements

$$-F^{(m)}/kT = \sum_{[m',\tau']} A^{(m)}_{m',\tau'}(p) \ln Q(m', \tau'). \qquad (26)$$

In these equations, the partition function $Q(m', \tau')$ refers to an isolated physical cluster of structure $[m', \tau']$. The coefficients $A^{(m)}_{m',\tau'}(p)$ are polynomials in p of degree m (or less), given by the expression

$$\sum_s N_{[m',\tau'],s} p^{m'} (1 - p)^s$$

truncated after the mth power of p; see (17) and (18). This expression, untruncated, can be interpreted, for sufficiently small p, as the number of isolated clusters of type $[m', \tau']$ to be found on the randomly dilute physical lattice. But $A^{(m)}_{m',\tau'}(p)$ is still given by this rule even when p is not small. In particular,

$$A^{(m)}_{m',\tau'} = A^{(m)}_{m',\tau'}(1).$$

Since $F^{(m)}$ does justice to all graphs with not more than m vertices, and these vertices are necessarily connected, if we evaluate the partition functions $Q(m', \tau')$ for $H = 0$ and then expand the expression

$$\sum_{[m',\tau']} A^{(m)}_{m',\tau'}(p) \ln Q(m', \tau')$$

in powers of $1/\theta$, we shall necessarily reproduce the true high-temperature expansion of F/kT when $H = 0$, up to and including the term in $1/\theta^{m-1}$; there are, of course, basic graphs of m links and $m + 1$ vertices (trees).[8]

We have worked in terms of the free energy, but almost identical statements can be made concerning the zero-field susceptibility, which is simply $-\partial^2 F/\partial H^2$ at $H = 0$. Thus, to find $\alpha_n(\theta)$ in (4), we require the zero-field susceptibilities of isolated physical clusters of $n + 1$ and fewer sites [$n + 1$ because α_n is associated with p^{n+1} in the expression (4) for χ]. And if we curtail (4) at $\alpha_n(\theta)p^n$ and expand in powers of $1/\theta$, we shall obtain $a_1(p) \cdots a_n(p)$ in (3) correctly for all values of p, including $p = 1$, but not the correct expressions for $a_{n+1}(p) \cdots$.

For the particular case $p = 1$ (pure ferromagnet) these results are not particularly new. Thus Domb,[9] without giving a formal proof of the validity of the method, has advocated what amounts to the use of (25) for obtaining high-temperature series expansions, and very recently Domb and Wood[10] have used computer calculations of small cluster partition functions to extend the zero-field susceptibility high-temperature expansions for open lattices, in the case $s = \frac{1}{2}$, from $n = 6$ to $n = 8$.

Again for the pure ferromagnetic case, $p = 1$, reference must also be made to the recent work of Strieb, Callen, and Horwitz.[11] These authors have considered what amount to the first three approximations in our scheme, namely $F^{(1)}$, $F^{(2)}$, and $F^{(3)}$, in a generalized version of the theory in which $\mathcal{H}$ is first expressed not in terms of the spin variables (S_1, S_2, S_3) at each lattice site, but in terms of spin-deviation operators S_1, S_2, $S_3 - S^*$; S^* is finally chosen so as to minimize $F^{(m)}$. Apart from minor modifications to the $F^{(1)}$ approximation, the formalism of the theory is essentially unchanged; the only major difference is that the partition function $Q(m, \tau)$ has now to be evaluated not relative to the undressed Hamiltonian $\mathcal{H}$ but relative to the modified Hamiltonian

$$\mathcal{H}^* = -2J \sum_{(ij)} \mathbf{S}^{(i)} \cdot \mathbf{S}^{(j)} - \sum_i [g\beta H - 2JS^*(z - l_i)] S_3^{(i)}, \qquad (27)$$

where, in (27), i, j refer to the sites of the physical cluster $[m, \tau]$, z is the coordination number of the underlying physical lattice, and l_i is the number of bonds meeting at site i in the cluster $[m, \tau]$. Apart from making more sense at low temperatures outside the domain of validity of normal high-temperature expansions, this modified procedure appears to carry the advantage that one extra term in the high-temperature expansion of the zero-field susceptibility derived from $F^{(m)}$ is thereby given correctly. Unfortunately this treatment based on spin-devia-

[8] Actually, the term in $1/\theta^m$ is given correctly from $F^{(m)}/kT$. This is because the cumulants associated with the corresponding mth-order Brout graphs necessarily vanish when $H = 0$. But when we turn to the zero-field susceptibility this is not so. The linear chain basic graph of m links and $m + 1$ vertices contributes to the coefficient of $1/\theta^m$ in $\chi(0)kT$.

[9] See Ref. 7, Sec. 5.3.2.

[10] C. Domb and D. W. Wood, Phys. Letters **8**, 20 (1964).

[11] B. Strieb, H. B. Callen, and G. Horwitz, Phys. Rev. **130**, 1798 (1963).

tion operators does not generalize straightforwardly, along the lines of the present paper, to the case $p < 1$.

For the randomly dilute ferromagnetic problems, $p < 1$, the basic result of this paper concerning the inter-relationship of Eqs. (3) and (4) was known empirically to Rushbrooke and Morgan,[1] and, indeed, used as a checking procedure on their computations. However, the formal proof is new. We do not wish to discuss here the use of these equations for estimating the dependence of Curie temperature on concentration. This has been done elsewhere.[1,2] It is arguable that the use of Eq. (4) still requires some further justification, but at least it is reassuring that this equation makes sense even for values of p too large for isolated physical clusters of magnetic systems to play any role in determining the macroscopic properties of the assembly.

ACKNOWLEDGMENTS

The author is indebted to the Chemistry Department of the University of Oregon, Eugene, Oregon, for hospitality during the academic year 1962–63, during which this work was done; and to the National Science Foundation for financial support.

APPENDIX

Suppose that a physical quantity Z has an expansion, in powers of a parameter x, which can be written

$$Z = 1 + xM_1 + (x^2/2!)M_2 + \cdots + (x^n/n!)M_n + \cdots . \quad (1A)$$

We shall refer to the coefficients M_1, M_2 $\cdots$ as *moments*. Then ln Z has the expansion

$$\ln Z = xK_1 + (x^2/2!)K_2 + \cdots + (x^n/n!)K_n + \cdots , \quad (2A)$$

and we shall refer to the coefficients K_1, K_2 $\cdots$ as *cumulants*. K_n is determined by the moments $M_1, \cdots M_n$; in fact, from the expansion of ln $(1 + y)$ in powers of y,

$$K_n = n! \sum_k (-1)^{k-1}(k-1)! \times \sum_{(\sum ln_l = n,\ \sum n_l = k)} \prod_l M_l^{n_l}/(l!)^{n_l} n_l! . \quad (3A)$$

But $n! \sum 1/\prod_l (l!)^{n_l} n_l!$, subject to $\sum ln_l = n$, $\sum n_l = k$, is the number $p(n, k)$ of *partitions* of n quantities into k groups (where the sequence within any group and the sequence of the groups are of no consequence). Let us denote a partition of n quantities $\alpha, \beta \cdots \omega$ by $\langle \alpha \cdots \beta \rangle \langle \gamma \cdots \delta \rangle \cdots \langle \nu \cdots \omega \rangle$ and write

$$[\alpha \cdots \omega] = \sum_k \sum_{p(n,k)} (-1)^{k-1}(k-1)! \times \langle \alpha \cdots \beta \rangle \langle \gamma \cdots \delta \rangle \cdots \langle \nu \cdots \omega \rangle, \quad (4A)$$

where $\sum_{p(n,k)}$ means that we sum over all partitions of $\alpha \cdots \omega$ into k groups. Then if, in (4A), we replace all $\langle\ \rangle$'s containing l symbols by M_l, and $[\alpha \cdots \omega]$ by K_n, we obtain Eq. (3A) above, i.e., the connection between K_n and $M_1 \cdots M_n$.

The converse of (4A) is easily shown to read

$$\langle \alpha \cdots \omega \rangle = \sum_k \sum_{p(n,k)} [\alpha \cdots \beta][\gamma \cdots \delta] \cdots [\nu \cdots \omega]. \quad (5A)$$

The most important property of $[\alpha \cdots \omega]$ is expressed by the "inside and outside" rule, best explained by illustration:

$$[\alpha] = \langle \alpha \rangle,$$
$$[\alpha\beta] = \langle \alpha\beta \rangle - \langle \alpha \rangle \langle \beta \rangle,$$
$$[\alpha\beta\gamma] = \langle \alpha\beta\gamma \rangle - \langle \alpha\beta \rangle \langle \gamma \rangle - \langle \alpha\gamma \rangle \langle \beta \rangle + \langle \alpha \rangle \langle \gamma \rangle \langle \beta \rangle - \langle \alpha \rangle \langle \beta\gamma \rangle + \langle \alpha \rangle \langle \beta \rangle \langle \gamma \rangle,$$

and so on. We put each new symbol "inside" and "outside" each existing $\langle\ \rangle$, associating a minus sign with the "outside" placings. This rule is an immediate consequence of the structure of the right-hand side of (4A). It is formally the same as the rule for differentiating a quotient, in the sense that

$$[\alpha \cdots \omega] = \frac{\partial}{\partial x_\omega} \cdots \frac{\partial}{\partial x_\alpha} \times \ln \langle (1 + x_\alpha \alpha)(1 + x_\beta \beta) \cdots (1 + x_\omega \omega) \rangle \Big]_{\text{all } x's = 0} . \quad (6A)$$

We turn now to our particular application of this mathematics. We start with a Hamiltonian of the form $\mathcal{H} = \mathcal{H}_1 + \mathcal{H}_0$ where

$$\mathcal{H}_1 = -2J\mathcal{P} = -2J \sum_{(ij)} p_{ij} = -2J \sum_{s=1}^{X} p_s$$

and

$$\mathcal{H}_0 = \sum_{i=1}^{N} h_i .$$

Here p_{ij} is a pair interaction operator, and h_i a single-particle operator. The suffices $i, j \cdots$ run over the members of a given enumerable set, $1 \cdots N$, among which neighbors are defined. $\sum_{(ij)}$ is a sum

over pairs of neighbors, and we can label these pairs $s = 1, \cdots X$. $\mathcal{H}_1$ and $\mathcal{H}_0$ commute.

We are concerned with

$$\mathrm{tr}\,(e^{-\mathcal{H}/kT})/\mathrm{tr}\,(e^{-\mathcal{H}_0/kT}),$$

which we shall here call Z. Then, expanding the exponential of $-\mathcal{H}_1/kT$ we have

$$Z = 1 + xM_1 + (x^2/2!)M_2 + \cdots + (x^n/n!)M_n + \cdots,$$

with $x = 2J/kT$ and

$$M_n = \mathrm{tr}\,(\mathcal{P}^n e^{-\mathcal{H}_0/kT})/\mathrm{tr}\, e^{-\mathcal{H}_0/kT} \equiv \langle \mathcal{P}^n \rangle.$$

Consequently,

$$\ln Z = xK_1 + (x^2/2!)K_2 + \cdots + (x^n/n!)K_n + \cdots,$$

where K_n is computed by rule (4A); i.e.,

$$K_n = \sum_{p(n)} (-1)^{k-1}(k-1)!\,\langle \mathcal{P}_\alpha \cdots \mathcal{P}_\beta \rangle \times \langle \mathcal{P}_\gamma \cdots \mathcal{P}_\delta \rangle \cdots \langle \mathcal{P}_\nu \cdots \mathcal{P}_\omega \rangle, \quad (7A)$$

where the partitions refer to the n suffices $\alpha, \beta \cdots \omega$ which, after forming the right-hand side of (7A) are then suppressed. In view of the structure of $\mathcal{P}$, this can be written

$$K_n = \sum_{p(n)} (-1)^{k-1}(k-1)! \times \langle (p_1 + \cdots + p_X)_\alpha \cdots (p_1 + \cdots + p_X)_\beta \rangle \times \langle (p_1 + \cdots + p_X)_\gamma \cdots (p_1 + \cdots + p_X)_\delta \rangle \cdots \times \langle (p_1 + \cdots + p_X)_\nu \cdots (p_1 + \cdots + p_X)_\omega \rangle. \quad (8A)$$

At this point we introduce the concept of interaction graphs. Let us group together those terms in (8A), resulting from expansion of the products on the right-hand side, which, before taking the traces, involve p_1, n_1 times; p_2, n_2 times; $\cdots$ p_X, n_X times. The only restriction on the n's is $\sum_s n_s = n$. For a given set of n's we can represent the resulting contribution to K_n by a graph, at present localized on the lattice, in which the bond s has multiplicity n_s. But since all p operators are the same, the suffix serving only to specify between which two sites (or spins) it acts, the contribution from a graph depends only on the topological structure of this graph, and not on the labeling of its vertices. We can therefore classify these graphs, as abstract entities, into distinct topological types, (n, t), and we let $N_{(n,t)}$ be the number of times such an (n, t) graph can be located on the connected point set $1 \cdots N$. In the language of the text, these, of course, are multiline graphs, not basic graphs. Then

$$\ln Z = \sum_{n \geq 1} \left(\frac{2J}{kT}\right)^n G_{(n,t)} N_{(n,t)},$$

where $n!G_{(n,t)}$ is the contribution to (8A) from a given localized (n, t) interaction graph. All we really need now is to prove that $G_{(n,t)}$ vanishes for unconnected graphs; but it is useful to be more specific, and give the general expression for $G_{(n,t)}$ in terms of traces of products of interaction operators.

It is convenient to return to (3A), and write (8A) in the form

$$K_n = \sum_k (-1)^{k-1}(k-1)!\,n! \times \sum \prod_l \langle \mathcal{P}^l \rangle^{n_l}/(l!)^{n_l} n_l!, \quad (9A)$$

where the second $\sum$ is over all sets l, n_l satisfying $\sum ln_l = n$ and $\sum n_l = k$. Consider $\langle \mathcal{P}^l \rangle$ where $\mathcal{P} = \sum_{s=1}^{X} p_s$. We do not wish to assume that the operators p_s commute. Suppose we pick out from $\mathcal{P}^l$ p_1, l_1-times; p_2, l_2 times; $\cdots$, p_X, l_X times; where $\sum_s l_s = l$. These l p factors can be permuted in $l!/\prod_s l_s!$ ways. We shall denote the set $(l_1 \cdots l_x)$ by $\{l\}$, and use $\{l\}!$ for $\prod_s l_s!$ We shall denote by $\langle\langle \{l\} \rangle\rangle$ the trace of the product of these l p factors averaged over all permutations of them: e.g., for $l = 3$, $\langle\langle 1, 2, 0, 0 \cdots 0 \rangle\rangle = \frac{1}{3}\langle p_1 p_2^2 + p_2 p_1 p_2 + p_2^2 p_1 \rangle$. Then

$$K_n = \sum_k (-1)^{k-1}(k-1)!\,n! \times \sum \prod_{\{l\}} \langle\langle \{l\} \rangle\rangle^{n_{\{l\}}}/(\{l\}!)^{n_{\{l\}}} n_{\{l\}}!, \quad (10A)$$

where the second summation is again subject to $\sum ln_l = n$ and $\sum n_l = k$, but now $n_{\{l\}}$ is the number of times the factor $\langle\langle \{l\} \rangle\rangle$ occurs.

To compute $G_{(n,t)}$ we may suppose, without loss of generality, that a bond of (n, t) of multiplicity n_1 can be brought into coincidence with the physical-lattice bond 1, a bond of multiplicity n_2 can be brought into coincidence with the physical-lattice bond 2, $\cdots$ a bond of multiplicity n_X can be brought into coincidence with the physical-lattice bond X. The set of numbers $(n_1, \cdots n_X)$, subject to $\sum n_s = n$, will be denoted by $\{n\}$. In the main text we used k_1, k_2 $\cdots$ rather than these symbols n_1, n_2 $\cdots$; more precisely, the set $(k_1, k_2 \cdots)$ is the set $(n_1, n_2 \cdots)$. Then for $n!G_{(n,t)}$, we require to pick out that part of K_n for which $\sum \{l\}n_{\{l\}} = \{n\}$. Thus

$$G_{(n,t)} = \sum_k (-1)^{k-1}(k-1)! \times \sum \prod_{\{l\}} \frac{\langle\langle \{l\} \rangle\rangle^{n_{\{l\}}}}{(\{l\}!)^{n_{\{l\}}} n_{\{l\}}!}, \quad (11A)$$

where the second $\sum$ is over all sets $\{l\}$, $n_{\{l\}}$ satisfying $\sum \{l\}n_{\{l\}} = \{n\}$ and $\sum n_{\{l\}} = k$.

Now consider n symbols,

$$\alpha^{(1)} \cdots \alpha^{(n_1)}, \beta^{(1)} \cdots \beta^{(n_2)}, \cdots \omega^{(1)} \cdots \omega^{(n_X)}.$$

The number of partitions of these having the structure $n_{\{l\}}$, $\{l\}$ where l_1 is the number of α's, l_2 the number of β's, $\cdots$, l_X the number of ω's in any bracket of the partition, is

$$\{n\}!/\prod_{\{l\}} (\{l\}!)^{n_{\{l\}}} n_{\{l\}}!.$$

Consequently,

$$G_{(n,t)} = \frac{1}{\{n\}!} \sum_{p(n)} (-1)^{k-1}(k-1)! \times \langle\langle \cdots \rangle\rangle\langle\langle \cdots \rangle\rangle \cdots \langle\langle \cdots \rangle\rangle, \qquad (12A)$$

where the $\langle\langle\ \rangle\rangle$'s denote a partition of the symbols $\alpha^{(1)} \cdots \omega^{(n_X)}$, the sum is over all such partitions, and having formed this we then replace any $\langle\langle\ \rangle\rangle$ containing l_1 α's, $l_2\beta$'s $\cdots$ by $\langle\langle\{l\}\rangle\rangle$; i.e., by the corresponding normalized trace of the symmetrized product of p operators.

This final prescription can, of course, be put rather more simply. Take an (n, t) interaction graph, whose bonds have multiplicities k_1, k_2 $\cdots$, where $\sum k_s = n$. Label these bonds with n symbols $\alpha, \beta \cdots \omega$, and label the vertices $1, 2 \cdots m$. Form

$$[\alpha, \beta \cdots \omega] = \sum_{p(n)} (-1)^{k-1}(k-1)! \times \langle \alpha \cdots \beta\rangle\langle\gamma \cdots \delta\rangle \cdots \langle \nu \cdots \omega\rangle, \qquad (13A)$$

where the sum is over all partitions of these symbols, k being the number of $\langle\ \rangle$'s concerned. Now replace each symbol $\alpha, \beta \cdots$ by the corresponding p_{ii}, where i and j are vertex labels of the interaction graph: and replace each $\langle\ \rangle$ by the corresponding mean trace of the product of these interaction operators, i.e., the trace averaged over all permutations of these operators. Then $[\alpha, \beta \cdots \omega]$ becomes $(\prod_s k_s!)G_{(n,t)}$ and it is this which we denote by $K_{(n,t)}$. Consequently, writing $1/\prod_s k_s! = w_{(n,t)}$ we have

$$\ln Z = \sum_{n \geq 1} \left(\frac{2J}{kT}\right)^n w_{(n,t)} K_{(n,t)} N_{(n,t)},$$

which is Eq. (11) of the text.

That $K_{(n,t)}$ vanishes for unconnected graphs follows from Eq. (6A). For if the symbols $\alpha, \beta \cdots \omega$ fall into two groups, $\alpha_1, \beta_1 \cdots \omega_1$; $\alpha_2, \beta_2 \cdots \omega_2$, such that

$$\langle \alpha_1 \cdots \beta_1, \alpha_2 \cdots \beta_2\rangle = \langle \alpha_1 \cdots \beta_1\rangle\langle\alpha_2 \cdots \beta_2\rangle$$

for all such choices of symbols from the two groups, then, since in this case

$$\ln \langle(1 + x_\alpha\alpha)(1 + x_\beta\beta) \cdots (1 + x_\omega\omega)\rangle = \ln \langle(1 + x_{\alpha_1}\alpha_1) \cdots (1 + x_{\omega_1}\omega_1)\rangle + \ln \langle(1 + x_{\alpha_2}\alpha_2) \cdots (1 + x_{\omega_2}\omega_2)\rangle,$$

it follows from (6A) that $[\alpha, \beta \cdots \omega]$ vanishes. And for disconnected graphs, our symmetrized traces of products of p operators certainly have this factorization property. Expressed formally, if $\{l\} = \{l_1\} + \{l_2\}$ where $\{l_1\}$ and $\{l_2\}$ refer to two separate parts of (n, t), then $\langle\langle\{l\}\rangle\rangle = \langle\langle\{l_1\}\rangle\rangle\langle\langle\{l_2\}\rangle\rangle$, which is all we require.

Exact Critical Percolation Probabilities for Site and Bond Problems in Two Dimensions*

M. F. SYKES
Wheatstone Physics Laboratory, King's College, London, England
AND
J. W. ESSAM
Westfield College, London, England
(Received 3 March 1964)

An exact method for determining the critical percolation probability, p_c, for a number of two-dimensional site and bond problems is described. For the site problem on the plane triangular lattice $p_c = \frac{1}{2}$. For the bond problem on the triangular, simple quadratic, and honeycomb lattices, $p_c = 2\sin(\frac{1}{18}\pi)$, $\frac{1}{2}$, $1 - 2\sin(\frac{1}{18}\pi)$, respectively. A matching theorem for the mean number of finite clusters on certain two-dimensional lattices, somewhat analogous to the duality transformation for the partition function of the Ising model, is described.

1. INTRODUCTION

PERCOLATION processes and their applications have been discussed by many authors,[1-5] and for a general introduction, reference should be made to the recent review by Frisch and Hammersley[6] who give an extensive bibliography. In this paper we shall derive some exact critical percolation probabilities for site and bond problems in two dimensions.

A study of the series expansions for the mean number of finite clusters on the plane triangular lattice leads to the discovery of a "matching" property somewhat analogous to the duality transformation for the partition function of the Ising model introduced by Kramers and Wannier[7] and interpreted geometrically by Onsager.[8] We shall introduce the series method, notice the matching property, and show that it depends essentially on a result sometimes known as Euler's Law of the Edges. We have been able to define a general class of two-dimensional lattices for which a matching property can be established. In certain special cases the property suffices to locate the critical probability. More generally we establish that the critical probabilities of certain pairs of lattices (matching pairs) are complementary. We shall locate the critical probability for one such matching pair, the bond problem on the triangular and honeycomb lattices, by a star–triangle substitution analogous to that introduced by Onsager for the corresponding Ising problem.

Apart from the theoretical interest of these exact results, a knowledge of p_c is an invaluable aid in the interpretation of power series that arise in a study of these problems.[9] We shall examine the general problem of deriving such expansions in a subsequent paper. A brief outline of the salient results in this paper has already been given.[10]

2. MEAN NUMBER OF CLUSTERS ON A FINITE GRAPH

We consider the site problem on a general linear graph G whose sites are colored at random, being black with probability p and white with probability $q = 1 - p$. For some purposes it is convenient to emphasise the symmetry of the problem and we shall then write

$$p = p_B = 1 - p_W, \tag{2.1}$$

$$q = p_W = 1 - p_B. \tag{2.2}$$

In most applications our interest in the problem will be asymmetric in that we shall consider the black sites as the primary species and refer to small p as low density and large p as high density. We shall adopt the convention of coloring a bond joining two nearest-neighbor black sites black (black bond),

* This research has been supported in part by U. S. Dept. of the Army through its European Research Office.
[1] S. R. Broadbent, and J. M. Hammersley, Proc. Cambridge Phil. Soc. **53**, 629 (1957).
[2] V. A. Vyssotosky, S. B. Gordon, H. L. Frisch, and J. M. Hammersley, Phys. Rev. **123**, 1566 (1961).
[3] H. L. Frisch, E. Sonnenblick, V. A. Vyssotosky, and J. M. Hammersley, Phys. Rev. **124**, 1021 (1961).
[4] R. J. Elliot, B. R. Heap, D. J. Morgan, and G. S. Rushbrooke, Phys. Rev. Letters **5**, 366 (1960).
[5] C. Domb and M. F. Sykes, Phys. Rev. **122**, 77 (1961).
[6] H. L. Frisch and J. M. Hammersley, J. Soc. Ind. Appl. Math. **11**, 894 (1963).
[7] H. A. Kramers and G. H. Wannier, Phys. Rev. **60**, 252 (1941).
[8] See G. H. Wannier, Rev. Mod. Phys. **17**, 50 (1950).
[9] M. F. Sykes and J. W. Essam, Phys. Rev. **133**, A 310 (1964).
[10] M. F. Sykes and J. W. Essam, Phys. Rev. Letters **10**, 1 (1963).

TABLE I. Finite clusters on the octahedron.

Cluster type	Size	Number	Probability
	1	6	pq^4
	2	12	p^2q^4
	3	12	p^3q^3
	3	8	p^3q^3
	4	3	p^4q^2
	4	12	p^4q^2
	5	6	p^5q
	6	1	p^6

and that joining two nearest-neighbor white sites white (white bond). Bonds joining sites of opposite color will be called uncolored. A connected linear graph formed by sites and bonds of one color is called a cluster.

Any particular realization R of the probability distribution on G defines two further linear graphs R_B and R_W which are, respectively, the ensemble of black and white clusters. If we denote the number of black clusters in R_B by $n(R_B)$ and the probability of occurrence of R_B by $P(R_B; G)$ then we shall define the mean number of black clusters as the weighted average

$$K(p_B; G) = \langle n(R_B) \rangle = \sum_{R_B} P(R_B; G) n(R_B), \qquad (2.3)$$

and by symmetry the mean number of white clusters is

$$K(p_W; G) = \langle n(R_W) \rangle. \qquad (2.4)$$

For a finite graph of N sites the 2^N realizations are enumerable and for small N the function $K(p; G)$ can readily be obtained explicitly. We illustrate the perimeter method described by Domb[11] by applying it to the octahedron. If we denote the mean number of clusters of size r (that is, with r sites) by a subscript, then

$$K = \sum_r K_r. \qquad (2.5)$$

In Table I. we list all the possible connected clusters on the octahedron and group them according to size and topology together with their respective probabilities of occurrence. It will be seen that there are six possible unit clusters each with probability pq^4, and therefore

$$K_1 = 6pq^4. \qquad (2.6)$$

Likewise,

$$K_2 = 12p^2q^4. \qquad (2.7)$$

For clusters of size 3 there are two possible types but the classification is mutually disjoint. By collecting all the contributions from Table I and substituting in (2.5), we obtain the mean number of black clusters as a polynomial in p and q and we shall denote this polynomial by $K(p, q; G)$. We find

$$K(p, q; G) = 6pq^4 + 12p^2q^4 + 20p^3q^3 + 15p^4q^2 + 6p^4q + p^6, \qquad (2.8)$$

where of course G is the octahedron. It is implicit in (2.8) that $q = 1 - p$. We thus obtain the mean number of the primary or black species by substitution[12] as a function of p only,

$$K(p; G) = 6p - 12p^2 + 8p^3 + 3p^4 - 6p^5 + 2p^6. \qquad (2.9)$$

By symmetry the mean number for the secondary or white species is obtained by writing q for p in (2.9).

For some applications it is convenient to express $K(p; G)$ as a function of q; that is, to express the mean number of black clusters in terms of the probability for white sites. This substitution is particularly appropriate to an investigation of the high-density region where q is small, and a suitable variable for

[11] C. Domb, Nature **184**, 509 (1959). (Report of the Physical Society Conference on Fluctuation Phenomena and Stochastic Processes held at Birkbeck College, London.)

[12] The function $K(p; G)$ can be written as a polynomial in p and q in more than one way since these are dependent variables. We shall reserve the symbol $K(p, q; G)$ to denote the polynomial that results from application of the perimeter method and which will be fundamental to our subsequent treatment of series expansions. By reversing the roles of p and q on both sides of (2.8), the mean number of white clusters is obtained.

the derivation of series developments. We shall write

$$K(p, 1-p; G) \quad \text{as} \quad K_L(p; G) = K(p; G), \tag{2.10}$$

$$K(1-q, q; G) \quad \text{as} \quad K_H(q; G), \tag{2.11}$$

and in general we shall omit the specification of G in the brackets unless it contributes anything essential to the argument. We have chosen the subscripts because K_L is most appropriate to low densities and K_H to high densities. It is clear that for any finite graph these two functions are finite polynomials in p and q, respectively, and

$$K_L(p) = K_H(1-p), \tag{2.12}$$

$$K_H(q) = K_L(1-q). \tag{2.13}$$

For the octahedron we have

$$K_L(p) = 6p - 12p^2 + 8p^3 + 3p^4 - 6p^5 + 2p^6, \tag{2.14}$$

$$K_H(q) = 1 \qquad + 3q^4 - 6q^5 + 2q^6. \tag{2.15}$$

It will be seen that the last three coefficients in (2.14) and (2.15) are identical. A similar phenomenon is found for the corresponding functions for the icosahedron for which the last eight coefficients are identical:

$$K_L(p) = 12p - 30p^2 + 20p^3 + 12p^5 + 28p^6 - 120p^7 + 75p^8 + 80p^9 - 126p^{10} + 60p^{11} - 10p^{12}, \tag{2.16}$$

$$K_H(q) = 1 \qquad + 12q^5 + 28q^6 - 120q^7 + 75q^8 + 80q^9 - 126q^{10} + 60q^{11} - 10q^{12}. \tag{2.17}$$

This "matching" property of K_L and K_H, which remains to be defined precisely, is not found to be a general property of all finite graphs. We interpret the property in the next section.

3. MEAN NUMBER OF CLUSTERS ON AN INFINITE GRAPH

It is convenient on an infinite lattice to define the mean number of clusters per site, and we shall write for a lattice of N sites

$$K(p, q) = k(p, q)N, \tag{3.1}$$

and generally write k for K where required through all the equations of the previous section. When we apply Eq. (2.5) to an infinite crystal lattice, such as the plane triangular lattice, the summation cannot be performed. However, when p is small the mean number of very large black clusters will be very small and, following Domb, we shall suppose that the double series in p and q that replaces the right-hand side of (2.5) will converge to $k(p, q)$ for small enough p. A similar observation holds when p is close to unity, for in the limit there is only one cluster, of infinite size, which fills the whole lattice. For $q > 0$ there will be a few finite black clusters surrounded by white sites, and again we shall suppose that $k(p, q)$ converges. The problem is now characterized by the existence of a critical probability p_c above which there is a nonzero probability of a site being a member of the cluster of infinite extent. For an infinite structure, "edge effects" may be supposed negligible and we find by direct enumeration of the possible clusters on the triangular lattice:

$$\begin{aligned} k_1 = pq^6, \quad k_2 = 3p^2q^8, \quad k_3 = p^3(2q^9 + 9q^{10}), \\ k_4 = p^4(3q^{10} + 12q^{11} + 29q^{12}), \\ k_5 = p^5(6q^{11} + 21q^{12} + 66q^{13} + 93q^{14}). \end{aligned} \tag{3.2}$$

We have derived further k_r and by substitution obtained the low- and high-density expansions for the number of finite black clusters as

$$k_L(p) = p - 3p^2 + 2p^3 + p^6 - p^7 + 3p^8 - 4p^9 + 9p^{10} - 15p^{11} + \cdots, \tag{3.3}$$

$$k_H(q) = \qquad q^6 - q^7 + 3q^8 - 4q^9 + 9q^{10} - 15q^{11} + \cdots. \tag{3.4}$$

These two series are valid in two separate regions (3.3) defining k_L for $p < p_c$, and (3.4) defining k_H for $p > p_c$. Again we shall suppose that these expansions converge for small values of their argu-

ment.[13] The remarkable matching of the coefficients suggests that we may write for this lattice

$$k_L(p) = \phi(p) + k_H(p), \tag{3.5}$$

$$\phi(p) = p - 3p^2 + 2p^3. \tag{3.6}$$

We shall call $\phi(p)$ the matching polynomial, and the interpretation of (3.5) is the following:

At density p the mean number of black clusters differs from the mean number of white clusters by $\phi(p)$. (3.7)

The importance of this result lies in the observation that $\phi(p)$ is a finite polynomial. We show in Sec. 7 that the property (3.5) which we shall describe as a self-matching enables the critical probability to be located as $p_c = \frac{1}{2}$. Self-matching is a very special property confined to a very limited class of infinite lattices. It is noticed to occur on any infinite 2-dimensional lattice that is fully triangulated[14] and the fitting together of the triangular faces need not form a regular pattern. We illustrate one such lattice for which the triangles do form a regular pattern in Fig. 1(a). It is also noticed for the lattice illustrated in Fig. 1(b). This lattice, which is two-dimensional but not planar, yields a site problem that, by the well known bond-to-site transformation,[15] is isomorphic with the bond problem on the simple quadratic lattice. We shall usually find it convenient to regard bond problems as site problems on the corresponding *covering lattice*. (We give a short account of the bond-on-site transformation in Appendix II.) In this way we avoid the complication of a special notation to distinguish the two problems. For the simple quadratic bond problem we find

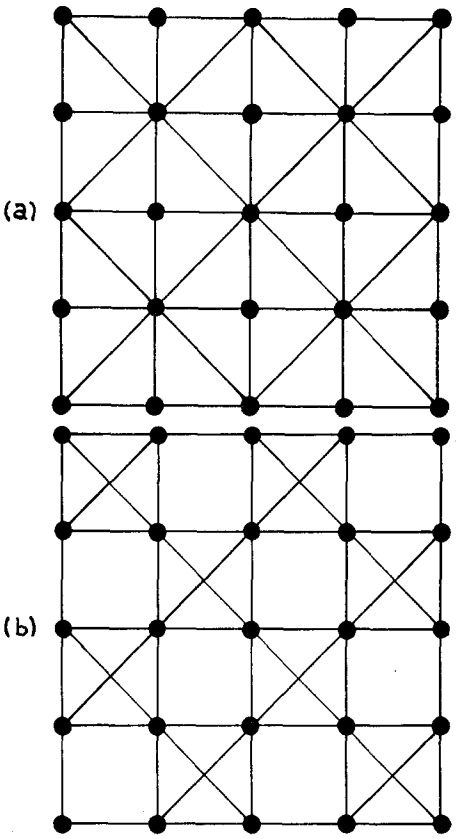


FIG. 1. Self-matching lattices. (a) Fully triangulated planar lattice; (b) covering lattice of the bond problem on the simple quadratic lattice.

$$k_L(p) = p - 3p^2 + 2p^3 + p^6 - p^7 + 3\tfrac{1}{2}p^8 - 6p^9 + 14p^{10} - 27p^{11} + 57\tfrac{1}{2}p^{12} - 118p^{13} + \cdots, \tag{3.8}$$

$$k_H(q) = q^6 - q^7 + 3\tfrac{1}{2}q^8 - 6q^9 + 14q^{10} - 27q^{11} + 57\tfrac{1}{2}q^{12} - 118q^{13} + \cdots, \tag{3.9}$$

and thus for this problem also

$$\phi(p) = p - 3p^2 + 2p^3. \tag{3.10}$$

When the mean number of clusters is expanded for the simple quadratic lattice we find

$$k_L(p) = p - 2p^2 + p^4 + p^8 - p^9 + 2p^{10} - 4p^{11} + 11p^{12} + \cdots, \tag{3.11}$$

$$k_H(q) = q^4 - q^5 + 2q^6 + 2q^8 - 3q^9 + 20q^{10} + \cdots, \tag{3.12}$$

and these two expansions do not match. However, if we also expand the mean number of clusters on the simple quadratic lattice with first- and second-neighbor bonds, we find, denoting the quantities for this case by an asterisk,

[13] It must not be assumed however that the radii of convergence of these expansions are p_c and $1 - p_c$, respectively.

[14] An infinite multiply connected planar graph all of whose finite faces are triangular.

[15] J. W. Essam and M. E. Fisher, J. Math. Phys. 2, 609 (1961).

$$k_L^*(p) = p - 4p^2 + 4p^3 + (-p^4 + p^4) - p^5 + 2p^6 + 2p^8 - 3p^9 + 20p^{10} + \cdots, \quad (3.13)$$

$$k_H^*(q) = q^8 - q^9 + 2q^{10} - 4q^{11} + 11q^{12} + \cdots, \quad (3.14)$$

and again these expansions do not match. However, it seems that (3.14) matches (3.11) and (3.13) matches (3.12). This suggests that we may write

$$k_L(p) = \phi(p) + k_H^*(p), \quad (3.15)$$

$$k_L^*(p) = \phi^*(p) + k_H(p), \quad (3.16)$$

$$\phi(p) = p - 2p^2 + p^4, \quad (3.17)$$

$$\phi^*(p) = p - 4p^2 + 4p^3 - p^4. \quad (3.18)$$

The statement equivalent to (3.7) is now

At density p the mean number of black clusters on the simple quadratic lattice differs from the mean number of white clusters on the simple quadratic lattice with first and second neighbors by $\phi(p)$. (3.19)

We show in Sec. 7 that this property which we shall call a cross-matching enables the critical probbilities of the matching pair p_c, p_c^* to be related by

$$p_c + p_c^* = 1. \quad (3.20)$$

We shall show in a subsequent paper that the coefficients in the series expansions of k_L and k_H can be related to certain enumerative problems on the lattice. By a closer examination of these enumerative problems, the matching properties illustrated in this section can be proved to hold term by term. In the following sections we develop a proof of the matching theorems that is more direct and derive them from certain simple results of the theory of linear graphs.

4. APPLICATION OF THE THEORY OF LINEAR GRAPHS TO THE SITE PROBLEM ON THE ICOSAHEDRON

In this section we establish the matching polynomial for the icosahedron as

$$12\phi(p) = 12p - 30p^2 + 20p^3 - 1. \quad (4.1)$$

We have already obtained this polynomial in Sec. 2 by a method that requires a complete enumeration of all possible clusters.

An account of the theory of linear graphs should be sought in the literature and, in particular, in the book of Berge.[16] We simply recall here by means of an example the results we require and illustrate our terminology. For a precise treatment, reference should be made to Berge, Chap. 4 (Cyclomatic Index), and Chap. 21 (Euler's Law of the Edges). In Fig. 2 we illustrate a typical planar linear graph (G). It has 12 sites, 11 bonds, and 2 finite faces. There is also an infinite face and there are 3 connected components. Denoting the number of sites by s, of bonds by b, of finite faces by f, and of the total number of faces including the infinite face by F $(= f + 1)$, and the number of connected components by n, then the *cyclomatic index* of the graph is defined to be

$$C(G) = b - s + n. \quad (4.2)$$

In our example $C(G) = 11 - 12 + 3 = 2$. The definition is not restricted to planar graphs, but for these we have an important result often known as Euler's Law of the Edges, which in its modern form states that, for a planar graph, the cyclomatic index is equal to the number of finite faces. For a planar graph we can thus write

$$f = b - s + n, \quad (4.3)$$

and we shall use this result in the form

$$n = s - b + F - 1. \quad (4.4)$$

In Fig. 3. we draw the icosahedron as a planar graph. It has one infinite (triangular) face and 19 finite triangular faces. The faces are all polygons but for a more general planar graph we have seen that this is not necessarily the case. For example, in the particular realization in which the four sites A, B, C, D are black and all the others white, the planar graph R_B has one finite face which is not a simple polygon. We shall call the sites adjacent to the face contour sites of that face, and the bonds joining two contour sites the contour bonds. In our example R_B, the points A, B, C, D are contour sites and AB, BD, AC, BD are contour bonds. To exploit the result (4.4) we require the following important Lemma:

For any realization R_B on the icosahedron, every face is either empty or contains one and only one connected component (white cluster) of R_W. (4.5)

For the present we shall regard this lemma as proved by an examination of all the possible realizations. A more general result, of which the present lemma is a particular case, is proved in Appendix I.

[16] C. Berge, *Theorie des Graphes et ses Applications* (Dunod Cie., Paris, 1958).

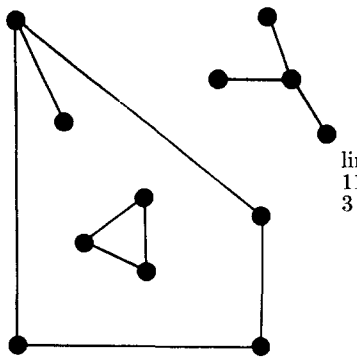


FIG. 2. Typical planar linear graph with 12 sites, 11 bonds, 2 finite faces, and 3 connected components.

Applying Euler's Law to the two graphs R_B and R_W we have from (4.4)

$$n_B = s_B - b_B + F_B - 1, \tag{4.6}$$

$$n_W = s_W - b_W + F_W - 1. \tag{4.7}$$

To apply the lemma, we observe that it follows that the number of white clusters is equal to the number of faces of R_B (black faces) that are not empty. Denoting the number of empty faces by $F_B(0)$, we can thus write

$$n_B = s_B - b_B + F_B(0) - 1 + n_W. \tag{4.8}$$

This result holds for all realizations, and we can therefore write the average sign through it and substitute

$$\langle s_B \rangle = 12p; \quad \langle b_B \rangle = 30p^2; \quad \langle n_B \rangle = K(p), \quad \langle n_W \rangle = K(q); \quad \langle F_B(0) \rangle = 20p^3. \tag{4.9}$$

The last entry results from the observation that there are 20 faces (all triangles and including the infinite face) that could be empty. We obtain

$$K(p) = 12p - 30p^2 + 20p^3 - 1 + K(q). \tag{4.10}$$

Thus the difference between the mean number of black and white clusters is the matching polynomial (4.1).

The method of this section can be applied without modification to the tetrahedron and the octahedron for which the lemma holds. For the tetrahedron

$$4\phi(p) = 4p - 6p^2 + 4p^3 - 1. \tag{4.11}$$

For any finite section of the plane triangular lattice the lemma will be found to hold except for the infinite face. The general treatment of the next section enables a proper account of the infinite face to be taken if required, but for our present purposes we assume, as it is certainly reasonable to do, that, as we require the mean number per site, we may neglect these edge effects. We then have at once as $N \to \infty$

$$\langle s_B/N \rangle = p, \quad \langle b_B/N \rangle = 3p^2, \quad \langle F_B(0)/N \rangle = 2p^3,$$

and therefore for this lattice,[17]

$$\phi(p) = p - 3p^2 + 2p^3, \tag{4.12}$$

as already surmised in Sec. 3.

We observe that this polynomial vanishes at $p = 0$, $\frac{1}{2}$, and 1—a result that is readily understood from (3.7).

5. GENERAL MATCHING PROPERTY FOR DECORATED MOSAICS

We shall use the term mosaic[18] to describe a planar graph or infinite planar lattice which is connected and has no articulation points. Such a graph has finite faces whose contour bonds form non-self-intersecting polygons. For such a graph the infinite face also has a polygonal contour, and we shall use the term face without qualifications to include the infinite face.

We now choose a mosaic M (the parent mosaic) and "decorate" it by drawing in all the possible diagonals on some of its polygonal faces, inside those faces, to form a new graph L (which is not necessarily planar). We shall call the operation of drawing in all the possible diagonals of a selected face close packing, and it results in the polygonal cluster formed by the n_c contour sites becoming the complete graph[16] of n_c sites or a close-packed cluster of n_c sites.[19] The graph L will be called a *decorated mosaic*.

We define the matching graph[20] L^* of the graph L to be the decorated mosaic graph which results from close packing all those faces of the parent mosaic M of L which were not close-packed to form L.[21] It follows from the symmetry of the defi-

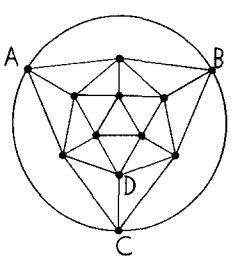


FIG. 3. The Icosahedron drawn as a planar graph.

[17] On an infinite lattice we shall write ϕ for the matching polynomials per site.

[18] A mosaic is simply a multiply-connected planar graph. We shall find the word mosaic conveniently short.

[19] C. Domb, Phil. Mag. Suppl. 9, 149 (1960).

[20] We anticipate in these definitions the results of previous sections.

[21] Triangular faces remain invariant under close packing.

nition that if L^* is the matching graph of L then L is also the matching graph of L^*. We shall call L and L^* a matching pair.

We shall take it as obvious that no bond of L crosses any bond of L^*, and the converse. For the infinite face the diagonals that result from close packing will be supposed drawn in the infinite face. A mosaic M can usually give rise to a variety of matching pairs. For example, the mosaic formed by the infinite simple quadratic lattice yields two important matching pairs[22]:

(i) L = Simple quadratic lattice. (5.1)
L^* = Simple quadratic matching lattice or simple quadratic lattice with first- and second-neighbor bonds.

(ii) L = Simple quadratic lattice in which alternate squares are replaced by tetrahedra [Sec. 3, Fig. 1 (b)]. (5.2)
$L^* = L$.

In this last example the two lattices are topologically identical and will be described as self-matching. (For our present purpose of examining percolation on graphs all of whose sites are occupied with equal probability, this definition will suffice. A slightly more detailed examination is required if the probabilities are not distributed equally and the sites are therefore labeled).

At this juncture we notice that all bond problems on mosaic lattices without multiple bonds can be made to correspond by the well known bond-to-site transformation to a site problem on a "covering lattice" which is a decorated mosaic. We notice further that the covering lattices of a lattice and its dual lattice form a matching pair of decorated mosaics. The proof of this is elementary.

As an example, the bond problem on the simple quadratic lattice is isomorphic with the site problem on the lattice L of (5.2). The property $L^* = L$ corresponds to the self-dual property of the simple quadratic lattice. We further illustrate these properties in Appendix II.

For a particular realization of the probability distribution on the parent mosaic M and the corresponding realizations on L and L^*, we have four graphs,

R_B and R_W on L, R_B^* and R_W^* on L^*,

and these graphs are not necessarily planar. However any realization, R_B say, will be made up of connected components some of which may contain subsets of points forming complete close-packed graphs. We shall extend our definition of a face to include these complete graphs as faces (close-packed faces). It is clear that these complete graphs contain no white sites and are therefore always empty.

We now state an important property of a matching pair of decorated mosaics.

Any two white sites in a face of R_B on L are connected on R_W^* on L^* and conversely if two white sites are connected on R_W^* on L^* then they both lie in the same face of R_B on L. (5.3)

The reader will readily satisfy himself of the truth of this statement by drawing a few examples. We relegate the rather tedious proof to the Appendix I.

For a decorated mosaic, Euler's Law of the Edges is not immediately applicable since the graph is not planar. In general, a realization R_B on a decorated mosaic will contain some close-packed faces. The result (4.4) will hold for the graph formed by unpacking all the close-packed faces of R_B. If we close-pack the faces in turn, each face of α points will increase the number of bonds by $\frac{1}{2}\alpha\ (\alpha - 3)$ without alteration of the total number of faces. We must therefore write

$$n = s - b + \sum \tfrac{1}{2}(\alpha^2 - 3\alpha + 2) + F - 1, \quad (5.4)$$

where the summation is taken over all the faces that are close-packed and F is the number of faces that are not close-packed. We shall make the substitution

$$F = F(0) + F', \quad (5.5)$$

where $F(0)$ is the number of empty faces (i.e., contain no white sites) that are not close-packed and write the resulting rather cumbersome expression as

$$n = \Phi + F'. \quad (5.6)$$

This equation is the modified form of Euler's Law of the Edges applicable to realizations on decorated mosaics.

6. GENERAL MATCHING THEOREM FOR A DECORATED MOSAIC

We now combine the results (5.3) and (5.6) of the previous section for the four graphs

R_B and R_W on L, R_B^* and R_W^* on L^*,

and the argument is simply a generalization of that used for the icosahedron in Sec. 4. Applying first Euler's Law (5.6) we have

$$n_B = \Phi_B + F'_B, \quad (6.1)$$

$$n_B^* = \Phi_B^* + F'^*_B, \quad (6.2)$$

[22] We neglect the infinite face.

TABLE II. Matching polynomials for the more usual lattices.

Triangular lattice (self-matching)	$\phi(p) = p - 3p^3 + 2p^3$
Simple quadratic lattice	$\phi(p) = p - 2p^2 + p^4$
Simple quadratic matching lattice (S.Q. with first- and second-neighbor bonds)	$\phi(p) = p - 4p^3 + 4p^3 - p^4$
Honeycomb lattice	$\phi(p) = \frac{1}{2}(2p - 3p^2 + p^6)$
Honeycomb matching lattice (honeycomb with first-, second- and third-neighbor bonds)	$\phi(p) = \frac{1}{2}(2p - 12p^2 + 20p^3 - 15p^4 + 6p^5 - p^6)$
Kagomé lattice (the covering lattice of the honeycomb bond problem)	$\phi(p) = \frac{1}{3}(3p - 6p^2 + 2p^3 + p^6)$
Kagomé matching lattice (the covering lattice of the triangular bond problem)	$\phi(p) = \frac{1}{3}(3p - 15p^2 + 22p^3 - 15p^4 + 6p^5 - p^6)$
Kagomé covering lattice (covering lattice of the Kagomé bond problem)	$\phi(p) = \frac{1}{6}(6p - 18p^2 + 14p^3 - 3p^4 + p^6)$
Matching lattice for Kagomé covering lattice (covering lattice of the dice lattice bond problem)	$\phi(p) = \frac{1}{6}(6p - 21p^2 + 22p^3 - 12p^4 + 6p^5 - p^6)$
Covering lattice for the simple quadratic bond problem (self-matching)	$\phi(p) = p - 3p^2 + 2p^3$

and from the matching property (5.3), we have

$$n_W^* = F'_B, \tag{6.3}$$

$$n_W = F'^*_B. \tag{6.4}$$

By substitution of (6.3) in (6.1), and (6.4) in (6.2),

$$n_B = \Phi_B + n_W^*, \tag{6.5}$$

$$n_B^* = \Phi_B^* + n_W. \tag{6.6}$$

On averaging over all the possible realizations and writing

$$\langle \Phi_B \rangle = N\phi(p), \tag{6.7}$$

$$\langle \Phi_B^* \rangle = N\phi^*(p), \tag{6.8}$$

we obtain

$$K(p; L) = N\phi(p) + K(q; L^*), \tag{6.9}$$

$$K(p; L^*) = N\phi^*(p) + K(q; L), \tag{6.10}$$

or

$$K_L(p; L) = N\phi(p) + K_H(p; L^*), \tag{6.11}$$

$$K_L(p; L^*) = N\phi^*(p) + K_H(p; L). \tag{6.12}$$

The result is conveniently expressed in words:

At density p the mean number of black clusters on L differs from the mean number of white clusters on L^* by $\phi(p)$.

The polynomial $\phi(p)$ we have called the matching polynomial of L. It follows from the relations (2.12) and (2.13) that for a matching pair

$$\phi(p) = -\phi^*(1 - p). \tag{6.13}$$

The matching polynomial is readily obtained from the definitive equation

$$N\phi(p) = \langle s - b + F(0) + \sum \tfrac{1}{2}(\alpha^2 - 3\alpha + 2) - 1 \rangle. \tag{6.14}$$

For example, for the simple quadratic covering lattice (5.2) we have, working per site, $\langle s/N \rangle = p$ $\langle b/N \rangle = 3p^2$, and the possible empty faces are a quadrilateral ($\frac{1}{2}N$) which has expectation $\frac{1}{2}p^4$, a tetrahedron ($\frac{1}{2}N$) which has weight 3 and therefore contributes $1\frac{1}{2}p^4$, or a triangle which does not occur in a tetrahedron and this has expectation $2p^3 - 2p^4$. Thus

$$\phi(p) = p - 3p^2 + 2p^3, \tag{6.15}$$

and we notice that for this lattice, which is self matching,

$$\phi(1 - p) = -p + 3p^2 - 2p^3; \tag{6.16}$$

as required by (6.13).

We summarize in Table II the matching polynomials for the more usual lattices.

7. CRITICAL PERCOLATION PROBABILITIES

The matching property (6.11–12) makes it possible to derive a property of the critical percolation probabilities of certain infinite lattices. Suppose first that the lattice is self-matching. Then from (6.9)

$$k(p; L) = \phi(p) + k(q; L). \tag{7.1}$$

Since $\phi(p)$ is a finite polynomial, its behavior is nonsingular. We shall suppose, without offering proof, that for real p $(0 \leq p \leq 1)$ the function K is singular at $p = p_c$, but nowhere else. This is to be expected in the light of exact results for closely related problems, and in particular, for percolation problems on lattices of the Bethe type for which K has been given exactly.[15] Now following closely the method of Kramers and Wannier in their derivation of the critical temperature of the Ising model we argue as follows.

From (7.1), if K is singular at p_c then it is also singular at $1 - p_c$, and if there is only one singularity these must be identical points, or

$$p_c = \tfrac{1}{2}. \tag{7.2}$$

This establishes two important percolation probabilities as $\frac{1}{2}$—that for the site problem on the triangular lattice and that for the bond problem on the simple quadratic lattice. The result (7.2) holds for any fully triangulated lattice.

For a matching pair we observe that if $k(L)$ is singular at p_c, then $k(L^*)$ is singular at $1 - p_c$. Thus if $k(L^*)$ has only one singularity at p_c^*, we must have

$$p_c + p_c^* = 1, \tag{7.3}$$

or the critical points are complementary.

To determine p_c, p_c^* we need a second relation. We have been able to find such a relation in only one case—the matching pair formed by the bond problem on the honeycomb and triangular lattices. To derive this relation we shall depart from our practice hitherto and study this pair directly as a bond problem. (This is not essential but is visually simpler.) We suppose that the bonds of the triangular lattice are occupied with probabilities ξ, η, ζ, along the usual three directions and that those of the honeycomb are occupied with probabilities x, y, z so orientated that ξ crosses x on the dual, etc. By an obvious extension of the preceding argument, if we assume there is only one critical locus for the honeycomb

$$U(x, y, z) = 0, \tag{7.4}$$

then the corresponding locus for the triangular lattice must be

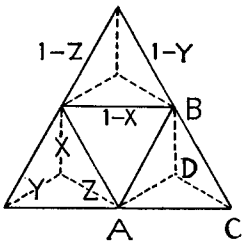


FIG. 4. Star–triangle overlapping of the triangular and honeycomb lattices with the probabilities as assigned in the text.

$$U(1 - \xi, 1 - \eta, 1 - \zeta) = 0. \tag{7.5}$$

We can satisfy this condition in the treatment that follows by supposing the bonds of the triangular lattice to be occupied with the complementary probabilities $1 - x$, $1 - y$, $1 - z$, respectively.

In Fig. 4 we draw the well known star–triangle overlapping of the two lattices. We calculate for one individual "star–triangle" the probabilities (a) of A being connected to neither B nor C; (b) of A being connected to B but not C; (c) of A being connected to C but not B; (d) of A being connected to both B and C. We find

	Honeycomb	Triangular	
(a)	$1 - xy - yz + xyz$	xz	(7.6)
(b)	$xy(1 - z)$	$xy(1 - z)$	
(c)	$zy(1 - x)$	$zy(1 - x)$	
(d)	xyz	$1 - xy - yz - xz + 2xyz$	

We notice that the conditions (b) and (c) are satisfied with equal probabilities on the two lattices as a result of our choice of complementary probabilities. We can obtain equality for both the remaining conditions if we select x, y, z to satisfy

$$1 - xy - yz - xz + xyz = 0. \tag{7.7}$$

On this locus the connectivity of each individual star–triangle will be identical. Thus the occurrence of an infinite cluster on one lattice would imply such an occurrence on the other, and by (7.5) and (7.4) these are mutually exclusive events except on the critical locus. Thus (7.7) is the critical locus and by substitution we obtain it as a function of the probabilities on the triangular lattice:

$$1 - \xi - \eta - \zeta + \xi\eta\zeta = 0. \tag{7.8}$$

From (7.8) by setting $\zeta = 0$ we obtain for the simple quadratic bond problem with two different probabilities at right angles the locus[23]

$$\xi + \eta = 1. \tag{7.9}$$

[23] The result for the asymmetric simple quadratic may be obtained directly by exploiting the self-dual property but, as remarked earlier, the matching lattice must be carefully defined and is the same lattice with the roles of ξ and η reversed.

For this case the matching polynomial factorizes as

$$\phi(\xi, \eta) = (1 - \xi - \eta)(\xi + \eta - 2\xi\eta), \qquad (7.10)$$

and this vanishes along the critical locus. Thus K is continuous there. The result can be established for any self-matching lattice.

For the symmetric triangular lattice we obtain from (7.8) the cubic

$$1 - 3p + p^3 = 0, \qquad (7.11)$$

which has only one root between 0 and 1 and yields

$$p_c = 2 \sin \tfrac{1}{18}\pi = 0.347296 \quad \text{(triangular)}, \qquad (7.12)$$
$$p_c = 0.652704 \quad \text{(honeycomb)}.$$

We notice that these results apply equally to the isomorphic site problem on the Kagomé lattice for which $p_c = 0.652704$.

8. CONCLUSIONS

We have defined a class of two-dimensional lattices (decorated mosaics) and proved a matching theorem which relates the mean number of black clusters on such a mosaic to the mean number of white clusters on another mosaic, the matching mosaic. The mean number of clusters for a matching pair is related by the results (6.9–12) which can be summarized symmetrically as

$$k(p; L) - \tfrac{1}{2}\phi(p) = k(q; L^*) - \tfrac{1}{2}\phi^*(q). \qquad (8.1)$$

When the matching pair are identical the critical probability is a $\frac{1}{2}$, and we have established this result for the site problem on the plane triangular lattice and the bond problem on the simple quadratic lattice.

We have also found that the mean number of clusters is continuous at p_c for a self-matching lattice. We shall examine the continuity of higher derivatives of K, and the continuity of K for cross-matching lattices in a subsequent paper.

The class of decorated mosaics for which (8.1) applies includes all bond problems on multiply-connected planar graphs without multiple bonds. This can be proved separately by repeating the arguments of this paper directly for the bond problem on a graph and its dual. We have preferred to work on the covering lattice so as to include the more general class of matching decorated mosaics, some of which do not correspond to bond problems.

ACKNOWLEDGMENTS

We are grateful to Professor C. Domb for criticisms of earlier drafts of this paper, and to Professor P. W. Kasteleyn and Dr. M. E. Fisher for most helpful discussions.

We are indebted to the Department of Scientific and Industrial Research for their financial support of this research.

APPENDIX I. PROOF OF THE MATCHING PROPERTY (5.3)

Any individual face of R_B on L is an area which we shall regard as made up of smaller units of area which can be classified as follows:

(1) Primary units, which are faces of the parent mosaic, and which arise from undecorated faces of M occurring in faces of R_B on L.

(2) Secondary units, which correspond to decorated faces of M or fractions thereof.

Each unit will have a polygonal contour which we shall call its edge, and we shall refer to the bonds and sites that form the polygonal boundary as its edge bonds and edge sites. These edge bonds can be black, white or uncolored for a particular realization R_B, R_W on L. Edge bonds that are white or uncolored for a particular realization will be called *internal* edges. We shall assume without proof that if two units A and B are units of the same face of R_B, then it is possible to find at least one path from A to B in the plane which crosses only internal edges. (This is equivalent to the elementary definitive property that a face has a connected interior).

We now show that it follows that any white edge site of A is connected to any white edge site of B on R_W^*. It will suffice to prove the result for two adjacent units (i.e., units having at least one internal edge in common).

In Fig. 5 let PQ be a common internal edge of A and B. Then at least one of P and Q must be white. Suppose P is white. Now suppose that X and Y are white edge sites of A and B, respectively. (We shall assume X $\neq$ P, Y $\neq$ P; if X = P and or Y = P then the proof is shorter).

Now

(1) if A is a primary unit, then on L^* this face will be close-packed and therefore X connected to P on L^*. Since both X and P are white, X is connected to P on R_W^* on L^*.

(2) If A is a secondary unit, then X must be connected to P along the edge sites of A. For if not then there is at least one black site along each of the two possible routes, and since the unit is secondary, these black sites must be connected on L which vitiates our assumption that A is a unit of one face. The connection along the edge will not be destroyed

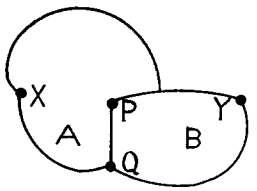


FIG. 5. Two adjacent units of area with a common internal edge PQ.

by "unpacking" the face.[24] Therefore X is connected to P on L^*. Likewise, Y is connected to P on L^*, and therefore X is connected to Y on L^*; and since X, Y, P and any sites employed on secondary units are white, X is connected to Y on R_W^*.

Thus any two white sites in a face of R_B on L are connected on R_W^* on L^*.

To establish the converse we notice that if two white sites X, Y are in different faces of R_B on L and are connected on R_W^* on L^* then at least one bond on R_W^* must cross a black contour. But no bond of L^* crosses any bond of L.

Thus if two white sites are connected on R_W^* on L^* they both lie in the same face of R_B on L.

This completes the proof of (5.3).

APPENDIX II. ILLUSTRATION OF THE BOND-TO-SITE TRANSFORMATION AND THE MATCHING CLUSTER FOR THE TRIANGULAR PRISM

The bond-to-site transformation introduced by Fisher and Essam[15] has also been described by Fisher,[25] who introduced the term covering lattice,

[24] This is true because the edge bonds of the secondary face used for the connection are also bonds of the parent mosaic.

[25] M. E. Fisher, J. Math. Phys. **4**, 620 (1961).

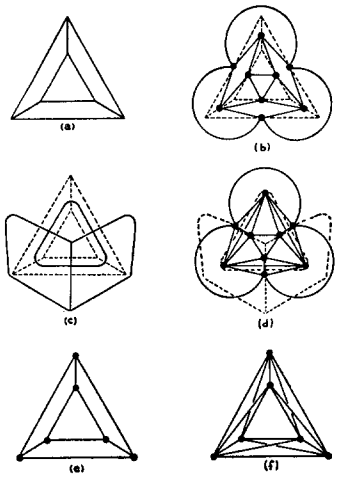


FIG. 6. Illustration of the bond-to-site transformation and the matching cluster for the triangular prism. (a) Bond problem on prism; (b) covering cluster (site problem); (c) bond problem on dual of prism; (d) covering cluster of dual of prism [this is the matching cluster for (b)]; (e) site problem on prism; (f) matching cluster for the site problem on prism.

and also by Dean[26] and Hammersley and Frisch.[6] The covering graph for the bond problem on a graph is constructed by replacing each bond of the graph by a site (placed at its center) and linking these sites together by sufficient new bonds to ensure that if two bonds of the original graph meet, then the corresponding two sites of the covering graph are joined by a direct bond and vice versa. We illustrate the transformation in Fig. 6.

[26] P. Dean, Proc. Cambridge Phil. Soc. **59**, 397 (1963).

On the Number of Self-Avoiding Walks. II

HARRY KESTEN*
Department of Mathematics, Cornell University, Ithaca, New York
(Received 11 February 1964)

Let $\chi_n(d)$ [resp. $\gamma_{2n-1}(d)$] be the number of self-avoiding walks (polygons) of n ($2n-1$) steps on the integral points in d dimensions. It is known that $\beta_d = \lim_{n\to\infty}[\chi_n(d)]^{1/n} = \lim_{n\to\infty}[\gamma_{2n-1}(d)]^{1/2n-1}$. In this paper β_d is compared with $\beta_{d,2r} = \lim_{n\to\infty}[\chi_{n,2r}(d)]^{1/n}$ where $\chi_{n,2r}(d)$ is the number of n-step walks on the integral points in d dimensions with no loops of $2r$ steps or less. In other words the walks counted in $\chi_{n,2r}(d)$ may visit the same point more than once as long as there are more than $2r$ steps between consecutive visits. It turns out that $\beta_{d,2r} - \beta_d = O(d^{-r})(d \to \infty)$ and it follows in particular that $\beta_d = 2d - 1 - 1/2d + O(1/d^2)(d \to \infty)$. It is also shown that, for suitable constants $\alpha_6 = \alpha_6(d)$ and $\alpha_7 = \alpha_7(d)$,

$$\chi_n(d) \leq \beta_n \exp[\alpha_6 n^{2/(d+2)} \log n] \text{ and } \beta^{2n-1} \exp[-\alpha_7 n^{2/(d+1)} \log n] \leq \gamma_{2n-1}(d).$$

1. INTRODUCTION

THE notation of our first paper[1] is used throughout. For completeness we repeat that w^n stands for a walk $X_0(w^n) = 0, X_1(w^n), \cdots, X_n(w^n)$ of n steps on the d-dimensional lattice such that each step goes from a point to one of its nearest neighbors and no point is visited more than once. The number of such walks will be denoted by $\chi_n = \chi_n(d)$ (even though we did not do so before, it is necessary in this paper to indicate explicitly the dependence of χ_n on d). By $W_{n,2r}$ we shall denote the class of n-step walks which have no loops of length $2r$ or less, but may visit the same point more than once if the time between visits exceeds $2r$, i.e., $W_{n,2r} = W_{n,2r}(d) =$ class of all paths $X_0 = 0, X_1, \cdots, X_n$ on the d-dimensional lattice with $|X_{i+1} - X_i| = 1$, $0 \leq i \leq n-1$, and $X_i \neq X_j$ if $0 < |i-j| \leq 2r$. $\chi_{n,2r}(d)$ is the number of elements in $W_{n,2r}(d)$. The behavior of $\chi_{n,2r}(d)$ for large n is much easier to determine than that of $\chi_n(d)$ because in the first case any step depends only on the previous $(2r-1)$ steps and not on the complete past as in the second case.[2] In principle, $\chi_{n,2r}$ can be found by computing the nth power of a certain matrix $M(d, r)$.[3,4] In practice this is not feasible but at least it allows one to conclude the existence of

$$\beta_{d,2r} = \lim_{n\to\infty}[\chi_{n,2r}(d)]^{1/n} \tag{1.1}$$

and of constants $A = A(d, r)$ and $\epsilon = \epsilon(d, r) > 0$ such that

$$\chi_{n,2r}(d) = A[\beta_{d,2r}]^n + O(\beta_{d,2r} - \epsilon)^n \quad (n \to \infty). \tag{1.2}$$

Moreover, $\beta_{d,2r}$ is the eigenvalue of largest absolute value of $M(d, r)$ which is necessarily positive and simple (by the Perron–Frobenius theorem[5]; some power of M have all entries strictly positive). In the simplest case, $r = 1$,

$$\chi_{n,2}(d) = \text{number of } n\text{-step walks without immediate reversals} = 2d(2d-1)^{n-1}, \tag{1.3}$$

and

$$\beta_{d,2} = 2d - 1. \tag{1.4}$$

For $r = 2$, corresponding to walks without immediate reversals or loops of size 4, it is still easy to write down $M(d, 2)$ and to obtain

$$\beta_{d,4} = 2d - 1 - 1/2d + O(1/d^2). \tag{1.5}$$

For higher r it becomes already cumbersome to write out $M(d, r)$ and expansions for $\beta_{d,2r}$ explicitly.

For $\chi_n(d)$ and $\beta_d = \lim_{n\to\infty}[\chi_n(d)]^{1/n}$ the best results in the literature, as far as the author knows, are[6-8]

$$\beta_d^n \leq \chi_n(d) \leq e^{\alpha_1 n^1}\beta_d^n \tag{1.6}$$

and

$$2d - 1 - \log(2d-1) \leq \beta_d \leq 2d - 1 \tag{1.7}$$

for a certain constant α_1. In addition, numerical results about β_2 and β_3 are known.[3] This ignorance

---

* Alfred P. Sloan Research Fellow.

[1] H. Kesten, J. Math. Phys. 4, 960 (1963).

[2] For the walks in $W_{n,2r}$ there is only Markov dependence (of order $2r-1$) between the steps. Walks with Markov dependence have also been considered by E. W. Montroll, J. Chem. Phys. 18, 734 (1950). The author is indebted to the referee for this reference.

[3] M. E. Fisher and M. F. Sykes, Phys. Rev. 114, 45 (1959).

[4] For some more details see also the end of Sec. 2 and Ref. 15.

[5] F. R. Gantmacher, *Applications of the Theory of Matrices* (Interscience Publishers, Inc., New York, 1959), Chap. III, Sec. 2.

[6] J. M. Hammersley and D. J. A. Welsh, Quart. J. Math., Ser. 2 13, 108 (1962).

[7] B. C. Rennie, Magy. Tud. Akad. Mat. Kut. Int. Kozlemen A6, 263 (1961).

[8] J. M. Hammersley, Sankhya, A25, 269 (1963).

justifies to some extent the derivation of asymptotic results for β_d as $d \to \infty$, even though $d > 3$ does not correspond to real physical problems (of also Fisher and Gaunt[9]). In Secs. 2 and 3 we prove

Theorem 1. For each fixed $r \geq 0$; $0 \leq \beta_{d,2r} - \beta_d = O(d^{-r})$ $d \to \infty$. (A more precise estimate of the error-term in the right-hand side is given in Secs. 2 and 3). By combining this theorem with (1.5) one obtains

Corollary 1. $\beta_d = 2d - 1 - 1/2d + O(1/d^2)$ $(d \to \infty)$. One might say that in high dimensions the main effect of the condition that the walks be self-avoiding is the exclusion of immediate reversals.

Since $\beta_{d,2r}$ is a solution of the characteristic equation for $M(d, r)$, whose coefficients are polynomials in d, it is algebraic in d. Moreover, $\beta_d \leq \beta_{d,2r} \leq \beta_{d,4}$ $(r > 2)$ so that $\beta_{d,2r}$ has an expansion[10]

$$\beta_{d,2r} = 2d - 1 - \frac{1}{2d} + \sum_{n=2h}^{\infty} a_{n,r} d^{-n/h}$$

for some positive integer $h = h(r)$ and large d. It is not obvious, however, that $h = 1$, and even though we arrive at some kind of asymptotic expansion of β_d, we cannot conclude that it involves only integral powers of d. It should be pointed out that the existence of such an asymptotic expansion for β_d was surmized by Fisher and Gaunt,[9] and even with considerable optimism computed to the coefficient of $1/d^5$. Since we only went as far as $\beta_{d,4}$ one can compare Fisher and Gaunt's expansion with Corollary 1 only up till terms of order d^{-2} (of course they agree so far). Fisher and Gaunts method is only heuristic and differs entirely from ours and the only way to check the coefficient in reference 8 would be to compute $M(d, 6)$ and then $\beta_{d,12}$ up till terms of order d^{-6}.

In Sect. 4, which is independent of Sects. 2 and 3 we improve the estimates (1.6). The exact result is given in

Theorem 2. There exist constants $\alpha_6(d)$ and $\alpha_7(d)$ such that

$$\beta_d^n \leq \chi_n(d) \leq e^{\alpha_6 n^{2/(d+2)} \log n} \beta_d^n \tag{1.8}$$

and

$$e^{-\alpha_7 n^{2/(d+2)} \log n} \beta_d^{2n-1} \leq \gamma_{2n-1}(d) \leq 4(d-1) n \beta_d^{2n-1}, \tag{1.9}$$

where $\gamma_{2n-1}(d)$ is the number of self-avoiding polygons of $2n - 1$ steps in d dimensions (i.e., the number of w^{2n-1} with $|X_{2n-1}(w^{2n-1})| = 1$; cf. Ref. 1). The left-hand side of (1.8) and the right-hand side of (1.9) where proved before by Hammersley.[11,12] The right-hand side of (1.8) improves (1.6) for all $d \geq 3$ and the left-hand side of (1.9) improves Theorem 3 of Ref. 1 for $d \geq 4$. Theorem 2 is, however, still far from the expected formulas

$$\chi_n \approx A_1 n^{\alpha_2} \beta^n, \qquad \gamma_{2n-1} \approx A_2 n^{\alpha_2} \beta^{2n-1}.$$

Actually, the following theorem follows easily from the results in Ref. 1.

Theorem 3. There exists a sequence $n_1 < n_2 < \cdots$ (depending on d, possibly) for which

$$\gamma_{4n_i+1}(d) \geq 2^{-4}(2n_i + 1)^{-6d-6} \beta_d^{4n_i}, \qquad i = 1, 2, \cdots.$$

In view of this it does not seem worthwhile to give more than an indication of the proof of Theorem 2. Finally we mention that, one can improve Theorem 4 of Ref. 1, by copying its original proof with the present Theorem 2 replacing (1.6) and its analog for γ_n. We only state the result.

Corollary 2. There exist constants $A_8(d) - A_9(d)$ such that

$$-A_8(d)(\log n)^{(d+1)/2d} n^{-(d-1)/2d} \leq \frac{\chi_{n+2}(d)}{\chi_n(d)} - \beta_d^2 \leq A_7(d)(\log n)^{(d+2)/(2d+2)} n^{-d/(2d+2)}, \tag{1.10}$$

and

$$-A_8(d)(\log n)^{(d+1)/2d} n^{-(d-1)/2d} \leq \frac{\gamma_{2n+3}(d)}{\gamma_{2n+1}(d)} - \beta^2 \leq A_9(d)(\log n)^{\frac{1}{2}} n^{-(d-1)/(2d+2)}. \tag{1.11}$$

2. THE ASYMPTOTIC BEHAVIOR OF $\beta_{d,2r} - \beta_d$

The idea of the proof of Theorem 1 is to build self-avoiding walks by "removing loops" from paths with double points. We do this in the following canonical way. Let $v = \{X_0(v) = 0, X_1(v), \cdots, X_n(v)\}$ be a walk on the d-dimensional lattice ($|X_{i+1} - X_i| = 1$, but v not necessarily self-avoiding). Let i_1 be the smallest index for which X_i is a double point, i.e., for which there exists a $j > i$ such that $X_i = X_j$, and let j_1 be the largest such j. Then the points $0, X_1, X_2, \cdots, X_{i_1-1}$ are never visited again, X_{i_1} is visited at the j_1st step but never after, and hence the walk $0, X_1, \cdots, X_{i_1}, X_{j_1+1}, X_{j_1+2}, \cdots, X_n$ (which still moves at each step from a point to

[9] M. E. Fisher and D. S. Gaunt, Phys. Rev. **133A**, 224 (1964).

[10] L. V. Ahlfors, *Complex Analysis* (McGraw-Hill Book Company, Inc., New York, 1953), Chap VI, Sec. 2.

[11] J. M. Hammersley, Quart. J. Math., Ser. 2 **12**, 250 (1961).

[12] J. M. Hammersley, Proc. Cambridge Phil. Soc. **57**, 516 (1961).

one of its nearest neighbors) has no double points in the first i_1 steps. We now determine i_2, the smallest index exceeding j_1 for which X_{i_2} is a double point of this new walk and j_2, the largest j with $X_j = X_{i_2}$. By removing the loop[13] $X_{i_2+1}, \cdots, X_{j_2}$ we obtain the walk $0, X_1, \cdots, X_{i_1}, X_{j_1+1}, \cdots, X_{i_2}, X_{j_2+1}, \cdots, X_n$ with no double points before X_{j_2+1}. Continuing in this manner we obtain a self-avoiding walk after a finite number of steps, say after removal of m loops $X_{i_k+1}, \cdots, X_{j_k}$ $k = 1, \cdots, m$ with total number of points $L = \sum_{k=1}^{m} (j_k - i_k)$. The resulting path is then a w^{n-L}. Of course $m = m(v)$ and $L = L(v)$ depend on v but $0 \leq m \leq L \leq n$, and there are at most $(n+1)^2$ possible pairs m, L. We now restrict v to $W_{n,2r}$ which has $\chi_{n,2r}$ members. One can then find $m_0 = m_0(r, n)$ and $L_0 = L_0(r, n)$ such that there are at least $\chi_{n,2r}/2(n+1)^2$ paths v in $W_{n,2r}$ which can be turned into a w^{n-L_0} by removal of m_0 loops of total number of points L_0. In general, a fixed w^{n-L_0} is obtained by removal of loops from several v but an upper bound for the number of v which lead to the same w is given in

Lemma 1. For each w^{n-L_0} there are at most

$$\binom{n - L_0 + 1}{m_0} (2d)^{L_0} \left[\frac{p(d,r)}{1 - p(d,r)}\right]^{m_0}$$

paths $v \in W_{n,2r}$ which turn into w^{n-L_0} by removal of m_0 loops, where[14] $p(d, r) = P\{S_i \neq 0, 1 \leq j \leq 2r$ but $S_j = 0$ for some $j > 2r\}$ and S_j is a simple random walk in d dimensions.

Proof: It is simpler to prove the lemma by reversing the process of loop removal, i.e., we use the fact that any $v \in W_{n,2r}$ which leads to w^{n-L_0} by removal of m_0 loops can be reconstructed by inserting m_0 loops of total number of points L_0 in w^{n-L_0}. Moreover, since $v \in W_{n,2r}$, each one of these loops must have at least $2r + 2$ points. The m_0 places where the loops should be inserted in w can be selected in $\binom{n - L_0 + 1}{m_0}$ ways. After the m_0 places have been selected, one still has to choose the m_0 loops and we need an upper bound for the number of choices for these m_0 loops. Represent the kth loop by a sequence of points

$$Y_{1,k}, Y_{2,k}, \cdots, Y_{r_k,k}, 0$$

with

$$|Y_{i+1,k} - Y_{i,k}| = 1,$$

$$|Y_{1,k}| = |Y_{r_k,k}| = 1, \qquad Y_{i,k} \neq Y_{j,k} \quad \text{for } 0 < |i - j| \leq 2r.$$

This representation is meant to indicate that insertion of this loop at the place X_i of $0, X_1, \cdots, X_{n-L_0}$ changes the path to $0, X_1, \cdots, X_i, X_i + Y_{1,k}, X_i + Y_{2,k}, \cdots, X_i + Y_{r_k,k}, X_i, X_{i+1}, \cdots, X_{n-L_0}$. The r_k's have to satisfy $r_k + 1 > 2r$ and $\sum_{k=1}^{m_0} (r_k + 1) = L_0$. The path

$$0, Y_{1,1}, Y_{2,1}, \cdots, Y_{r_1,1}, 0, Y_{1,2}, \cdots, Y_{r_2,2}, 0, \cdots, 0, Y_{1,m_0}, \cdots, Y_{r_{m_0},m_0}, 0,$$

obtained by juxtaposing the m_0 loops is then a walk of $L_0 + 1$ points which passes at least $(m_0 + 1)$ times through the origin (counting the initial and end point). However, any two visits to 0 are separated by at least $2r + 1$ steps. The number of such paths is at most

$(2d)^{L_0} P$ {a simple random walk of L_0 steps returns at least m_0 times to 0 with at least $2r + 1$ steps between successive returns} $\leq (2d)^{L_0}[p(d, r)]^{m_0}$.

This is not the required upper bound though, because a given path, $0, Y_{1,1}, \cdots, Y_{r_1,1}, 0, \cdots, Y_{r_{m_0},m_0}, 0$ can be broken up in several ways into m_0 loops. Assume that we have a path of $L_0 + 1$ points starting and ending at 0 and returning exactly $s \geq m_0 - 1$ times to 0 in between, with at least $2r + 1$ steps between successive visits to 0. Such a path can be broken up in $\binom{s}{m_0 - 1}$ ways into m_0 loops, each of length at least $2r + 1$. In fact we have to choose $m_0 - 1$ points for end points of the loops (and add the last point as end point of the m_0th loop), and these points can be chosen from the s visits to 0. By the same argument as above, the number of paths with s intermediate visits to 0 as described is at most $(2d)^{L_0}[p(d, r)]^{s+1}$, and since each such path corresponds to $\binom{s}{m_0 - 1}$ ways of choosing the loops, the total number of choices for the m_0 loops is bounded by

$$\sum_{s=m_0-1}^{\infty} \binom{s}{m_0 - 1} (2d)^{L_0}[p(d, r)]^{s+1} = (2d)^{L_0} \left[\frac{p(d,r)}{1 - p(d,r)}\right]^{m_0}.$$

[13] We denote by "loop" the path $X_{i_2+1}, \cdots, X_{j_2}$ where $|X_{i_2+1} - X_{i_2}| = 1$ and not $X_{i_2+1} = X_{j_2}$. The number of points in this loop is $j_2 - i_2$, which must be even.

[14] The symbol $P\{E\}$ means the probability of the event E. $P\{E|F\}$ will be used for the conditional probability of E, given F. The simple random walk is a Markov chain on the integral points. It starts at $S_0 = 0$ and moves at each step from a point to one of its $2d$ nearest neighbors, with probability $1/2d$ for each neighbor. The individual steps are independent.

This completes the proof, since the number of times w^{n-L_0} can be obtained by loop removal is at most equal to the number of v which can be reconstructed by inserting loops in w^{n-L_0}, which in turn is at most the number of possible selections for the places where the loops should be inserted times the number of ways of choosing the loops.

Lemma 2.

$$\chi_{n-L_0}(d) \geq \frac{\chi_{n,2r}(d)}{2(n+1)^2}\binom{n-L_0+1}{m_0}^{-1} \times (2d)^{-L_0}\left[\frac{1-p(d,r)}{p(d,r)}\right]^{m_0}. \quad (2.1)$$

Proof: By the choice of m_0 and L_0 there are at least $\chi_{n,2r}(d)/2(n+1)^2$ paths in $W_{n,2r}$, which, by removal of m_0 loops of total number of points L_0, go over in a self-avoiding walk of length $n - L_0$. But there are $\chi_{n-L_0}(d)$ w^{n-L_0} and each one is obtained at most

$$\binom{n-L_0+1}{m_0}(2d)^{L_0}\left[\frac{p(d,r)}{1-p(d,r)}\right]^{m_0}$$

times by Lemma 1. Thus

$$\frac{\chi_{n,2r}(d)}{2(n+1)^2} \leq \binom{n-L_0+1}{m_0} \times (2d)^{L_0}\left[\frac{p(d,r)}{1-p(d,r)}\right]^{m_0}\chi_{n-L_0}(d),$$

which is precisely the lemma.

Lemma 3. There exist constants C_1 and C_2 such that

$$0 \leq p(d,0) - \frac{1}{2d} \leq \frac{C_1}{d^2}, \quad d = 3, 4, \cdots, \quad (2.2)$$

$$p(d,r) \leq \frac{C_2}{d^{r+1}}\frac{(r+2)^{2r+4}}{(r+2)!}, \quad r = 1, 2, \cdots, d = 3, 4, \cdots. \quad (2.3)$$

Moreover, for suitable constants C_3 and C_4, $L_0(r, n)$ can be chosen in such a way that

$$\limsup_{n\to\infty}\frac{L_0(0,n)}{n} \leq \frac{C_3}{d} \quad \text{for} \quad d = 5, 6, \cdots \quad (2.4)$$

and

$$\limsup_{n\to\infty}\frac{L_0(r,n)}{n} \leq \frac{C_4 d!\,(d-8r+4)!\,(r+3)^{2r+10}}{d^{r+1}[(d-4r+2)!]^2(r+3)!} \quad (2.5)$$

for $r = 1, 2, \cdots, d \geq \max(8r-4, 4r+4)$.

This lemma is of a computational nature and we prefer to postpone its proof until the next section.

Theorem 1. There exists a constant C_0 such that

$$0 \leq \beta_{d,2r} - \beta_d \leq \frac{C_0}{d^r} \times \left[\frac{(r+2)^{2r+4}}{(r+2)!} + \frac{1}{d}\frac{d!\,(d-8r+4)!}{[(d-4r+2)!]^2}\frac{(r+3)^{2r+10}}{(r+3)!}\right]$$

for $r = 0, 1, \cdots$, and $d \geq \max(26r - 13, 5)$.

Proof: The left-hand inequality is obvious. To obtain the right-hand inequality one takes $(n - L_0)$th roots of both sides in (2.1) and applies Stirling's formula. There results the inequality

$$[\chi_{n-L_0}(d)]^{1/n-L_0} = \beta_d(1+\epsilon_1(n)) \geq [\chi_{n,2r}(d)]^{1/n(1+L_0/n-L_0)}2^{-1/n-L_0}(n+1)^{-2/n-L_0} \times \binom{n-L_0+1}{m_0}^{-1/n-L_0}(2d)^{-L_0/n-L_0} \times \left[\frac{1-p(d,r)}{p(d,r)}\right]^{m_0/n-L_0} = \beta_{d,2r}\left(\frac{\beta_{d,2r}}{2d}\right)^{-L_0/n-L_0} \times (1-x_n)\left[\frac{x_n}{1-x_n}\frac{1-p(d,r)}{p(d,r)}\right]^{x_n}(1+\epsilon_2(n)),$$

where

$$x_n = m_0/(n-L_0)$$

and $\epsilon_i(n) \to 0$ $(i = 1, 2)$ as $n \to \infty$ through a sequence for which $\limsup m_0/n \leq \limsup L_0/n < \frac{1}{2}$. It is a trivial minimization problem to show that

$$(1-x)[x/(1-x)(1-p)/p]^x \geq (1-p), \quad 0 \leq x \leq 1$$

[recall that $0 \leq p(d,r) < 1$] and, therefore, when we let $n \to \infty$ through a sequence for which

$$\frac{L_0(r,n)}{n - L_0(r,n)} \to A(d,r) = \liminf_{n\to\infty}\frac{L_0(r,n)}{n-L_0(r,n)}$$

and for which $A(d,r) < \frac{1}{2}$, we obtain

$$\beta_d \geq \beta_{d,2r} - \beta_{d,2r}\left[1 - \left(\frac{\beta_{d,2r}}{2d}\right)^{A(d,r)}(1-p(d,r))\right] \geq \beta_{d,2r} - 2d\left[1 - \left(\frac{\beta_{d,2r}}{2d}\right)^{A(d,r)}(1-p(d,r))\right]. \quad (2.6)$$

This is first applied for $r = 0$, which corresponds to a simple walk without conditions on the loops. Then $\beta_{d,0} = 2d$ and therefore, as soon as (2.2) and (2.4) are established, it follows from (2.6) that

$$\beta_d \geq 2d - 2dp(d,0) = 2d - 1 + O(1/d)$$

for sufficiently large d, say $d \geq d_0$, so that

$$\beta_{d,2r} \geq \beta_d \geq 2d - 2 \quad (d \geq d_0). \quad (2.7)$$

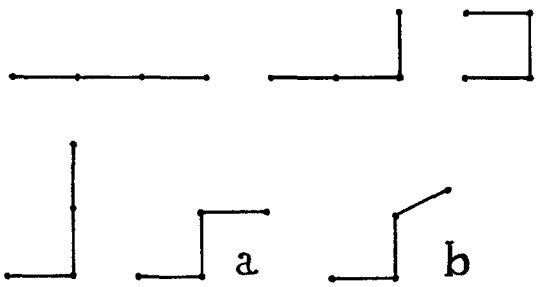


FIG. 1. The six types for three steps of a path in $W_{n,4}$. In (a), the three steps are in one plane and in (b) they are three dimensions.

[From previous estimates[3] one knows that (2.7) holds for $d = 2, 3$ and probably for all $d \geq 2$. Our estimates for C_1 and C_3 are not sharp enough to show this]. Substituting (2.7) in (2.6) one finally has for $d \geq d_0$

$$\beta_d \geq \beta_{d,2r} - 2d\left[1 - \left(\frac{2d-2}{2d}\right)^{A(d,r)}(1 - p(d,r))\right].$$

Together with (2.3) and (2.5) this implies the theorem as soon as $d \geq d_0$ and the right-hand side of (2.5) is less than $\frac{1}{2}$. It follows easily from Stirling's formula that the right-hand side of (2.5) will be less than $\frac{1}{2}$ for large d, $d \geq 26r - 13$. But then the theorem follows in general by taking C_0 sufficiently large.

Corollary 1. $\beta_d = 2d - 1 - 1/2d + O(1/d^2)$.

As stated in the introduction, Corollary 1 follows from

$$\beta_{d,4} = 2d - 1 - 1/2d + O(1/d^2). \tag{2.8}$$

We shall not actually prove (2.8) here, but do want to indicate how to find $M(d, 2)$ (cf. Sec. 1). Whether or not a path will have an immediate reversal or will close a loop of size four at a given step depends on the three preceding steps. The six possible types for three consecutive steps of a path in $W_{n,4}$ are shown in Fig. 1. Thus $W_{n,4}$ splits into six subclasses according to the type of the last three steps. It is easy to express the number of paths of a given type of size $(n + 1)$ as a linear combination of the number of paths of size n of each type. $M(d, 2)$ is the matrix corresponding to this set of linear recurrence relations of the first order and $\beta_{d,4}$ is its largest eigenvalue.[15] As described, $M(d, 2)$ would be a 6×6 matrix. Actually a closer look reveals that one does not need to divide the paths in six types but that certain types can be lumped together and a 3×3 matrix will suffice.

[15] An analogous example is treated in detail in R. Bellman, *Introduction to Matrix Analysis* (McGraw-Hill Book Company, Inc., New York, 1960), Chap. 16, Sec. 2–11.

3. PROOF OF LEMMA 3

It is well known[16] that $P\{S_n = 0\} = 0$ for odd n and thus by the definition of $p(d, r)$ (cf. Lemma 1), one has

$$p(d, r) \leq \sum_{n=r+1}^{\infty} P\{S_{2n} = 0\} \tag{3.1}$$

and

$$0 \leq p(d, 0) - P\{S_2 = 0\} = p(d, 0) - \frac{1}{2d} \leq \sum_{n=2}^{\infty} P\{S_{2n} = 0\}. \tag{3.2}$$

Moreover, for $Z = \{Z^{(1)}, \cdots, Z^{(d)}\}$,[16]

$$P\{S_{2n} = Z\} = \left(\frac{1}{2\pi}\right)^d \int_{-\pi}^{+\pi} e^{-i\theta\cdot Z}\phi^{2n}(\theta)\, d\theta \leq \left(\frac{1}{2\pi}\right)^d \int_{-\pi}^{+\pi} \phi^{2n}(\theta)\, d\theta = P\{S_{2n} = 0\} \tag{3.3}$$

where $\int_{-\pi}^{+\pi} d\theta$ stands for the d-fold integral

$$\int_{-\pi}^{+\pi} \cdots \int_{-\pi}^{+\pi} d\theta_1 \cdots d\theta_d, \qquad \theta\cdot Z = \sum_{j=1}^{d} \theta_j Z^{(j)},$$

and

$$\phi(\theta) = \frac{1}{d}\sum_{j=1}^{d} \cos\theta_j,$$

the characteristic function for one step of the simple random walk in d dimensions. It is easy to see from (3.3) that[16]

$$P\{S_{2n} = 0\} \geq P\{S_{2n+2} = 0\}. \tag{3.4}$$

Finally,[8]

$$P\{S_{2n} = 0\} \leq P(\text{first } 2n \text{ steps of random walk are confined to a } n\text{-dimensional subspace}) \tag{3.5}$$

$$\leq \binom{d}{n}\left(\frac{n}{d}\right)^{2n} \leq \frac{n^{2n}}{n!\, d^n},$$ [17]

since return to 0 at the $2n$th step requires that for each step in the jth direction there is an opposite step in the jth direction. By (3.4), (3.5), and (3.3) one has

$$\sum_{n=r+1}^{\infty} P\{S_{2n} = 0\} \leq P\{S_{2r+2} = 0\} + \sum_{n=r+2}^{d-1} + \sum_{n=d}^{\infty} P\{S_{2n} = 0\} \leq \frac{(r+1)^{2r+2}}{(r+1)!\, d^{r+1}} + \frac{d(r+2)^{2r+4}}{(r+2)!\, d^{r+2}} + \left(\frac{1}{2\pi}\right)^d \int_{-\pi}^{+\pi} \frac{\phi(\theta)^{2d}}{1 - \phi(\theta)}\, d\theta. \tag{3.6}$$

If $0 \leq \theta_{j'} \leq \pi/2 \leq \theta_{j''} < \pi$ for $j' = 1, \cdots, k$,

[16] G. Polya, Math. Ann. **84**, 149 (1921).

[17] The next-to-last member of (3.5) is correct only for $n \leq d$, but the last member can obviously be used for all $n \geq 0$.

$j'' = k+1, \cdots, d$, then

$$|\phi(\theta)| \le \frac{1}{d}\sum_{j=1}^{d}|\cos\theta_j| \le 1 - \frac{4}{\pi^2 d}\sum_{j=1}^{k}\theta_j^2 - \frac{4}{\pi^2 d}\sum_{j=k+1}^{d}(\pi-\theta_j)^2 \le \exp -\frac{4}{\pi^2 d}\left\{\sum_{j=1}^{k}\theta_j^2 + \sum_{j=k+1}^{d}(\pi-\theta_j)^2\right\}.$$

Using the evenness of $\cos\theta$ and the symmetry of $\phi(\theta)$ in its arguments, we may write[18]

$$\begin{aligned}
&\left|\left(\frac{1}{2\pi}\right)^d \int_{-\pi}^{+\pi} \frac{\phi(\theta)^{2d}}{1-\phi(\theta)}\,d\theta\right| \\
&\le \frac{1}{\pi^d}\sum_{k=0}^{d}\binom{d}{k}\int_0^{\pi/2}\cdots\int_0^{\pi/2} d\theta_1\cdots d\theta_k \\
&\quad\times\int_{\pi/2}^{\pi}\cdots\int_{\pi/2}^{\pi} d\theta_{k+1}\cdots d\theta_d \left|\frac{\phi(\theta)^{2d}}{1-\phi(\theta)}\right| \\
&\le \frac{1}{\pi^d}\sum_{k=0}^{d}\binom{d}{k}\int_0^{\pi/2}\cdots\int_0^{\pi/2} d\theta_1\cdots d\theta_k \\
&\quad\times\int_{\pi/2}^{\pi}\cdots\int_{\pi/2}^{\pi} d\theta_{k+1}\cdots d\theta_d \qquad (3.7) \\
&\times\left\{\frac{4}{\pi^2 d}\sum_{j=1}^{k}\theta_j^2 + \frac{4}{\pi^2 d}\sum_{j=k+1}^{d}(\pi-\theta_j)^2\right\}^{-1} \\
&\quad\times\exp -\frac{8}{\pi^2}\left\{\sum_{j=1}^{k}\theta_j^2 + \sum_{j=k+1}^{d}(\pi-\theta_j)^2\right\} \\
&\le \left(\frac{1}{\pi}\right)^d \frac{\pi^2 d}{4}\int_{-\infty}^{+\infty}\left\{\sum_{j=1}^{d}\theta_j^2\right\}^{-1}\exp\left\{-\frac{8}{\pi^2}\sum_{j=1}^{d}\theta_j^2\right\}d\theta \\
&= 2d\left(\frac{\pi}{8}\right)^{\frac{1}{2}d}\frac{\Gamma[(d-2)/2]}{\Gamma(\frac{1}{2}d)} = 4\,\frac{d}{d-2}\left(\frac{\pi}{8}\right)^{\frac{1}{2}d}.
\end{aligned}$$

Combining (3.7) with (3.6), (3.1), and (3.2), one easily derives (2.2) and (2.3).

For the proof of (2.4) we need the following estimate, which is proved in exactly the same way as (3.7):

$$\begin{aligned}
&\sum_{n=d}^{\infty} 2nP\{S_{2n}=0\} \\
&= \left(\frac{1}{2\pi}\right)^d\int_{-\pi}^{+\pi}\frac{2d\phi(\theta)^{2d} - (2d-1)\phi(\theta)^{2d+1}}{[1-\phi(\theta)]^2}\,d\theta \\
&\le \left(\frac{1}{\pi}\right)^d\frac{\pi^4 d^3}{4}\int_{-\infty}^{+\infty}\left\{\sum_{j=1}^{d}\theta_j^2\right\}^{-2} \\
&\quad\times\exp\left\{-\frac{8}{\pi^2}\sum_{j=1}^{d}\theta_j^2\right\}d\theta \\
&= 64\,\frac{d^3}{(d-2)(d-4)}\left(\frac{\pi}{8}\right)^{\frac{1}{2}d}, \quad d = 5, 6, \cdots. \qquad (3.8)
\end{aligned}$$

[18] R. Courant, *Differential and Integral Calculus* (Interscience Publishers, Inc., New York), Vol. II, 2d ed., Appendix to Ch. IV, Sec. 3.3.

One proves (2.4) and (2.5) by means of Markov's inequality,[19] namely, if

$$\frac{\sum_{v\in W_{n,2r}} L(v)}{\sum_{v\in W_{n,2r}} 1} = \frac{\sum_{v\in W_{n,2r}} L(v)}{\chi_{n,2r}} \le \tfrac{1}{2}C(d,r), \qquad (3.9)$$

then

$$\sum_{\substack{v\in W_{n,2r}\\ L(v)\ge C(d,r)}} 1 \le \tfrac{1}{2}\chi_{n,2r},$$

and hence the number of $v \in W_{n,2r}$ with $L(v) < C(d, r)$ is at least $\frac{1}{2}\chi_{n,2r}$, and thus there exists an $L_0 = L_0(r, n) < C(d, r)$ such that the number of $v \in W_{n,2r}$ with $L(v) = L_0$ is at least $\chi_{n,2r}/2(n+1)$. For this L_0 there exists an m_0 such that the number of $v \in W_{n,2r}$ with $L(v) = L_0$, $m(v) = m_0$ is at least $\chi_{n,2r}/2(n+1)^2$, as required. For the second part of Lemma 3 it therefore suffices to prove (3.9) with n times the right-hand side of (2.4) substituted for $C(d, 0)$, respectively, n times the right-hand side of (2.5) for $C(d, r)$. It is useful to realize that the left-hand side of (3.9) is the expectation of $L(v)$ if v is uniformly distributed over $W_{n,2r}$. But this distribution is the same as the conditional distribution for a simple random-walk path of n steps, given that it has no loops of size less than $2r+1$. Denoting this conditional probability distribution by P_r and its corresponding expectation by E_r, we have

$$\begin{aligned}
\sum_{v\in W_{n,2r}} L(v)/\chi_{n,2r} &= E_r\{L(v)\} \\
&\le E_r\ \{\text{number of } k \text{ for which there exist } i, j, i < k \le j, \text{ such that } X_i(v) = X_j(v)\} \\
&= \textstyle\sum_{k=0}^{n} P_r\ \{\text{there exist } i, j, i < k \le j \text{ such that } X_i(v) = X_j(v)\}. \qquad (3.10)
\end{aligned}$$

The inequality follows from the fact that a point $X_k(v)$ is removed with one of the loops from v only if there exists at least one $i < k$ and one $j \ge k$ such that $X_i(v) = X_j(v)$. Of course if there are no loops of fewer than $2r+1$ steps, such i and j must satisfy $|j-i| \ge 2r+2$, and consequently

$$\begin{aligned}
&\sum_{k=0}^{n} P_r\{\text{there exist } i, j, i < k \le j, \text{ such that } X_i(v) = X_j(v)\} \\
&\le \sum_{k=0}^{n}\sum_{s=r+1}^{\infty}\sum_{i=k-2s}^{k-1} P_r\{X_i(v) = X_{i+2s}(v)\} \\
&\le \sum_{i=0}^{n-1}\sum_{s=r+1}^{\infty} 2sP_r\{X_i(v) = X_{i+2s}(v)\}. \qquad (3.11)
\end{aligned}$$

[19] A. Renyi, *Wahrscheinlichkeitsrechnung* (VEB Deutscher Verlag der Wissenschaften, Berlin, 1962), p. 183.

The easiest case is of course $r = 0$, for then there are no conditions on loops and P_0 is the probability for simple random walk so that [cf. (3.4), (3.5), (3.8)]

$$\sum_{s=1}^{\infty} 2sP_0\{X_i(v) = X_{i+2s}(v)\} = \sum_{s=1}^{\infty} 2sP\{S_{2s} = 0\}$$

$$\leq 2P\{S_2 = 0\} + 4P\{S_4 = 0\} + d^2P\{S_6 = 0\} + \sum_{s=d}^{\infty} 2sP\{S_{2s} = 0\}$$

$$\leq \frac{1}{d} + 4\frac{2^4}{2!\,d^2} + d^2\frac{3^6}{3!\,d^3} + 64\frac{d^3}{(d-2)(d-4)}\left(\frac{\pi}{8}\right)^{\frac{1}{2}d} = O\left(\frac{1}{d}\right).$$

Thus for $r = 0$, (2.4) holds and, as observed in the proof of Theorem 1, this together with (2.2) implies (2.7) for $d \geq d_0$, a fact which we need in order to prove (2.5). For general r we split $v \in W_{n,2r}$ with $X_i(v) = X_{i+2s}(v)$ in three pieces: its first $i + 2(s - t)$ steps, its last $n - i - 2s$ steps and a middle piece of $2t$ steps ($t \leq s$). These pieces are paths belonging to $W_{i+2(s-t),2r}$, $W_{n-i-2s,2r}$, and $W_{2t,2r}$, respectively. If $X_i(v)$ is to equal $X_{i+2s}(v)$, the middle piece has to satisfy certain conditions, once the first piece is given. In fact when the first part is given, the end point of the middle piece is fixed and there are at most

$$\sup_Z \{\text{number of } v \in W_{2t,2r} \text{ which end in } Z\} \leq \sup_Z \{\text{number of } v \in W_{2t,0} \text{ which end in } Z\} = (2d)^{2t} \sup_Z P\{S_{2t} = Z\} \leq (2d)^{2t} P\{S_{2t} = 0\} \text{ [cf. (3.3)]} \tag{3.12}$$

possible choices for the middle piece. Consequently,

$$P_r\{X_i(v) = X_{i+2s}(v)\} = \chi_{n,2r}^{-1} \times \{\text{number of } v \in W_{n,2r} \text{ with } X_i(v) = X_{i+2s}(v)\} \leq \chi_{n,2r}^{-1}\,\chi_{i+2(s-t),2r}\,\chi_{n-i-2s,2r} \times \sup_Z \{\text{number of } v \in W_{2t,2r} \text{ which end in } Z\}. \tag{3.13}$$

On the other hand, let P be the probability that three paths, v_1, v_2, v_3, chosen at random from $W_{i+2(s-t),2r}$, $W_{2t,2r}$, $W_{n-i-2s,2r}$, can be juxtaposed (by translating the initial point of v_2 so that it coincides with the endpoint of v_1 and similarly translating v_3) to form a path in $W_{n,2r}$. Then,

$$\chi_{n,2r} \geq \chi_{i+2(s-t),2r} \cdot \chi_{2t,2r} \cdot \chi_{n-i-2s,2r} \cdot P. \tag{3.14}$$

For $d \geq 8r - 4$ and $t \geq r$ one can easily give a lower bound for P. In fact the last $(2r - 1)$ steps of v_1 and the first $(2r - 1)$ steps of v_3 are together contained in a $(4r - 2)$-dimensional subspace, and the same holds for the first and last $(2r - 1)$ steps of v_2 together. If these two subspaces are disjoint, i.e., if the first and last $(2r - 1)$ steps of v_2 are all in directions different from the last $(2r - 1)$ steps of v_1 and first $(2r - 1)$ steps of v_3, then there can be no loops of $2r$ or fewer steps with an initial point in $v_1(v_2)$ and end point in $v_2(v_3)$. Since v_1, v_2, v_3 themselves have no loops of $2r$ or fewer steps and $t \geq r$, the path constructed by juxtaposing v_1, v_2, v_3 will indeed belong to $W_{n,2r}$ if the above subspaces are disjoint. If the first $(2r - 1)$ and last $(2r - 1)$ steps of v_2 together are along exactly $k \leq 4r - 2$ coordinates axes, then, since all possible k tuples of k directions are equally likely, the probability that these k directions are different from all the directions of the last $(2r - 1)$ steps of v_1 and first $(2r - 1)$ steps of v_3 is at least

$$\binom{d - 4r + 2}{k} \Big/ \binom{d}{k} \geq \frac{(d - 4r + 2)!\,(d - 4r + 2)!}{d!\,(d - 8r + 4)!} = P(d, r), \tag{3.15}$$

say. It is therefore also true that

$$P \geq P(d, r), \tag{3.16}$$

and from (3.12)–(3.14) and (2.7),[20] it now follows that

$$P_r\{X_i(v) = X_{i+2s}(v)\} \leq [P(d, r)\chi_{2t,2r}]^{-1}, \tag{3.17}$$

$$(2d)^{2t}P\{S_{2t} = 0\} \leq [P(d, r)]^{-1}\left(\frac{2d}{2d - 2}\right)^{2t} P\{S_{2t} = 0\}.$$

We apply (3.17) with $t = s$ for $s \leq [e^{-2}d]$,[21,22] and $t = [e^{-2}d]$ for $[e^{-2}d] \leq s \leq 4rd^2 - 1$.

From (3.4), (3.5), (3.17), and Stirling's formula, one concludes, with suitable constants C_5, C_6,

$$\sum_{s=r+1}^{4rd^2-1} 2sP_r\{X_i(v) = X_{i+2s}(v)\} \leq [P(d, r)]^{-1}\left(\frac{2d}{2d - 2}\right)^{2d/e^2} [(2r + 2)P\{S_{2r+2} = 0\} + (2r + 4)P\{S_{2r+4} = 0\} + e^{-4}d^2P\{S_{2r+6} = 0\} + 4r^2d^4P\{S_{2[e^{-2}d]} = 0\}] \leq C_5[P(d, r)]^{-1}$$

[20] We, also, use the inequality $\chi_{n,2r} \geq (\beta_{d,2r})^n$ which is proved in exactly the same manner as the inequality $\chi_n \geq (\beta_d)^n$ in Ref. 11 [left-hand inequality of (3)].

[21] In two places we use $[a]$ to denote the largest integer not exceeding a. In this meaning it occurs only as $[e^{-2}\,d]$ and as $[s/r]$. Even though we use square brackets in other meanings, the risk of confusion is small.

[22] The cutoff point is chosen at $[e^{-2}\,d]$ because the estimate $n^{2n}/n!\,d^n$ roughly takes its minimum for $n = e^{-2}\,d$.

$$\times \left\{ \frac{(r+3)^{2r+6}}{(r+3)!\, d^{r+1}} + \frac{r^2 d^4}{[e^{-2}d]!} \left(\frac{[e^{-2}d]^2}{d} \right)^{[e^{-2}d]} \right\}$$

$$\leq C_6 [P(d,r)]^{-1} \frac{(r+3)^{2r+10}}{(r+3)!\, d^{r+1}}. \tag{3.18}$$

Unfortunately these estimates are not sharp enough in the region $s \geq 4rd^2$. We have to go back to (3.13), and (3.14) which we now use (with $t = s$) in the form

$$P_r\{X_i(v) = X_{i+2s}(v)\} \leq [P(d,r)]^{-1} \chi_{2s,2r}^{-1}$$
$$\times \sup_Z \{\text{number of } v \in W_{2s,2r} \text{ which end in } Z\}$$
$$= [P(d,r)]^{-1} \sup_Z P\{v \text{ ends in } Z \mid v \in W_{2s,2r}\}.^{14} \tag{3.19}$$

Of course, if one replaces $W_{2s,2r}$ by $W_{2s,0}$, then[16] $\sup_Z P\{v \text{ ends in } Z \mid v \in W_{2s,0}\} = \sup_Z P\{S_{2s} = Z\} \leq C_7 (2s)^{-\frac{1}{2}d}$. At present we can not draw this conclusion because under the condition $v \in W_{2s,2r}$ the steps of v are not independent. However, they still have enough independence properties for our purposes. More precisely, we use the fact that even under the condition $v \in W_{2s,2r}$ and even if $X_p(v)$, $p < i$, and $X_q(v) - X_{i+1}(v)$ $q \geq i+1$, are fixed, then there are still at least $2(d - 4r + 2)$ possible choices for $X_{i+1}(v) - X_i(v)$, and all possibilities are equally likely under the given conditions. The argument is the same as in the derivation of (3.15), (3.16). If the steps $X_t(v) - X_{t-1}(v)$ $t = i - 2r + 2$, $i - 2r + 3, \cdots, i$ and $t = i+2, i+3, \cdots, i + 2r$ are along the coordinate axes $l_1, \cdots l_m$, where necessarily $m \leq 4r - 2$, then $X_{i+1}(v) - X_i(v)$ can be any step along an axis $l \notin \{l_1, \cdots, l_m\}$ for then there is only one step along the axis l between $X_{i-2r+1}(v)$ and $X_{i+2r}(v)$, and hence v can have no loop which contains the step $X_{i+1}(v) - X_i(v)$ and has $2r$ or fewer steps in total. This argument is easily generalized for a number of steps at least $2r - 1$ steps removed from each other. That is, if v has to belong to $W_{2s,2r}$ and $X_j(v) - X_{j-1}(v)$ is already given for all j not of the form $2rp$, $p = 1, 2, \cdots$, then there are at least $2(d - 4r + 2)$ possible choices for each one of the steps $X_{2rp}(v) - X_{2rp-1}(v)$, $p = 1, 2, \cdots, [s/r]$,[21] and these choices can be made independent of each other. Let us assume that under the given conditions the possible choices for $X_{2rp}(v) - X_{2rp-1}(v)$ are $(0, \cdots, 0, \pm 1, 0, \cdots, 0)$ with $+1$ or -1 in one of the places $k_1, k_2, \cdots, k_g$, $g \geq d - 4r + 2$, and only a $+1$ or only a -1 in the places $k'_1, \cdots, k'_h$. Then $\phi_p(\theta)$, the conditional characteristic function of $X_{2rp}(v) - X_{2rp-1}(v)$ given these conditions, is

$$\phi_p(\theta) = \frac{1}{2g+h} \sum_{q=1}^{h} \exp(\eta_q i \theta_{k'_q}) + \frac{2}{2g+h} \sum_{q=1}^{g} \cos \theta_{k_q},$$

where η_q is either $+1$ or -1. Since $2g + h \leq 2d$, this implies

$$|\phi_p(\theta)| \leq 1 - \frac{1}{d} \sum_{q=1}^{g} (1 - |\cos \theta_{k_q}|)$$
$$\leq \exp - \left\{ \frac{4}{\pi^2 d} \sum_{q=1}^{g} \tilde{\theta}_{k_q}^2 \right\} \qquad |\theta_{k_q}| \leq \pi, \tag{3.20}$$

where $\tilde{\theta} = \min\{\theta^2, (\pi - |\theta|)^2\}$.
Moreover, because the $X_{2rp}(v) - X_{2rp-1}(v)$, $p = 1, 2, \cdots, [s/r]$, are independent when $X_j(v) - X_{j-1}(v)$ is given for $j \neq 2rp$, it follows that

$$\sup_Z P\{v \text{ ends in } Z \mid v \in W_{2s,2r},$$
$$X_j(v) - X_{j-1}(v),\ j \neq 2rp\}$$
$$= \sup_Z P\Big\{ \sum_{p=1}^{[s/r]} X_{2rp}(v) - X_{2rp-1}(v)$$
$$= Z - \sum_{j \neq 2rp} (X_j(v) - X_{j-1}(v)) \tag{3.21}$$
$$v \in W_{2s,2r},\ X_j(v) - X_{j-1}(v),\ j \neq 2rp \Big\}$$
$$= \sup_Z P\Big\{ \sum_{p=1}^{[s/r]} X_{2rp}(v) - X_{2rp-1}(v) = Z \mid$$
$$v \in W_{2s,2r},\ X_j(v) - X_{j-1}(v),\ j \neq 2rp \Big\}$$
$$\leq \frac{1}{(2\pi)^d} \int_{-\pi}^{+\pi} \prod_{p=1}^{[s/r]} |\phi_p(\theta)|\, d\theta. \tag{3.21}$$

From (3.20) one sees that

$$\frac{1}{(2\pi)^d} \int_{-\pi}^{+\pi} \prod_{p=1}^{[s/r]} |\phi_p(\theta)|\, d\theta \leq \frac{1}{(2\pi)^d}$$
$$\times \int_{-\pi}^{+\pi} \exp - \left\{ \frac{4}{\pi^2 d} \sum_{j=1}^{d} n_j \tilde{\theta}_j^2 \right\} d\theta \leq \prod_{\substack{j=1 \\ n_j \neq 0}}^{d} \left(\frac{\pi d}{4 n_j} \right)^{\frac{1}{2}} \tag{3.22}$$

where n_j is the number of p for which $\tilde{\theta}_j$ appears in the sum $\sum_{q=1}^{g} \tilde{\theta}_{k_q}^2$ in the right-hand side of (3.20). Since $g \geq d - 4r + 2$, for each p one must have

$$\sum_{j=1}^{d} n_j \geq (d - 4r + 2) \left[\frac{s}{r} \right] \quad \text{and} \quad n_j \leq \left[\frac{s}{r} \right],$$

which easily implies

$$\prod_{\substack{j=1 \\ n_j \neq 0}}^{d} n_j \geq \left[\frac{s}{r} \right]^{d-4r+1} \tag{3.23}$$

as soon as $[s/r] \geq d$. For $s \geq rd$ one finally obtains from (3.21)–(3.23)

$$\sup_z P\{v \text{ ends in } Z \mid v \in W_{2s,2r},\ X_j(v) - X_{j-1}(v),\ j \neq 2rp\} \leq \left(\frac{\pi d}{3}\right)^{\frac{1}{2}d}\left(\frac{r}{s}\right)^{\frac{1}{2}(d-4r+1)}.$$

This holds for every permissible combination of $X_j(v) - X_{j-1}(v)$, $j \neq 2rp$ and therefore also

$$\sup_z P\{v \text{ ends in } Z \mid v \in W_{2s,2r}\} \leq \left(\frac{\pi d}{3}\right)^{\frac{1}{2}d}\left(\frac{r}{s}\right)^{\frac{1}{2}(d-4r+1)}. \tag{3.24}$$

When one combines this with (3.19) one sees that

$$\sum_{s=4rd^2}^{\infty} 2sP_r\{X_i(v) = X_{i+2s}(v)\} \leq C_8[P(d,r)]^{-1}\left(\frac{\pi d}{3}\right)^{\frac{1}{2}d} r^2(2d)^{2-d+4r} \leq C_9[P(d,r)]^{-1}\frac{(r+3)^{2r+10}}{(r+3)!\,d^{r+1}} \tag{3.25}$$

for certain constants C_8 and C_9, and $d \geq \max(8r-4, 4r+4)$, $r = 1, 2, \cdots$.

(2.5) follows from (3.10), (3.11), (3.18), (3.25), and (3.15).

4. BOUNDS FOR χ_n AND γ_n

We indicate here how (1.9) can be proved. As pointed out in the introduction we do not want to give the proof in all its details because we hope that Theorem 2 will be improved at some time. For the same reason we do not want to give the proof of (1.8) which is a refinement of the proof of (1.9) and the ideas in Ref. 6. Actually we prove

$$\chi_n^2 \leq \exp(\alpha_8 n^{2/(d+1)} \log n) \sum_{\lambda=0}^{n+dn^{2/(d+1)}} \gamma_{2\lambda-1}, \tag{4.1}$$

for some constant $\alpha_8 = \alpha_8(d)$. Since the left-hand inequality of (1.8) is known,[11] and since

$$2d(\lambda_2+1)\gamma_{\lambda_1+\lambda_2+1} \geq \gamma_{\lambda_1}\gamma_{\lambda_2} \quad [(3.2), (3.3) \text{ of Ref. 1}], \tag{4.2}$$

this implies the left hand inequality of (1.9) [with $n + dn^{2/(d+1)}$ taking the place of n]. But in view of the right-hand inequality of (1.9), which is also known,[12,1] (4.1) also implies (1.8) with $n^{2/(d+2)}$ in the exponent replaced by $n^{2/(d+1)}$.

Since there are at most $(2n+1)^d$ possibilities for the end point of w^n, we can find a $Z = Z(n)$ such that the number of w^n with $X_n(w^n) = Z$ is at least $(2n+1)^{-d}\chi_n(d)$. Consider two such self-avoiding walks w_i with $X_n(w_i) = Z$ $i = 1, 2$, and choose a Y such that $Y + X_j(w_2) = X_k(w_1)$ for at most n^β pairs (j, k), $0 \leq j, k \leq n$ (β will be determined later). Since there are only $(n+1)^2$ pairs (j, k), it is not hard to see that there are at most $(n+1)^2 n^{-\beta}$ points Y' for which $Y' + X_j(w_2) = X_k(w_1)$ has more than n^β solutions. Consequently we can choose Y such that its ith coordinate satisfies $|Y^{(i)}| \leq n^{(2-\beta)/d}$ for $i = 1, \cdots, d$. (There are $[2n^{(2-\beta)/d}+1]^d > (n+1)^2 n^{-\beta}$ such points.) There exists a self-avoiding path $u = \{X_i(u), 0 \leq i < m\}$ of length $m = \sum_{i=1}^d |Y^{(i)}| < dn^{(2-\beta)/d}$ from 0 to Y. Form a path v which traverses first w_1, then u, then w_2 in the reverse, and finally u in the reverse, where each piece is translated so that its initial point coincides with the last point of the preceding piece in v. Formally v consists of the following points in the given order: $X_i(w_1)$, $0 \leq i \leq n$, $X_n(w_1) + X_i(u)$, $1 \leq i \leq m$, $X_n(w_1) + X_m(u) + X_i(w_2) - X_n(w_2)$, $i = n-1, n-2, \cdots, 0$, $X_n(w_1) + X_m(u) - X_n(w_2) + X_i(u) - X_m(u)$, $i = m-1, m-2, \cdots, 1$. v has length $2n + 2m - 1 \leq 2n + 2dn^{(2-\beta)/d} - 1$, and since $X_n(w_1) = X_n(w_2) = Z$, its last point is $X_1(u)$ with $|X_1(u)| = 1$ so that v is a polygon. It is not necessarily self-avoiding, but by the choice of Y we know that there are at most n^β pairs (j, k) for which $X_k(w_1) = X_n(w_1) + X_m(u) + X_j(w_2) - X_n(w_2) = Y + X_j(w_2)$. Any further double points of v must occur in the piece $X_n(w_1) + X_i(u)$, $1 \leq i \leq m$, or in the piece $X_n(w_1) + X_m(u) - X_n(w_2) + X_i(u) - X_m(u) = X_i(u)$, $i = m-1, m-2, \cdots, 1$, and one can show that if v has r double points, then $r \leq n^\beta + 5m \leq n^\beta + 5dn^{(2-\beta)/d}$ [a point which is visited s times counts as $(s-1)$ double points]. Since there are at least $(2n+1)^{-2d}\chi_n^2(d)$ choices for the pair w_1, w_2 it follows that

$$\chi_n^2/(2n+1)^{2d} \leq \text{number of polygons with length at most } 2n + 2dn^{(2-\beta)/d} - 1 \text{ and at most } n^\beta + 5dn^{(2-\beta)/d} \text{ double points.} \tag{4.3}$$

On the other hand, each v with r double points can be constructed by attaching successively r self-avoiding polygons to a self-avoiding polygon. If we want to attach a polygon of length λ' to a path y of length λ, then this cay be done in at most $(\lambda+1)\gamma_{\lambda'}$ ways. The factor $(\lambda+1)$ enters for the number of choices of the place of attachment and the factor $\gamma_{\lambda'}$ for the number of choices of the polygon (compare the argument in the proof of Lemma 1). If y varies through a class of ρ members one obtains at most $\rho(\lambda+1)\gamma_{\lambda'}$ paths (of length $\lambda + \lambda' + 1$) in this way. It is not hard to show by repeating this argument that the number of polygons with length at most $2n + 2dn^{(2-\beta)/d} - 1$ and at most $n^\beta + 5dn^{(2-\beta)/d}$ double points does not exceed

$$\sum_{r=0}^{n^{\beta}+5dn^{(2-\beta)/d}} \sum_{r+\lambda_0+\cdots+\lambda_r\leq 2n+2dn^{(2-\beta)/d}-1} (2n+2dn^{(2-\beta)/d})^r \prod_{j=0}^{r} \gamma_{\lambda_j}. \tag{4.4}$$

Using (4.2) one shows that (4.4) is at most

$$[2d(2n+2dn^{(2-\beta)/d})^3]^{n^{\beta}+5dn^{(2-\beta)/d}+1} \sum_{\lambda=0}^{n+dn^{(2-\beta)/d}} \gamma_{2\lambda-1}.$$

Together with (4.3) and (4.4) this proves (4.1) when one takes $\beta = 2/(d+1)$ (which gives the best bound).

We end with a proof of Theorem 3 which is merely a combination of two facts proved in Ref. 1. Firstly,

$$\gamma_{4n+1} \geq (2n+1)^{-6d-2}\nu_n^4, \tag{4.5}$$

and secondly,

$$\exp 2\sum_{n=0}^{\infty}\frac{\nu_n}{\beta^n}x^n \geq \sum_{n=0}^{\infty}\frac{\chi_n}{\beta^n}x^{n+1} \geq \frac{x}{1-x}, \qquad 0 \leq x \leq 1 \tag{4.6}$$

[See formula (3.7) and p. 966 in Ref. 1. ν_n is defined on p. 963 there]. Notice that we have $(\chi_n/\beta^n)x^{n+1}$ as summand rather than $(\chi_n/\beta^n)x^n$, as was falsely given in Ref. 1.

Thus

$$\sum_{n=0}^{\infty}\frac{\nu_n}{\beta^n}x^n \geq -\tfrac{1}{2}\log(1-x), \qquad x \uparrow 1,$$

and

$$\nu_n \geq \beta^n \cdot \frac{1}{4n}$$

for infinitely many n. Together with (4.5) this proves Theorem 3.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 8 AUGUST 1964

Product Property and Cluster Property Equivalence*

S. SHERMAN
Department of Mathematics, Wayne State University, Detroit, Michigan
(Received 17 March 1964)

Uhlenbeck and Ford in their presentation of the Ursell development express a sequence of symmetric functions W_N in terms of another sequence of symmetric functions U_N. They invert this sequence of equations and remark that from the first set of equations it follows that the product property of W_N sequence is equivalent to the cluster property for the U_N sequence. They ask for a simple proof of the equivalence. Here a modification of an algebra developed by Bohnenblust to prove Spitzer's formula on the fluctuations of the sums of independent, identically distributed, random variables is used systematically to invert the relations and to prove the equivalence under lighter assumptions than previously used.

UHLENBECK AND FORD (Ref. 1) consider a sequence of symmetric functions $W_1(r_1)$, $W_2(r_1, r_2)$, $W_3(r_1, r_2, r_3)$, $\cdots$ and another sequence of symmetric functions

$$U_1(r_1), \quad U_2(r_1, r_2), \quad U_3(r_1, r_2, r_3), \cdots$$

such that, in their language,

$$\begin{aligned} &\text{“}W_1(r_1) = U_1(r_1) \equiv 1, \\ &W_2(r_1, r_2) = U_2(r_1, r_2) + U_1(r_1)U_1(r_2), \\ &W_3(r_1, r_2, r_3) = U_3(r_1, r_2, r_3) + U_2(r_1, r_2)U_1(r_3) \\ &\quad + U_2(r_1, r_3)U_1(r_2) + U_2(r_2, r_3)U_1(r_1) \\ &\quad + U_1(r_2)U_1(r_3), U_1(r_1), \end{aligned} \tag{28}$$

and so on. The general rule is as follows. Divide the N particles which occur in W_N in a number of groups, and form the product of the functions U_M which depend on the particles in each group. Then W_N will be the sum of these products for all possible

* Research supported by the National Science Foundation, Grant No. G24334.

[1] Pp. 136–7 of G. E. Uhlenbeck and G. W. Ford, The Theory of Linear Graphs with Applications to the Theory of the Virial Development of the Properties of Gases, in *Studies in Statistical Mechanics*, edited by J. deBoer and G. E. Uhlenbeck (North-Holland Publishing Company, Amsterdam, 1962), Vol. I, referred to as UF.

$$\sum_{r=0}^{n^{\beta}+5dn^{(2-\beta)/d}} \sum_{r+\lambda_0+\cdots+\lambda_r\leq 2n+2dn^{(2-\beta)/d}-1} (2n+2dn^{(2-\beta)/d})^r \prod_{j=0}^{r} \gamma_{\lambda_j}. \tag{4.4}$$

Using (4.2) one shows that (4.4) is at most

$$[2d(2n+2dn^{(2-\beta)/d})^3]^{n^{\beta}+5dn^{(2-\beta)/d}+1} \sum_{\lambda=0}^{n+dn^{(2-\beta)/d}} \gamma_{2\lambda-1}.$$

Together with (4.3) and (4.4) this proves (4.1) when one takes $\beta = 2/(d+1)$ (which gives the best bound).

We end with a proof of Theorem 3 which is merely a combination of two facts proved in Ref. 1. Firstly,

$$\gamma_{4n+1} \geq (2n+1)^{-6d-2}\nu_n^4, \tag{4.5}$$

and secondly,

$$\exp 2\sum_{n=0}^{\infty}\frac{\nu_n}{\beta^n}x^n \geq \sum_{n=0}^{\infty}\frac{\chi_n}{\beta^n}x^{n+1} \geq \frac{x}{1-x}, \qquad 0 \leq x \leq 1 \tag{4.6}$$

[See formula (3.7) and p. 966 in Ref. 1. ν_n is defined on p. 963 there]. Notice that we have $(\chi_n/\beta^n)x^{n+1}$ as summand rather than $(\chi_n/\beta^n)x^n$, as was falsely given in Ref. 1.

Thus

$$\sum_{n=0}^{\infty}\frac{\nu_n}{\beta^n}x^n \geq -\tfrac{1}{2}\log(1-x), \qquad x \uparrow 1,$$

and

$$\nu_n \geq \beta^n \cdot \frac{1}{4n}$$

for infinitely many n. Together with (4.5) this proves Theorem 3.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 8 AUGUST 1964

Product Property and Cluster Property Equivalence*

S. SHERMAN
Department of Mathematics, Wayne State University, Detroit, Michigan
(Received 17 March 1964)

Uhlenbeck and Ford in their presentation of the Ursell development express a sequence of symmetric functions W_N in terms of another sequence of symmetric functions U_N. They invert this sequence of equations and remark that from the first set of equations it follows that the product property of W_N sequence is equivalent to the cluster property for the U_N sequence. They ask for a simple proof of the equivalence. Here a modification of an algebra developed by Bohnenblust to prove Spitzer's formula on the fluctuations of the sums of independent, identically distributed, random variables is used systematically to invert the relations and to prove the equivalence under lighter assumptions than previously used.

UHLENBECK AND FORD (Ref. 1) consider a sequence of symmetric functions $W_1(r_1)$, $W_2(r_1, r_2)$, $W_3(r_1, r_2, r_3)$, $\cdots$ and another sequence of symmetric functions

$$U_1(r_1), \quad U_2(r_1, r_2), \quad U_3(r_1, r_2, r_3), \cdots$$

such that, in their language,

$$\begin{aligned} &\text{“}W_1(r_1) = U_1(r_1) \equiv 1, \\ &W_2(r_1, r_2) = U_2(r_1, r_2) + U_1(r_1)U_1(r_2), \\ &W_3(r_1, r_2, r_3) = U_3(r_1, r_2, r_3) + U_2(r_1, r_2)U_1(r_3) \\ &\quad + U_2(r_1, r_3)U_1(r_2) + U_2(r_2, r_3)U_1(r_1) \\ &\quad + U_1(r_2)U_1(r_3), U_1(r_1), \end{aligned} \tag{28}$$

and so on. The general rule is as follows. Divide the N particles which occur in W_N in a number of groups, and form the product of the functions U_M which depend on the particles in each group. Then W_N will be the sum of these products for all possible

* Research supported by the National Science Foundation, Grant No. G24334.

[1] Pp. 136–7 of G. E. Uhlenbeck and G. W. Ford, The Theory of Linear Graphs with Applications to the Theory of the Virial Development of the Properties of Gases, in *Studies in Statistical Mechanics*, edited by J. deBoer and G. E. Uhlenbeck (North-Holland Publishing Company, Amsterdam, 1962), Vol. I, referred to as UF.

ways of division of the N particles. The U_M can also be expressed in terms of the W_N by inverting the relations given in (28). One finds

$$U_1(r_1) = W_1(r_1),$$
$$U_2(r_1, r_2) = W_2(r_1, r_2) - W_1(r_1)W_1(r_2),$$
$$U_3(r_1, r_2, r_3) = W_3(r_1, r_2, r_3) - W_2(r_1, r_2)W_1(r_3) - W_2(r_1, r_3)W_1(r_2) - W_2(r_2, r_3)W_1(r_1) + 2W_1(r_1)W_1(r_2)W_1(r_3), \quad (29)$$

and so on. The general rule is the same as that for expressing the W_N in the U_M except for a coefficient $(-)^{k-1}(k-1)!$, where k is the number of groups into which the M particles in U_M are divided.

The significance of the relations (28) and (29) lies in the following fundamental properties of the functions W_N and U_M. When we divide the particles into a number of groups containing $\alpha_1, \alpha_2, \cdots$, particles, then for configurations in which particles of different groups are so far away from each other that their interaction vanishes, we have:

$W_N \to W_{\alpha_1} W_{\alpha_2} \cdots$ (product property),
$U_M \to 0$ (cluster property).

This follows immediately from the definitions in (26) and (27). However, one can prove in general that if the W_N and U_M are related by (28) and (29) then the product property of the W_N implies the cluster property of the U_M, and vice versa. For a proof by induction of the first statement see B. Kahn and G. E. Uhlenbeck, Physica 5 (1938) 399. A simple direct proof of either statement is lacking." (which ends the quotation from UF).

While the cited proof in Kahn and Uhlenbeck[2] is simple enough and the reverse implication can be supplied in a few additional lines, there is a modification of an algebra, developed by Bohnenblust[3] to establish Spitzer's formula[4] on the fluctuations of sums of independent, identically distributed random variables, which can also be used systematically to invert (28) and to establish the equivalence between the product property for the W_N sequence and the cluster property for the U_N sequence. This approach is presented here deliberately ignoring the fact that the W_N and U_N are integrands of certain multiple integrals. Comments suggested by this fact are presented at the end.

Let $\Omega = \{1, 2, \cdots\}$. Let $A, B, \cdots, S, T, \cdots$ be finite subsets of Ω. With each such subset A associate symbol $z^{[A]}$. Define

$$z^{[A]}z^{[B]} \underset{\text{df}}{=} \begin{cases} z^{[A\cup B]}, & A \cap B = \phi, \\ 0, & A \cap B \neq \phi, \end{cases}$$

where $A \cup B$ is the union of A and B, $A \cap B$ is the intersection of A and B, and ϕ is the empty set. Note that the resulting multiplication is commutative and associative:

$$z^{[A]}z^{[B]} = z^{[B]}z^{[A]},$$

and

$$(z^{[A]}z^{[B]})z^{[C]} = z^{[A]}(z^{[B]}z^{[C]}).$$

Real linear combinations of $z^{[A]}$ define a real algebra. For each A let $|A|$ be the number of elements in A and let W_A and U_A be symmetric functions of $|A|$ variables. Then

$$\mathfrak{W}(z) \underset{\text{df}}{=} \sum_S W_S z^{[S]},$$
$$\mathfrak{U}(z) \underset{\text{df}}{=} \sum_{\phi \neq T} U_T z^{[T]},$$

are functions from $r = (r_1, r_2, \cdots)$ to the algebra. If $A = \{i_1, i_2, \cdots, i_{|A|}\}$, then

$$W_A \text{ maps } r \to \text{ into } W_A(r_{i_1}, r_{i_2}, \cdots, r_{i_{|A|}}).$$

Suppose

$$\mathfrak{W}(z) = \exp \mathfrak{U}(z) \underset{\text{df}}{=} \sum_{n=0}^{\infty} \frac{\mathfrak{U}^n(z)}{n!}. \quad (1)$$

Note that this implies

$$W_\phi \equiv 1$$

and by comparing coefficients of $z^{[T]}$ for $T \neq \phi$,

$$W_T = \sum U_A U_B \cdots$$
$$\text{partition of } T \text{ into } \{A, B, C, \cdots\}.$$

For each A and B such that $|A| = |B|$ suppose $U_A = U_B$. Then write $U_A = U_{|A|}$. This specialization (and the additional ones (1) $U_1(r) \equiv 1$ and (2) $W_A \to 1$ as all the arguments become infinite) is what occurs in UF.

Now return to investigate the consequences of Eq. (1) in the general case. Fix $r_1, r_2, \cdots$. Let $\Omega_1 \cup \Omega_2 = \Omega$ and $\Omega_1 \cap \Omega_2 = \phi$. Suppose

$$W_A = W_{A\cap\Omega_1} W_{A\cap\Omega_2} \quad \text{(product property)} \quad (2)$$

for each A. Let

$$\mathfrak{W}_{\Omega_i} \underset{\text{df}}{=} \sum_{A \subset \Omega_i} W_A z^{[A]}, \qquad i = 1, 2.$$

From Eq. (2)

$$\mathfrak{W} = \mathfrak{W}_{\Omega_1}\mathfrak{W}_{\Omega_2},$$

and conversely this implies (2).
Let

$$\mathfrak{U}_{\Omega_i} \underset{\text{df}}{=} \sum_{\phi \neq B \subset \Omega_i} U_B z^{[B]}, \qquad i = 1, 2.$$

[2] B. Kahn and G. E. Uhlenbeck, Physica 5, 399 (1938), referred to as KU.
[3] Presented at meeting of A.A.A.S. (December, 1962) by H. F. Bohnenblust of California Institute of Technology.
[4] F. Spitzer, Trans. Am. Math. Soc. 82, 323 (1956). See Theorem 3.1, p. 330.

From Eq. (1)

$$\mathfrak{W}_{\Omega_i} = \exp(\mathfrak{U}_{\Omega_i}), \qquad i = 1, 2.$$

Thus

$$\mathfrak{W} = \mathfrak{W}_{\Omega_1}\mathfrak{W}_{\Omega_2} = \exp(\mathfrak{U}_{\Omega_1} + \mathfrak{U}_{\Omega_2}) = \exp(\mathfrak{U}).$$

It follows that $\mathfrak{U}_{\Omega_1} + \mathfrak{U}_{\Omega_2} = \mathfrak{U}$. Further $U_A \not\equiv 0$ implies

$$A \subset \Omega_1 \text{ or } A \subset \Omega_2 \quad \text{(cluster property)}. \tag{3}$$

Thus the product property for $\mathfrak{W}$ implies the cluster property for $\mathfrak{U}$. Analogously reversing the steps shows that cluster property for $\mathfrak{U}$ implies the product property for $\mathfrak{W}$. What has been shown here involves weaker hypotheses than (KU, pp. 402–404) or (UF, pp. 136–137, see footnote p. 137) since it is *not* assumed that U_A and U_B define the same function if $|A| = |B|$. It is neither assumed that $U_1 \equiv 1$ nor $W_A(r_{i_1} \cdots r_{i_a}) \xrightarrow[r_{ij}\to\infty, 1\le i\le a]{} 1$. Further, the choice of Ω_1 and Ω_2 need not be on the basis that $i \in \Omega_1$, $j \in \Omega_2$ imply $|r_i - r_j|$ large.

Write

$$\mathfrak{W} = 1 + \mathfrak{W}_1$$

where

$$\mathfrak{W}_1 = \sum_{\phi \neq A} z^{[A]} W_A.$$

Then solve for $\mathfrak{U}$ to get

$$\mathfrak{U} = \sum_{j=1}^{\infty} (-1)^{j-1} \frac{\mathfrak{W}^j}{j}.$$

For $S \neq \phi$

$$U_S = \sum (-1)^{-1 + \text{No. sets in partition}} (-1 + \text{No. sets in partition})!\, W_A W_B \cdots$$

$$\text{partition of } S \text{ into } \{A, B, C, \cdots\}.$$

This generalizes the inversion of the W–U relations in the aforementioned references.

Note the similarity between Eq. (1) and II, 3 Eq. (14) of UF (p. 132), but bear in mind that Eq. (1) is an equation in the Bohnenblust algebra while II, 3 Eq. (14) is not. Nevertheless this suggests that the existence of a generating function identity for integrals should make one suspect the existence of an analogous generalized generating function identity for integrands. Also the argument above suggests that product property–cluster property equivalence depends on the defining identity (1). The last two sentences suggest for $U - V$ equations (UF III, 3, Eq. 30, p. 158) related to the generating function equation (UF III, 2 Eq. 15, p. 153) that no equivalence between product and cluster properties (see UF, p. 159) exist and further since no neat explicit inversion of UF III, 2 Eqs. 30, 15, p. 153 is available, none should be expected for UF III, 3 Eq. 30, p. 158.

B. Friedman has made an interesting suggestion which almost deduces our modification (just a renormalization) of the Bohnenblust algebra from more conventional considerations. Take a Grassmann algebra[5] of the real vector space M with dimension m. If $\{\zeta_1, \cdots, \zeta_m\}$ is a basis for the conjugate space M' then $\zeta_1, \cdots, \zeta_m$ form a set of generators for the Grassmann algebra. Then

$$\zeta_i \square \zeta_i = 0,$$

$$\zeta_i \square \zeta_j + \zeta_j \square \zeta_i, \qquad 1 \le i < j \le m$$

and the elements $\zeta_{i_1} \square \zeta_{i_2} \square \cdots \square \zeta_{i_r}$ corresponding to the various subsets $\{i_1, i_2, \cdots, i_r\}$ of the set $\{1, 2, \cdots, m\}$ (with $i_1 < i_2 < \cdots < i_r \le m$) are linearly independent. Each element of the algebra may be written as a linear combination of the unit element E and of such elements. Now to make a connection with Bohnenblust's work. Let $A = \{i_1, \cdots, i_r\}$, $i_1 < i_2 < \cdots i_r$, and $B = \{j_1, \cdots, j_s\}$, $j_1 < \cdots < j_s$, be subsets of $\{1, 2, \cdots, m\}$. Think of $z^{[A]}$ as corresponding to $\zeta_{i_1} \square \cdots \square \zeta_{i_r}$ and $z^{[B]}$ as corresponding to $\zeta_{j_1} \square \cdots \square \zeta_{j_s}$. Now $z^{[A]}z^{[B]}$ corresponds (up to a possible change of sign) to $(\zeta_{i_1} \square \cdots \square \zeta_{i_r}) \square (\zeta_{j_1} \square \cdots \square \zeta_{j_s})$ since if $A \cap B \neq \phi$, then $z^{[A]}z^{[B]}$ is zero while $(\zeta_{i_1} \square \cdots \square \zeta_{j_s})$ is in that case also zero. The Grassmann algebra preserves the ordering of the set up to even permutations, while the Bohnenblust algebra is not concerned with the ordering of the set. Possibly properties of Grassmann algebra can be used to explain the formal similarity between (1) and II, 3 Eq. 14 of UF (p. 132).[6]

Note added in proof. A conversation with G. -C. Rota revealed that the modified Bohnenblust algebra presented here is identical with Rota's Poisson algebra I (M. I. T. lecture notes). While he used the algebra to deduce the Möbius function of a finite boolean lattice, the algebra is used here for a different purpose and as a byproduct yields the Möbius function of the lattice of partitions of a finite set [Roberto Frucht and G. -C. Rota. A Möbius-type inversion formula for partitions, Notices Am. Math. Soc. **10**, 495 (1963)].

[5] C. Chevalley, *Theory of Lie Groups I* (Princeton University Press, Princeton, New Jersey, 1946), p. 145.

[6] After this paper had been submitted for publication, the author's attention was called to the "symmetric sequence algebra," pp. 52 *et seq.* in J. Schwartz, *Statistical Mechanics*, New York University mimeographed notes 1957–1958; and the corresponding development in Chap. 1 of D. Ruelle, *Rigorous Results in Statistical Mechanics*, mimeographed lectures at Theoretical Physics Institute, University of Colorado, Summer, 1963. While these tools are also designed to study the W–U relationship and are adequate for the physical purpose, they require stronger hypotheses than that required for the Bohnenblust algebra.

Kinetic Equation for an Inhomogeneous Plasma far from Equilibrium

R. BALESCU AND A. KUSZELL*
Faculté des Sciences, Université Libre de Bruxelles, Brussels, Belgium
(Received 11 February 1964)

A kinetic equation for a nonuniform plasma is derived from the Liouville equation by a general diagram technique. It describes the evolution of a small spatial inhomogeneity in a plasma whose velocity distribution is far from equilibrium (and hence time-dependent). The equation is valid for short and long times, within the ring approximation. Its explicit form is obtained by the exact closed solution of a singular integral equation. The kinetic equation is non-Markoffian and, contrary to the corresponding homogeneous equation, keeps a trace of this character even in the limit of long times. Only when the velocity distribution and the two-body correlation function reach thermal equilibrium does the equation reduce to a Markoffian limit. The latter is identical with the kinetic equation derived earlier by Guernsey. The treatment of unstable inhomogeneous plasmas is briefly indicated.

1. INTRODUCTION

IN recent years the problem of kinetic equations for plasmas has been extensively studied. In 1960 the kinetic equation for a stable spatially homogeneous plasma was derived independently by one of us,[1,2] Guernsey,[3] Lenard,[4] and Rostoker and Rosenbluth.[5] In all these papers the plasma is assumed to be characterized by a small value of its characteristic dimensionless parameter $e^2c^{\frac{1}{3}}\beta$ (where e is the electronic charge, c the average number density and β the inverse of the average kinetic energy); the resulting approximation will be called the ring approximation. The stability criterion is the well-known condition that the Vlassov dielectric constant has no zero in the upper half-plane (see, for example, Penrose,[6] and Balescu[2]).

These works have been generalized in the following two ways. First, the kinetic equation for an unstable homogeneous plasma has been derived by one of us[7] and, independently, in a less general case, by Rutherford and Frieman.[8] A second type of generalization, to which considerable effort has been devoted, is the derivation of a kinetic equation for inhomogeneous plasmas.

Before the discussion of previous work in this field, a clear statement of the problem is desirable. The inhomogeneity of a general physical system can be characterized by two parameters: the *maximum amplitude of the inhomogeneity*, μ, and the *rate of spatial variation of the inhomogeneity*, γ. In the case of a plasma, it turns out that the problem can be treated explicitly in the cases in which one or the other of these two parameters is small. (Of course the interactions are still treated in the ring approximation.) The case in which γ is assumed small can be characterized more precisely by introducing the following two characteristic lengths: the mean free path L_f, and the hydrodynamical length L_h defined, for instance, as $c/\max|\nabla c(\mathbf{x})|$, where c is the average density, $c(\mathbf{x})$ the local density, and $\nabla \equiv \partial/\partial\mathbf{x}$. The case in which

$$\gamma \equiv L_f/L_h \ll 1 \tag{1.1}$$

is important in the study of the typically hydrodynamical phenomena as well as of transport phenomena in which the constraints are externally imposed (e.g., temperature gradients which we are able to create experimentally usually extend over distances much longer than the mean free path). The solution of this problem is relatively very simple, once the corresponding problem for the homogeneous system has been solved. The kinetic equation in this approximation has been derived by one of us (Refs. 1 and 2).

The second type of approximation is characterized by the condition

$$\mu \equiv \max[|c(\mathbf{x}) - c|/c] \ll 1. \tag{1.2}$$

The gradients, on the other hand, can be as strong as one wishes. A schematical illustration of the two situations under discussion is shown in Fig. 1. The latter situation is important in the study of plasma

* On leave of absence from the Institute of Nuclear Research, Warsaw, Poland.
[1] R. Balescu, Phys. Fluids 3, 52 (1960).
[2] R. Balescu, *Statistical Mechanics of Charged Particles* (John Wiley & Sons, Inc., New York, 1963).
[3] R. Guernsey, Thesis, University of Michigan, Ann Arbor, Michigan, (1960).
[4] A. Lenard, Ann. Phys. (N. Y.) 10, 390 (1960).
[5] M. N. Rosenbluth and N. Rostoker, Phys. Fluids 3, 1 (1960).
[6] O. Penrose, Phys. Fluids 3, 288 (1960).
[7] R. Balescu, J. Math. Phys. 4, 1009 (1963).
[8] P. Rutherford and E. A. Frieman, Phys. Fluids 6, 1139 (1963).

oscillations. It turns out to be considerably more complex than the study of the first situation. Moreover, a study of the plasma oscillations requires a kinetic equation valid for short times (of order ω_p^{-1}, the period of plasma oscillations), whereas in the study of static transport coefficients an asymptotic equation valid for times much longer than ω_p^{-1} is sufficient. Let us remark however that the most interesting domain in the study of plasma oscillations is really the domain of long wavelengths; short wavelengths are indeed very rapidly killed by Landau damping before any action of the "collisions". It seems therefore that the domain in which the complete "μ equation" significantly departs from the corresponding "γ equation" is of a rather academic interest. This remark might be interesting in view of the fact that the complete equation is exceedingly complicated. However this equation throws a new light on many questions related to the theoretical structure of kinetic equations for plasmas, and this was the main motivation of our having treated this problem anew.

The first attempt in this direction appears in Guernsey's thesis,[3] in which however no definite answer was obtained. Ichikawa[9] went a step further by deriving a fundamental integral equation whose solution would provide the kinetic equation. However, he solved this equation only in a very rough approximation. Wu and Rosenberg[10] made very extensive calculations with a similar integral equation but the approximations they used were quite unclear. Fried and Wyld[11] also attacked the problem, yet could only discuss a few simple models. Meeron[12] also derived an integral equation for this problem, but did not solve it. The first complete and satisfactory solution of the problem is due to Guernsey.[13] By adjoining to the assumption of small μ the additional hypothesis that the plasma is near to the thermal equilibrium, he has been able to solve the fundamental integral equation exactly and hence to derive a rigorous kinetic equation for an inhomogeneous plasma, valid for all times. The technique used by him is a Laplace transformation of the first two equations of the BBGKY hierarchy, closed in the manner appropriate to the ring approximation (i.e., expressing the three-body correlations as a superposition of two-body correlations).

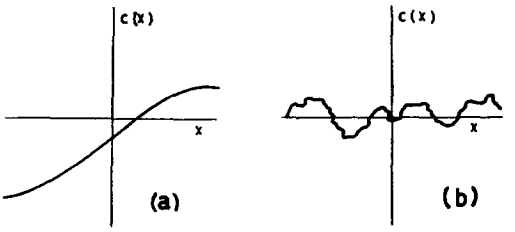


FIG. 1. The two extreme types of inhomogeneous systems (a) γ small, μ large; (b) γ large, μ small.

This method can only be applied if the velocity distribution is time-independent.

Our own main purpose in the present paper is to get rid of Guernsey's assumption of closeness to equilibrium. On one hand, the crucial role played by this assumption is rather uncomfortable from a theoretical point of view; on the other hand, it excludes any study of unstable systems which, by definition, are very far from equilibrium. Our method is based on an extensive use of *"partial Laplace transforms"* introduced by Résibois[14] and one of us.[7,2] The main functional $\mathfrak{F}$ occurring in the theory is simultaneously a function of time variables and of Laplace variables. As is seen from the formal part of the theory, presented in Secs. 2 and 3, the diagram technique developed by Prigogine and one of us[15,16,2] is a framework in which the concept of partial Laplace transformation enters quite naturally. As in the homogeneous case, the kinetic equation is determined by a fundamental functional of the one-body distribution function; in order to evaluate this functional one must solve a singular integral equation derived in Sec. 3. In Sec. 4 this equation is shown to reduce to Guernsey's corresponding equation when the velocity distribution and the homogeneous two-body correlation function have their equilibrium value. The fundamental integral equation is solved exactly in Sec. 5, where the main properties of the kinetic equation are discussed. In Sec. 6 we indicate briefly how these results can be extended to the unstable case. However, as this extension follows exactly the same line as in the homogeneous case,[7] we do not write out the calculations explicitly.

2. DERIVATION OF FORMAL EQUATION

The system studied in the present paper is an idealized plasma which consists of a slightly inhomogeneous (i.e., small μ) electron gas of mean density $c \equiv N/\Omega$, imbedded in a continuous neutralizing

[9] Y. H. Ichikawa, Progr. Theoret. Phys. (Kyoto) 24, 1083 (1960).
[10] T. Y. Wu and R. L. Rosenberg, Can. J. Phys. 40, 463 (1962).
[11] B. D. Fried and H. W. Wyld, Phys. Rev. 122, 1 (1961).
[12] E. Meeron (to be published).
[13] R. Guernsey, Phys. Fluids 5, 322 (1962).
[14] P. Résibois, Phys. Fluids 6, 817 (1963).
[15] I. Prigogine and R. Balescu, Physica 25, 281 (1959).
[16] I. Prigogine, *Non Equilibrium Statistical Mechanics* (John Wiley & Sons, Inc., New York, 1962).

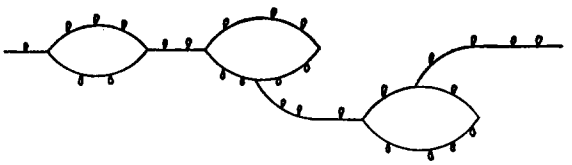


FIG. 2. A general contribution to $\rho_k(\alpha; t)$ in the (small μ) ring approximation.

positive background. The system is described by the Liouville equation

$$\mathfrak{L}f_N \equiv (\mathfrak{L}_0 + e^2\mathfrak{L}')f_N = 0, \tag{2.1}$$

where

$$\mathfrak{L}_0 = \partial_t + \sum_i \mathbf{v}_i \cdot \nabla_i,$$

$$\mathfrak{L}' = -\sum_{i<n} m^{-1}(\nabla V_{in}) \cdot (\partial_i - \partial_n),$$

$$\partial_t = \partial/\partial t; \quad \nabla_i = \partial/\partial \mathbf{x}_i; \quad \partial_i = \partial/\partial \mathbf{v}_i, \tag{2.2}$$

and

$$V_{in} = (|\mathbf{x}_i - \mathbf{x}_n|)^{-1} = \frac{1}{2\pi^2} \int d\mathbf{k} \frac{e^{i\mathbf{k}\cdot(\mathbf{x}_i - \mathbf{x}_n)}}{k^2}. \tag{2.3}$$

The one-particle distribution function in Fourier representation has the form

$$f(\mathbf{x}, \mathbf{v}; t) = c\left[\varphi(\mathbf{v}; t) + \int d\mathbf{k} \rho_{\mathbf{k}}(\mathbf{v}; t) e^{i\mathbf{k}\cdot\mathbf{x}}\right]. \tag{2.4}$$

The Fourier transform $\rho_{\mathbf{k}}(\mathbf{v}; t)$ will be called the *inhomogeneity factor*; $\varphi(\mathbf{v}; t)$ is the *velocity distribution*. In this paper we use the theory developed by Prigogine and collaborators.[2,15,16] In this theory the solution of Eq. (2.1) for the Fourier components of the N-body distribution function is expressed in the following form:

$$\rho_{\{k\}}(\mathbf{v}; t) = (2\pi)^{-1} \sum_{n=0}^{\infty} \int_c dz e^{-izt} (-e^2)^n \sum_{\{k'\}} (8\pi^3/\Omega)^{\nu'-\nu}$$

$$\times \langle \{k\} | R_0(z)[\mathfrak{L}' R_0(z)]^n | \{k'\} \rangle \rho_{\{k'\}}(v; 0), \tag{2.5}$$

where $\{k\}$ is the set of wave vectors corresponding to all the N particles of the system, $\mathbf{k}_1, \cdots, \mathbf{k}_N$, ν and ν' are the number of independent nonvanishing wave vectors in the sets $\{k\}$ and $\{k'\}$, respectively, and $R_0(z)$ is the resolvent operator for $\mathfrak{L}_0$ and has the form

$$\langle \{k\} | R_0(z) | \{k'\} \rangle$$

$$= \frac{1}{i(\sum_i \mathbf{k}_i \cdot \mathbf{v}_i - z)} \prod_i \delta(\mathbf{k}_i - \mathbf{k}'_i). \tag{2.6}$$

C is the usual Laplace contour consisting of a parallel to the real axis lying above all singularities of the integrand.

Each term in the expansion (2.5) can be represented by a diagram introduced by Prigogine and one of us.[2,15,16]

We will first discuss the choice of diagrams. This operation is somewhat less clear for a short-time equation (which we are looking for) than for an asymptotic equation. In the latter case it has been shown that the ring approximation is equivalent to retaining all diagrams of order $(e^2c)^m(e^2t)^n$ where m and n are arbitrary positive integers (Refs. 1 and 2). This criterion looses its absolute meaning when applied to short times. In order to keep classification of the terms, we shall therefore be guided by the asymptotic behavior and define the ring approximation as follows: All terms which behave asymptotically as $(e^2c)^m(e^2t)^n$ are retained, but the time dependence of the corresponding diagrams is evaluated exactly. The contributions to the inhomogeneity factor $\rho_{\mathbf{k}}(\alpha; t)$ [here and below, $f(\alpha) \equiv f(\mathbf{v}_\alpha)$] which is the most important quantity characterizing an inhomogeneous system, consist of diagrams having one external line at left; with the assumption that the inhomogeneity is initially small ($\mu \ll 1$), there is also a single external line at right. It can then be shown that there can be at most one inhomogeneity line running through the diagram. [An inhomogeneity line is a line representing a factor $\rho_{\mathbf{k}}(j)$ which is therefore not included in a diagonal fragment (see Severne[17])]. The most general diagram is then easily seen to be of the type drawn in Fig. 2: it is a succession of rings linked together by inhomogeneity lines; an arbitrary number of loops can be drawn on every line (Ref. 2).

We now go over to the summation of the diagrams. The method used here is a straightforward generalization of the one exposed in Chap. 2 and Appendix 8 of Ref. 2 (and also in Ref. 7). We shall therefore only indicate the main lines of the reasoning. The Fourier component of f_N with a single nonvanishing wave vector $\mathbf{k}$ can be written in the following form:

$$\rho_{\mathbf{k}}(\alpha \mid \cdots; t) = (2\pi)^{-1} \sum_{n=0}^{\infty} \int_c dz e^{-izt} \frac{1}{i(\mathbf{k}\cdot\mathbf{v}_\alpha - z)}$$

$$\times \{\rho_{\mathbf{k}}(\alpha \mid \cdots; 0) - (4\pi e^2/mk^2\Omega)$$

$$\times \sum_i i\mathbf{k} \cdot \partial_{\alpha i} \tilde{\rho}_{\mathbf{k}}(j \mid \alpha, \cdots; z)$$

$$+ \sum_i \mathfrak{R}_{\mathbf{k}}^{(\alpha i)}(z) \tilde{\rho}_{\mathbf{k}}(j \mid \alpha, \cdots; z)\}. \tag{2.7}$$

[17] G. Severne, thesis, Université Libre de Bruxelles, Brussels, Belgium (1963).

We use here the same notations as in Ref. 2. Let us recall that $\tilde{\rho}_{\mathbf{k}}(j \mid \cdots ; z)$ is the Laplace transform of $\rho_{\mathbf{k}}(j \mid \cdots ; t)$. $\mathfrak{R}_{\mathbf{k}}^{(\alpha i)}(z)$ is an abbreviation for the following expression:

$$\mathfrak{R}_{\mathbf{k}}^{(\alpha i)} = \sum_{n=1}^{\infty} (-e^2)^{n+1} \langle \mathbf{k}^{(\alpha)} \mid \mathfrak{L}'[R_0(z)\mathfrak{L}']^n \mid \mathbf{k}^{(i)} \rangle, \quad (2.8)$$

the sum being over all inhomogeneous rings of the type shown in Fig. 3. $|\mathbf{k}^{(i)}\rangle$ denotes a state in which particle j has a wave vector $\mathbf{k}$, whereas all other particles have a vanishing wave vector. In (2.7), the first term on the rhs corresponds to a bare line, the second to the sum of all diagrams beginning at left with a loop, and the third to the sum of all diagrams beginning with a ring. We now integrate this expression over all velocities but $\mathbf{v}_\alpha$ and take the time derivative in order to obtain the kinetic equation for $\rho_{\mathbf{k}}(\alpha; t)$ in the form

$$\partial_t \rho_{\mathbf{k}}(\alpha; t) + i\mathbf{k}\cdot\mathbf{v}_\alpha \rho_{\mathbf{k}}(\alpha; t) = \omega_p^2 k^{-2} i\mathbf{k}\cdot\partial_\alpha \varphi(\alpha; t)$$
$$\times \int d\mathbf{v}_j \rho_{\mathbf{k}}(j; t) + (2\pi)^{-1} \sum_i \int_{(\alpha)} (d\mathbf{v})^{N-1}$$
$$\times \int_c dz\, e^{-izt} \mathfrak{R}_{\mathbf{k}}^{(\alpha i)}(z) \tilde{\rho}_{\mathbf{k}}(j \mid \alpha, \cdots ; z). \quad (2.9)$$

We made use here of the factorization property[2]

$$\rho_{\mathbf{k}}(j \mid 1, \cdots, s; t) = \rho_{\mathbf{k}}(j; t) \prod_{i=1}^{s} \varphi(i; t). \quad (2.10)$$

A similar factorization can however not be applied to the Laplace transform $\tilde{\rho}_{\mathbf{k}}(j \mid \cdots ; z)$ appearing in the last term of Eq. (2.9). The summation procedure of the ring diagrams depends however crucially on such a property; we shall therefore take the inverse Laplace transform of $\tilde{\rho}_{\mathbf{k}}(j \mid \cdots ; z)$ and obtain

$$D_t \rho_{\mathbf{k}}(\alpha; t) \equiv (\partial_t + i\mathbf{k}\cdot\mathbf{v}_\alpha)\rho_{\mathbf{k}}(\alpha; t) - \omega_p^2 k^{-2} i\mathbf{k}\cdot\partial_\alpha \varphi(\alpha; t)$$
$$\times \int d\mathbf{v}_j \rho_{\mathbf{k}}(j; t) = \sum_i \int_0^t d\tau (2\pi)^{-1} \int_c dz\, e^{-iz\tau}$$
$$\times \int_{(\alpha)} (d\mathbf{v})^{N-1} \mathfrak{R}_{\mathbf{k}}^{(\alpha i)}(z) \rho_{\mathbf{k}}(j \mid \alpha, \cdots ; t - \tau). \quad (2.11)$$

One recognizes in the first term of the lhs the familiar *linearized Vlassov* term. The rhs represents

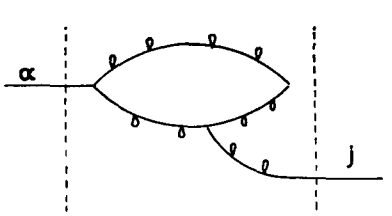


FIG. 3. A typical inhomogeneous ring.

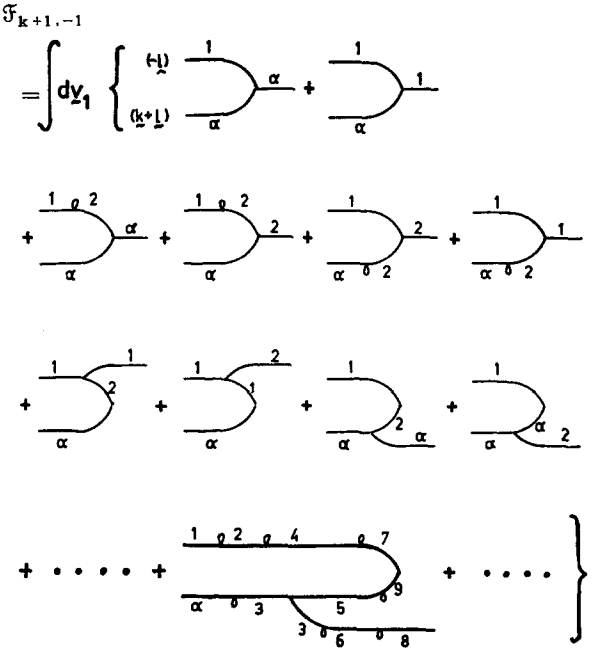


FIG. 4. The series for $\mathfrak{F}_{k+1,-1}(\alpha; z, t - \tau)$.

the effect of the *"collisions"*. Equation (2.11) is a typical non-Markoffian integral equation (in the time variable), relating the present evolution to the whole history of the system. This is a characteristic feature of all short-time equations. Equation (2.11) is the extension to inhomogeneous systems of Eq. (A8.6) of Ref. 2. Let us stress at this point the occurrence of the operation we called a *"partial Laplace transformation"*. The Fourier component $\tilde{\rho}_{\mathbf{k}}(j \mid \cdots ; z)$ has been transformed back to the time-dependent $\rho_{\mathbf{k}}(j \mid \cdots ; t)$, but the resolvent operator $\mathfrak{R}_{\mathbf{k}}^{(\alpha i)}(z)$ has *not* been transformed. This operation explicitly introduces distribution functions *retarded in time*. It will be seen that the concept of partial Laplace transforms allows us to avoid completely Guernsey's assumption of closeness to equilibrium.

3. THE FUNDAMENTAL INTEGRAL EQUATION

In order to perform the summation of the rings, we note that all ring diagrams begin with the same vertex at left; we therefore write this vertex explicitly, thereby obtaining the following equation:

$$D_t \rho_{\mathbf{k}}(\alpha; t) = -\int_0^t d\tau (2\pi)^{-1} \int_c dz\, e^{-iz\tau}$$
$$\times \int d\mathbf{l}\, \omega_p^2 l^{-2} i\mathbf{l}\cdot\partial_\alpha \mathfrak{F}_{\mathbf{k}+1,-1}(\alpha; z, t - \tau). \quad (3.1)$$

The function $\mathfrak{F}_{\mathbf{k}+1,-1}(\alpha; z, t - \tau)$ is defined by a series represented diagrammatically in Fig. 4. Its simultaneous dependence on a Laplace variable z

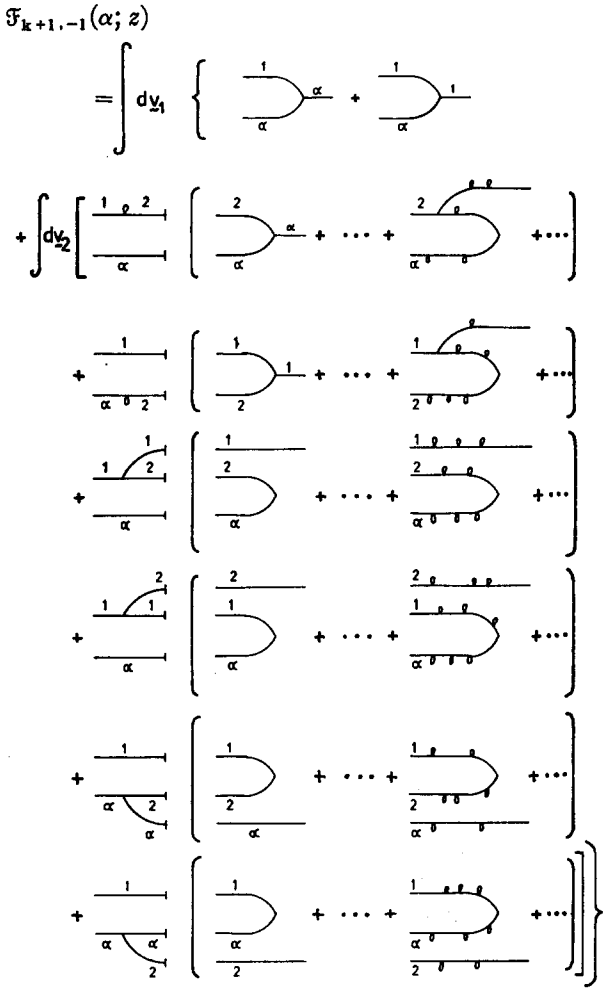


FIG. 5. The recurrence relation for $\mathfrak{F}_{k+1,-1}(\alpha; z, t-\tau)$.

and a time variable $t - \tau$ is characteristic of a "partial Laplace transform". [The argument $t - \tau$ will frequently not be written explicitly, unless required by clarity.] This series can be summed by establishing a recurrence relation as in the homogeneous case. This relation is obtained by splitting off the leftmost vertex of each type of diagram in the series: it is shown diagrammatically in Fig. 5. One easily sees that in the first two brackets the series for $\mathfrak{F}_{k+1,-1}(\alpha; z)$ is reproduced; in the remaining brackets there appears a new quantity denoted by $\Phi_{k,1}(\alpha, \beta, \gamma; z)$; it is defined by the series of Fig. 6. From Fig. 5 we can write the recurrence relation in analytical form:

$$\mathfrak{F}_{k+1,-1}(\alpha; z) = \int d\mathbf{v}_1 [i(\mathbf{k}+\mathbf{l})\cdot\mathbf{v}_\alpha - \mathbf{l}\cdot\mathbf{v}_1 - iz]^{-1}$$
$$\times \Big\{ (e^2/2\pi^2 m)[l^{-2} i\mathbf{l}\cdot\partial_{\alpha 1}\rho_k(\alpha)\varphi(1)$$
$$+ |\mathbf{k}+\mathbf{l}|^{-2} i(\mathbf{k}+\mathbf{l})\cdot\partial_{\alpha 1}\varphi(\alpha)\rho_k(1)]$$
$$- \omega_p^2 l^{-2} i\mathbf{l}\cdot\partial_1\varphi(1)\mathfrak{F}_{k+1,-1}(\alpha; z)$$
$$+ \omega_p^2 |\mathbf{k}+\mathbf{l}|^{-2} i(\mathbf{k}+\mathbf{l})\cdot\partial_\alpha\varphi(\alpha)\mathfrak{F}_{-1,k+1}(1; z)$$
$$+ \omega_p^2 \int d\mathbf{v}_2 [-|\mathbf{k}+\mathbf{l}|^{-2} i(\mathbf{k}+\mathbf{l})\cdot\partial_1\Phi_{k,k+1}(1, \alpha, 2; z)$$
$$+ k^{-2} i\mathbf{k}\cdot\partial_1\Phi_{k,k+1}(2, \alpha, 1; z) + l^{-2} i\mathbf{l}\cdot\partial_\alpha\Phi_{k,1}$$
$$\times (\alpha, 2, 1; z) + k^{-2} i\mathbf{k}\cdot\partial_\alpha\Phi_{k,1}(2, \alpha, 1; z)]\Big\}. \quad (3.2)$$

Note that here and in the following equations the functions $\varphi(j)$ and $\rho_k(j)$ are *always evaluated at time* $t - \tau$, except when otherwise stated explicitly.

We now evaluate the function $\Phi_{k,1}(\alpha, \beta, \gamma; z)$ in terms of known functions. This is very easily done by using a theorem due to Résibois[14] concerning a factorization property of disconnected diagrams. We use a form of this theorem which is explained in detail in Appendix 10 of Ref. 2, to which we refer the reader for a general statement of the property. The application of Résibois' theorem to our case is as follows.

Consider first all the diagrams consisting of loops on a single line, i.e., the upper part of the diagrams of Fig. 6. Their sum is nothing other than the well-known solution of the linearized Vlassov equation, obtained by Landau[18] (see also Chap. 3 of Ref. 2). More precisely, calling $\epsilon_k(z/k)$ the dielectric con-

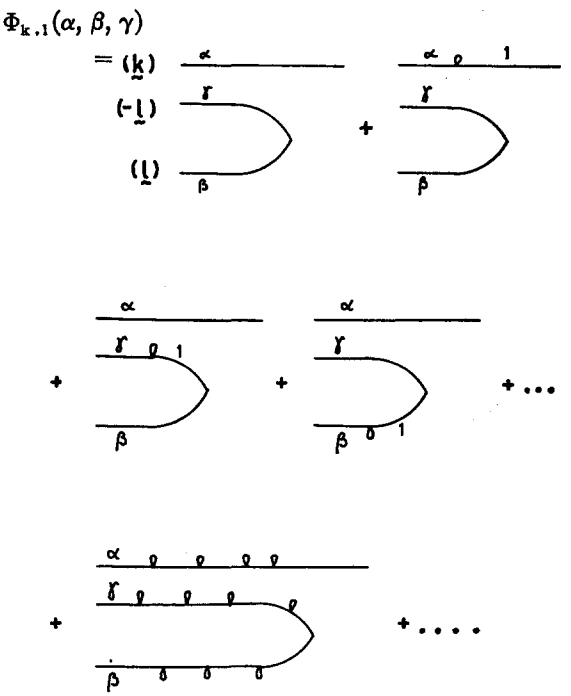


FIG. 6. The series for $\Phi_{k,1}(\alpha, \beta, \gamma; z, t-\tau)$.

[18] L. Landau, Soviet Phys.—Tech. Phys. 10, 25 (1946).

stant in the Vlassov approximation

$$\epsilon_{\mathbf{k}}(z/k) = 1 - \frac{\omega_p^2}{k^2}\int d\mathbf{v}\,\frac{\mathbf{k}\cdot\partial\varphi(\mathbf{v};\,t-\tau)}{\mathbf{k}\cdot\mathbf{v}-z}, \tag{3.3}$$

and letting the kernel $\mathcal{U}_{\mathbf{k}}(\mathbf{v}_\alpha \mid \mathbf{v}_\beta;\, z)$ be defined as

$$\mathcal{U}_{\mathbf{k}}(\mathbf{v}_\alpha \mid \mathbf{v}_\beta; z) = \frac{1}{i(\mathbf{k}\cdot\mathbf{v}_\beta - z)}\left[\delta(\mathbf{v}_\alpha - \mathbf{v}_\beta) - \frac{\omega_p^2}{k^2\epsilon_{\mathbf{k}}(z/k)}\frac{\mathbf{k}\cdot\partial_\alpha\varphi(\alpha;\,t-\tau)}{\mathbf{k}\cdot\mathbf{v}_\alpha - z}\right], \tag{3.4}$$

the sum of these diagrams, denoted by $R_{\mathbf{k}}(\mathbf{v}_\alpha; z, t-\tau)$ is

$$R_{\mathbf{k}}(\mathbf{v}_\alpha; z,\, t-\tau) = \int d\mathbf{v}_1 \mathcal{U}(\mathbf{v}_\alpha \mid \mathbf{v}_1; z)\rho_{\mathbf{k}}(\mathbf{v}_1;\, t-\tau). \tag{3.5}$$

It will be convenient for later use to introduce also the inverse Laplace transform of $R_{\mathbf{k}}(\mathbf{v}_\alpha;\, z,\, t-\tau)$,

$$\hat{R}_{\mathbf{k}}(\mathbf{v}_\alpha;\, \sigma,\, t-\tau) = (2\pi)^{-1}\int_c dz\, e^{-is\sigma}R_{\mathbf{k}}(\mathbf{v}_\alpha; z,\, t-\tau). \tag{3.6}$$

$\hat{R}_{\mathbf{k}}(\mathbf{v}_\alpha;\, \sigma,\, t-\tau)$ is clearly the solution of the linearized Vlassov equation for the initial condition,

$$\hat{R}_{\mathbf{k}}(\mathbf{v}_\alpha; 0,\, t-\tau) = \rho_{\mathbf{k}}(\mathbf{v}_\alpha;\, t-\tau). \tag{3.7}$$

Let us also note that the density excess, defined as the integral of $\tilde{\rho}_{\mathbf{k}}(\mathbf{v}; z)$ over the velocity is given, in the Vlassov approximation, by[2]

$$H_{\mathbf{k}}(z,\, t-\tau) = \int d\mathbf{v}\, R_{\mathbf{k}}(\mathbf{v}; z,\, t-\tau) = \frac{1}{\epsilon_{\mathbf{k}}(z/k)}\int d\mathbf{v}\,\frac{\rho_{\mathbf{k}}(\mathbf{v};\, t-\tau)}{i(\mathbf{k}\cdot\mathbf{v}-z)}. \tag{3.8}$$

As for $R_{\mathbf{k}}$, we define the function $\hat{H}_{\mathbf{k}}(\sigma,\, t-\tau)$ by

$$\hat{H}_{\mathbf{k}}(\sigma,\, t-\tau) = (2\pi)^{-1}\int_c dz\, e^{-is\sigma}H_{\mathbf{k}}(z,\, t-\tau). \tag{3.9}$$

The value of this function for $\sigma = 0$ is the exact density excess at time $t-\tau$,

$$\hat{H}_{\mathbf{k}}(0,\, t-\tau) = \int d\mathbf{v}\rho_{\mathbf{k}}(\mathbf{v};\, t-\tau). \tag{3.10}$$

Consider now the sum of all lower parts of the diagrams in Fig. 6. This sum is nothing other than the two-body correlation function in the ring approximation for a *homogeneous* plasma. From Ref. 2 we know that this function, denoted by $G_{\mathbf{k},-\mathbf{k}}(\beta, \gamma; z)$, is of the following form:

$$G_{\mathbf{k},-\mathbf{k}}(\beta, \gamma; z,\, t-\tau) = \frac{1}{i[\mathbf{k}\cdot(\mathbf{v}_\beta - \mathbf{v}_\gamma) - z]} \times \left\{\frac{e^2}{2\pi^2 mk^2}\, i\mathbf{k}\cdot\partial_{\beta\gamma}\varphi(\beta;\, t-\tau)\varphi(\gamma;\, t-\tau) - id_{\mathbf{k}}(\gamma;\, t-\tau)F_{\mathbf{k}}(\beta; z,\, t-\tau) + id_{\mathbf{k}}(\beta;\, t-\tau)F_{-\mathbf{k}}(\gamma; z,\, t-\tau)\right\}, \tag{3.11}$$

where

$$d_{\mathbf{k}}(\beta;\, t-\tau) = \omega_p^2 k^{-2}\mathbf{k}\cdot\partial_\beta\varphi(\beta;\, t-\tau), \tag{3.12}$$

and

$$F_{\mathbf{k}}(\mathbf{v}_\beta; z,\, t-\tau) = \int d\mathbf{v}_\gamma G_{\mathbf{k},-\mathbf{k}}(\mathbf{v}_\beta, \mathbf{v}_\gamma; z,\, t-\tau). \tag{3.13}$$

The function $F_{\mathbf{k}}(\mathbf{v}_\beta;\, z,\, t-\tau)$ is the fundamental function of the kinetic theory of homogeneous plasmas, where it appears as the solution of an integral equation which can be solved exactly in closed form (Appendix 8, Ref. 2): It is therefore considered here as a known function.

We now collect our partial results. The sum of all disconnected diagrams of Fig. 6, obtained by putting together the two components and permuting the relative order of their two sets of vertices in all possible ways is given, by Résibois' theorem, by the following convolution product:

$$\Phi_{\mathbf{k},1}(\alpha, \beta, \gamma; z) = (2\pi)^{-1}\int_{c'} dz' R_{\mathbf{k}}(\alpha; z')G_{1,-1}(\beta, \gamma; z - z'). \tag{3.14}$$

The contour C' lies above all singularities of $R_{\mathbf{k}}(z')$ and below all those of $G_{1,-1}(z - z')$.

We are now ready for the substitution of (3.14) into Eq. (3.2). Using also (3.8) and (3.13), we obtain

$$\epsilon_{-1}\left[\frac{z - (\mathbf{k}+1)\cdot\mathbf{v}_\alpha}{l}\right]\mathfrak{F}_{\mathbf{k}+1,-1}(\alpha; z) - d_{\mathbf{k}+1}(\alpha)\int d\mathbf{v}_1\,\frac{\mathfrak{F}_{-1,\mathbf{k}+1}(1; z)}{(\mathbf{k}+1)\cdot\mathbf{v}_\alpha - 1\cdot\mathbf{v}_1 - z} = q_{\mathbf{k}+1,-1}(\alpha; z), \tag{3.15}$$

where

$$q_{\mathbf{k}+1,-1}(\alpha; z) = \int d\mathbf{v}_1\,\frac{1}{(\mathbf{k}+1)\cdot\mathbf{v}_\alpha - 1\cdot\mathbf{v}_1 - z} \times \left\{\frac{e^2}{2\pi^2 ml^2}\, 1\cdot\partial_{\alpha 1}\rho_{\mathbf{k}}(\alpha)\varphi(1) + \frac{e^2}{2\pi^2 m\,|\mathbf{k}+1|^2}\,(\mathbf{k}+1)\cdot\partial_{\alpha 1}\rho_{\mathbf{k}}(1)\varphi(\alpha) + \omega_p^2(2\pi)^{-1}\int dz'[k^{-2}\mathbf{k}\cdot\partial_\alpha G_{1,-1}(\alpha, 1; z - z')H_{\mathbf{k}}(z')\right.$$

$$- |\mathbf{k}+\mathbf{l}|^{-2}(\mathbf{k}+\mathbf{l})\cdot\partial_1 R_{\mathbf{k}}(1;z')F_{\mathbf{k}+\mathbf{l}}(\alpha;z-z')$$
$$+ l^{-2}\mathbf{l}\cdot\partial_\alpha R_{\mathbf{k}}(\alpha;z')F_{-\mathbf{l}}(1;z-z')$$
$$+ k^{-2}\mathbf{k}\cdot\partial_1 G_{\mathbf{k}+\mathbf{l},-\mathbf{k}-\mathbf{l}}(\alpha,1;z-z')H_{\mathbf{k}}(z')]\Big\}. \quad (3.16)$$

Equation (3.16) is the fundamental integral equation whose solution provides the explicit form of the collision operator in the ring approximation.

4. THE FUNDAMENTAL INTEGRAL EQUATION NEAR THE EQUILIBRIUM

Before proceeding with the solution of Eq. (3.16), we specialize it first to the case in which the system is near equilibrium. In this way we can compare our results with Guernsey's.[13] We therefore assume that $\varphi(\alpha;t)$ is time-independent, and equal to the Maxwell distribution:

$$\varphi^0(\alpha) = (m\beta/2\pi)^{\frac{3}{2}}\exp(-\tfrac{1}{2}\beta m v_\alpha^2), \quad (4.1)$$

where $\beta = (kT)^{-1}$. In the subsequent text all quantities evaluated for $\varphi(\alpha) = \varphi^0(\alpha)$ will always be denoted by a superscript ^{0}.

Let $\rho_{\mathbf{k},-\mathbf{k}}(\alpha,\beta;t)$ be the (Fourier-transformed) two-body correlation function in a homogeneous system. It is related to the partial Laplace transform $G_{\mathbf{k},-\mathbf{k}}(\alpha,\beta;z,t-\tau)$ defined in (3.11) by the equation

$$\rho_{\mathbf{k},-\mathbf{k}}(\alpha,\beta;t) = (2\pi)^{-1}\int_c dz\, e^{-izt} \times \int_0^\infty d\tau\, e^{iz\tau}G_{\mathbf{k},-\mathbf{k}}(\alpha,\beta;z,t-\tau). \quad (4.2)$$

When the velocity distribution has its equilibrium value (4.1), it is easily seen that $G^0_{\mathbf{k},-\mathbf{k}}(\alpha,\beta;z)$ does not depend on $t-\tau$ (indeed, the dependence on $t-\tau$ comes entirely from the time dependence of the velocity distribution). Hence,

$$\rho^0_{\mathbf{k},-\mathbf{k}}(\alpha,\beta;t) = (2\pi)^{-1}\int_c dz e^{-izt}\frac{G^0_{\mathbf{k},-\mathbf{k}}(\alpha,\beta;z)}{-iz}. \quad (4.3)$$

We may stress at this point that $\rho^0_{\mathbf{k},-\mathbf{k}}(\alpha,\beta;t)$ is *not* time-independent. Indeed, $G^0_{\mathbf{k},-\mathbf{k}}(\alpha,\beta;z)$ regarded as a function of z, has singularities in the lower half-plane; hence $\rho^0_{\mathbf{k},-\mathbf{k}}(\alpha,\beta;t)$ consists of an asymptotic, time-independent term, equal to $G^0_{\mathbf{k},-\mathbf{k}}(\alpha,\beta;0)$, and of a number of transient contributions coming from the singularities of $G^0_{\mathbf{k},-\mathbf{k}}(\alpha,\beta;z)$. Guernsey's approximation consists of taking at the initial time not only $\varphi(\mathbf{v};t) = \varphi^0(\mathbf{v})$, but also of neglecting the transient contributions to the correlation function, and hence assuming $\rho^0_{\mathbf{k},-\mathbf{k}}(\alpha,\beta;t) = G^0_{\mathbf{k},-\mathbf{k}}(\alpha,\beta;0)$. This is quite natural in view of the nature of his problem. In the situation treated by this author, one first waits a very long time until complete equilibrium is established; hence $\varphi^0(\mathbf{v})$, and all correlation functions have their asymptotic value. One then introduces a perturbation consisting of an inhomogeneity which does not affect the correlations. Only in this (rather idealized) case is one allowed to neglect the transient contribution to $\rho^0_{\mathbf{k},-\mathbf{k}}(\alpha,\beta;t)$ in a short-time equation.

From (4.3) we conclude that in equilibrium

$$G^0_{\mathbf{k},-\mathbf{k}}(\alpha,\beta;z) = -iz\int_0^\infty d\sigma\, e^{iz\sigma}\rho^0_{\mathbf{k},-\mathbf{k}}(\alpha,\beta;\sigma). \quad (4.4)$$

We are now ready for the evaluation of the convolution integrals appearing in Eq. (3.16). The first of these, involving G and H, can be evaluated using (4.4) and (3.9),

$$(2\pi)^{-1}\int_{c'} dz' G^0_{1,-1}(\alpha,1;z-z')H_{\mathbf{k}}(z';t-\tau)$$
$$= -\int_0^\infty d\sigma\rho^0_{1,-1}(\alpha,1;\sigma) \times \int_{c'} dz' i(z-z')e^{i(z-z')\sigma}H_{\mathbf{k}}(z',t-\tau)$$
$$= -\int_0^\infty d\sigma\rho^0_{1,-1}(\alpha,1;\sigma)\frac{\partial}{\partial\sigma}\{e^{iz\sigma}\hat{H}_{\mathbf{k}}(\sigma,t-\tau)\}$$
$$= \rho^0_{1,-1}(\alpha,1;0)\hat{H}_{\mathbf{k}}(0,t-\tau) + \int_0^\infty d\sigma e^{iz\sigma}\hat{H}_{\mathbf{k}}(\sigma,t-\tau)\frac{\partial}{\partial\sigma}\rho^0_{1,-1}(\alpha,1;\sigma). \quad (4.5)$$

If the transient contributions to $\rho^0_{1,-1}(\alpha,1;\sigma)$ are neglected, the second term in the last equation vanishes. Using also (3.10) and the well-known Debye form for $G^0_{1,-1}(\alpha,\beta;0)$, we obtain the final result

$$(2\pi)^{-1}\int_{c'} dz' G^0_{1,-1}(\alpha,1;z-z')H_{\mathbf{k}}(z';t-\tau)$$
$$\simeq -\frac{e^2\beta}{2\pi^2}\frac{1}{k^2+\kappa^2}\varphi^0(\alpha)\varphi^0(1)\int d\mathbf{v}_2\rho_{\mathbf{k}}(2;t-\tau). \quad (4.6)$$

In a similar way, the other type of convolution integral appearing in (3.16) yields

$$(2\pi)^{-1}\int_{c'} dz' F^0_1(1;z-z')R_{\mathbf{k}}(\alpha;z',t-\tau) \simeq -\frac{e^2\beta}{2\pi^2}\frac{1}{k^2+\kappa^2}\varphi^0(1)\rho_{\mathbf{k}}(\alpha;t-\tau). \quad (4.7)$$

In these formulas, κ is the inverse Debye length,

$$\kappa^2 = 4\pi e^2 c\beta. \quad (4.8)$$

Substituting these results in (3.15) and (3.16) and

rearranging the terms, we finally obtain the equation

$$\epsilon^0_{-1}\left[\frac{z-(\mathbf{k}+\mathbf{l})\cdot\mathbf{v}_\alpha}{l}\right]\mathfrak{F}_{\mathbf{k}+\mathbf{l},-1}(\alpha;z,t-\tau)$$
$$-d^0_{\mathbf{k}+\mathbf{l}}(\alpha)\int d\mathbf{v}_1\frac{\mathfrak{F}_{-1,\mathbf{k}+\mathbf{l}}(1;z,t-\tau)}{(\mathbf{k}+\mathbf{l})\cdot\mathbf{v}_\alpha-\mathbf{l}\cdot\mathbf{v}_1-z}$$
$$=q^0_{\mathbf{k}+\mathbf{l},-1}(\alpha;z,t-\tau), \tag{4.9}$$

where

$$q^0_{\mathbf{k}+\mathbf{l},-1}(\alpha;z)=\int d\mathbf{v}_1\frac{1}{(\mathbf{k}+\mathbf{l})\cdot\mathbf{v}_\alpha-\mathbf{l}\cdot\mathbf{v}_1-z}$$
$$\times\left\{\frac{e^2}{2\pi^2 m}[-l^{-2}\mathbf{l}\cdot\partial_1\varphi^0(1)\rho_{\mathbf{k}}(\alpha;t-\tau)\right.$$
$$+|\mathbf{k}+\mathbf{l}|^{-2}(\mathbf{k}+\mathbf{l})\cdot\partial_\alpha\varphi^0(\alpha)\rho_{\mathbf{k}}(1;t-\tau)$$
$$-(|\mathbf{k}+\mathbf{l}|^2+\kappa^2)^{-1}(\mathbf{k}+\mathbf{l})\cdot\partial_1\rho_{\mathbf{k}}(1;t-\tau)\varphi^0(\alpha)$$
$$+(l^2+\kappa^2)^{-1}\mathbf{l}\cdot\partial_\alpha\rho_{\mathbf{k}}(\alpha;t-\tau)\varphi^0(1)]$$
$$-\frac{\kappa^2}{k^2}\left[\frac{\mathbf{k}\cdot\partial_1}{|\mathbf{k}+\mathbf{l}|^2+\kappa^2}+\frac{\mathbf{k}\cdot\partial_\alpha}{l^2+\kappa^2}\right]$$
$$\left.\times\varphi^0(\alpha)\varphi^0(1)\int d\mathbf{v}_2\rho_{\mathbf{k}}(z;t-\tau)\right\}. \tag{4.10}$$

Consider finally the function $\bar{\mathfrak{F}}_{\mathbf{k}+\mathbf{l},-1}(\alpha;z)$ obtained by completing the partial Laplace transform,

$$\bar{\mathfrak{F}}_{\mathbf{k}+\mathbf{l},-1}(\alpha;z)=\int_0^\infty d\tau e^{iz\tau}\mathfrak{F}_{\mathbf{k}+\mathbf{l},-1}(\alpha;z,\tau). \tag{4.11}$$

As the coefficients of $\mathfrak{F}$ in (4.9) are time-independent, the latter equation is readily Laplace transformed into an equation for $\bar{\mathfrak{F}}$. As, moreover, in each term of (4.10) the only time-dependent factor is $\rho_{\mathbf{k}}$, it is immediately seen that the resulting equation is identical with (4.10), except that each factor $\rho_{\mathbf{k}}(\alpha;t-\tau)$ is replaced by its Laplace transform. This equation is identical with Guernsey's equation (24) of Ref. 13.[19]

5. SOLUTION OF THE INTEGRAL EQUATION FOR $\mathfrak{F}_{\mathbf{k}+\mathbf{l},-1}(\alpha;z)$ IN THE STABLE CASE

As a result of the discussion of the previous section, the only difference between our equation (3.15) and Guernsey's corresponding equation [Eq. (21) of Ref. 13] is in the values of the coefficients $\epsilon_{\mathbf{k}}$, $d_{\mathbf{k}}$, $q_{\mathbf{k},\mathbf{k}'}$. But the validity of Guernsey's method of solution depends only on some very general properties of the coefficients. More precisely, the method depends crucially on the property that $\epsilon_{\mathbf{k}}(z/k)$, as a function of the complex variable z, has no zeros in the upper half-plane. This is the well-known stability condition of the plasma (see, e.g., Refs. 2 and 6). The connection between the integral equations of the type (3.15) and the stability condition has been extensively discussed in Refs. 2 and 7; this discussion is not repeated here. Assuming therefore that the plasma is stable, Guernsey's method of solution can be transposed immediately to our case.

The method depends crucially on the fact that the kernel of the integral equation involves the vector variables $\mathbf{v}_\alpha$, $\mathbf{v}_1$ only in the scalar combinations $(\mathbf{k}+\mathbf{l})\cdot\mathbf{v}_\alpha$ and $\mathbf{l}\cdot\mathbf{v}_1$. Let us then define barred quantities as follows:

$$\bar{f}_{\mathbf{k}}(\nu)=\int d\mathbf{v}\,\delta(\nu-\mathbf{k}\cdot\mathbf{v}/k)f_{\mathbf{k}}(\mathbf{v}), \tag{5.1}$$

$$\bar{g}_{\mathbf{k},\mathbf{k}'}(\nu)=\int d\mathbf{v}\,\delta\left(\nu-\frac{\mathbf{k}\cdot\mathbf{v}}{k}\right)g_{\mathbf{k},\mathbf{k}'}(\mathbf{v}). \tag{5.2}$$

Let us insist on the peculiarities of these definitions. The barred functions associated with functions depending on a single wave vector $\mathbf{k}$ are defined in the same way as in previous works on homogeneous systems.[1-4] In quantities depending on two wave vectors, the barring is performed, by definition, with respect to the first of these.[13]

Noting also that the coefficient of $\mathfrak{F}_{\mathbf{k}+\mathbf{l},-1}(\alpha;z)$ in Eq. (3.15) only depends on $(\mathbf{k}+\mathbf{l})\cdot\mathbf{v}_\alpha$, we can multiply this equation throughout by

$$\delta[\nu_\alpha-(\mathbf{k}+\mathbf{l})\cdot\mathbf{v}_\alpha/|\mathbf{k}+\mathbf{l}|]$$

and integrate over $\mathbf{v}_\alpha$; we thus obtain an equation for the barred function $\bar{\mathfrak{F}}_{\mathbf{k}+\mathbf{l},-1}(\nu_\alpha;z)$. Moreover, *in the stable case*, we let z approach the real axis from above and obtain a singular integral equation, which is equivalent to (3.11):

$$\bar{\epsilon}_1\left[\frac{-\omega+|\mathbf{k}+\mathbf{l}|\,\nu_\alpha}{l}\right]\bar{\mathfrak{F}}_{\mathbf{k}+\mathbf{l},-1}(\nu_\alpha;\omega)$$
$$-i\frac{|\mathbf{k}+\mathbf{l}|}{l}\epsilon^{(2)}_{\mathbf{k}+\mathbf{l}}(\nu_\alpha)\int d\nu_1$$
$$\times\delta_-\left[\frac{|\mathbf{k}+\mathbf{l}|\,\nu_\alpha-\omega}{l}-\nu_1\right]\bar{\mathfrak{F}}_{-1,\mathbf{k}+\mathbf{l}}(-\nu_1;\omega)$$
$$=\bar{q}_{\mathbf{k}+\mathbf{l},-1}(\nu_\alpha;\omega). \tag{5.3}$$

In the derivation of this equation, we used the well-known limiting formula[2]

$$\lim_{\epsilon\to+0}\frac{1}{\omega\pm i\epsilon}=\mp\pi i\delta_\pm(\omega)$$
$$\equiv\mp\pi i\left[\delta(\omega)\pm\frac{i}{\pi}P\left(\frac{1}{\omega}\right)\right]. \tag{5.4}$$

[19] The only difference is the occurrence of a term containing the initial value of the correlation function in Guernsey's equation. In our case such a term would appear through the destruction fragments; these are however neglected as being of higher order (see the discussion in Sec. 2).

The function ϵ_1^- is obtained from ϵ_1 through this limiting process. Finally we used the identity $\bar{d}_{\mathbf{k}}(\nu) = (k/\pi)\epsilon_{\mathbf{k}}^{(2)}(\nu)$, where $\epsilon_{\mathbf{k}}^{(2)}$ is the imaginary part of $\epsilon_{\mathbf{k}}^-$.

From here on, the method of solution is quite similar to Guernsey's.[13] The calculations will therefore not be repeated in detail here, and only the final result, i.e., the solution of Eq. (3.15), will be quoted:

$$\mathfrak{F}_{\mathbf{k}+1,-1}(\mathbf{v}_\alpha; \omega, t-\tau) = \frac{q_{\mathbf{k}+1,-1}(\mathbf{v}_\alpha; \omega, t-\tau)}{\epsilon_1^-\left(\frac{|\mathbf{k}+1|\,\nu_1 - \omega}{l}\right)}$$

$$+ \frac{2i}{|\mathbf{k}+1|} d_{\mathbf{k}+1}(\mathbf{v}_\alpha; t-\tau) \int d\nu_1\, \delta_-(\nu_\alpha - \nu_1)$$

$$\times \frac{\Phi_2^+(\nu_1; \omega, t-\tau) - \Phi_1^-(\nu_1; \omega; t-\tau)}{\epsilon_{\mathbf{k}+1}^+(\nu_1; t-\tau)\epsilon_1^-((|\mathbf{k}+1|\,\nu_1 - \omega)/l; t-\tau)}, \quad (5.5)$$

with

$$\Phi_2^+(\nu; \omega, t-\tau)$$

$$= \frac{1}{2}\int d\nu_1\, \delta_+(\nu - \nu_1)\bar{q}_{\mathbf{k}+1,-1}(\nu_1; \omega, t-\tau), \quad (5.6)$$

$$\Phi_1^-(\nu; \omega, t-\tau) = -\frac{1}{2}\frac{|\mathbf{k}+1|}{l}\int d\nu_1$$

$$\times \delta_-\left(\frac{|\mathbf{k}+1|\,\nu - \omega}{l} - \nu_1\right)\bar{q}_{-1,\mathbf{k}+1}(-\nu_1; \omega, t-\tau). \quad (5.7)$$

Substitution of Eq. (5.5) into (3.1) yields the kinetic equation for $\rho_{\mathbf{k}}(\alpha; t)$. It is a very complicated integro-differential equation, of non-Markoffian character, which is valid even for very short times (within the ring approximation). The general form of Eq. (5.5) is formally analogous to Guernsey's solution (Eq. 61 of Ref. 13). There is however a fundamental difference with Guernsey's equation: *all the coefficients are time-dependent.* The kinetic equation is linear in $\rho_{\mathbf{k}}$, the coefficients are however extremely complex functionals of $\varphi(\alpha; t)$, i.e., the velocity distribution. The time dependence of this function is determined by solving the corresponding kinetic equation for a homogeneous system (Eq. A8.37 of Ref. 2). Indeed, the evolution of $\varphi(\alpha; t)$ is the same in an inhomogeneous system as in the uniform system. The procedure to follow for finding the one-particle distribution is therefore the following: solve the homogeneous kinetic equation which is a *closed nonlinear equation* for $\varphi(\alpha; t)$; substitute the result into (5.5) and then solve the resulting *closed linear equation* for $\rho_{\mathbf{k}}(\alpha; t)$. The problem is therefore quite well defined mathematically. Of course, the practical finding of an exact solution to this couple of equations is beyond our present possibilities!

Guernsey's equation appears as a particularly simple case of the general system of equations (5.5) and (A8.37) of Ref. 2. By assuming that the velocity distribution is Maxwellian, and hence time-independent, it no longer becomes necessary to solve the homogeneous equation; we are left with a *linear* integro-differential equation with time-independent coefficients. The most difficult step in the procedure described above is avoided. Unfortunately, even the remaining equation is still too complicated to be solved exactly.

From our previous experience with the homogeneous case, we might hope that the kinetic equation (5.5) could be enormously simplified in the case where we are interested only in the long-time behavior of the equation (i.e., in times much longer than ω_p^{-1}). Let us therefore consider a stable system and let us assume that, during the effective duration of memory of the system, the variation of the quantities $\rho_{\mathbf{k}}(\alpha; t)$ and $\varphi(\alpha; t)$ is small. We can then replace the function $\mathfrak{F}_{\mathbf{k}+1,-1}(\alpha; z, t-\tau)$ by $\mathfrak{F}_{\mathbf{k}+1,-1}(\alpha; z, t)$ [i.e., replace in all the functions $\rho_{\mathbf{k}}$ and φ appearing in (5.5) the argument $t-\tau$ by t]. A more detailed discussion of this approximation can be found (e.g., in Refs. 2 and 15). We can then perform the integration over τ explicitly in Eq. (3.1), with the result

$$D_t\rho_{\mathbf{k}}(\alpha; t) = -\omega_p^2 \int_c dz \frac{e^{-izt}}{z}$$

$$\times \int d\mathbf{l}\, \frac{\omega_p^2}{l^2}\, i\mathbf{l}\cdot\partial_\alpha \mathfrak{F}_{\mathbf{k}+1,-1}(\alpha; z, t). \quad (5.8)$$

We are therefore left with a Markoffian equation. If moreover the poles of the function $\mathfrak{F}_{\mathbf{k}+1,-1}(\alpha; z, t)$ lie well below the real axis, at an average distance of order ω_p (i.e., if the plasma is sufficiently stable; see Ref. 7), we can limit ourselves to the residue of the integrand at the pole $z = 0$. All other contributions are damped out in a time of order ω_p^{-1}. Equation (5.8) then becomes simply

$$D_t\rho_{\mathbf{k}}(\alpha; t)$$

$$= -\omega_p^2 \int d\mathbf{l} l^{-2} i\mathbf{l}\cdot\partial_\alpha \mathfrak{F}_{\mathbf{k}+1,-1}(\alpha; 0, t). \quad (5.9)$$

If we want to evaluate $\mathfrak{F}_{\mathbf{k}+1,-1}(\alpha; 0, t)$ explicitly, we note that this is a linear functional of $q_{\mathbf{k},\mathbf{k}'}(\alpha; 0, t)$. The latter can be computed from Eq. (3.16). It is immediately seen from the latter equation that, by taking the values of this function in $z = 0$, we are still left with a convolution integral over z', which cannot be evaluated without further information about the function $\varphi(\alpha; t)$.

It follows, in particular, that for the determination

of the asymptotic form of the inhomogeneous correlation function $\mathfrak{F}_{k+1,-1}(\alpha; 0, t)$ a knowledge of the homogeneous correlation functions $G_{1,-1}(\alpha, 1; \omega, t - \tau)$ and $F_1(\alpha; \omega, t - \tau)$ for *all real values of the frequency* ω is in general necessary. This rather unexpected difficulty shows that much care must be taken when one tries to take over the results of the homogeneous case to the inhomogeneous systems. *The non-Markoffian character of the equation leaves a trace even in the long-time equation.* Only in Guernsey's situation, i.e., near equilibrium, can one get rid of the convolution integral.

6. KINETIC EQUATION IN THE UNSTABLE CASE

We conclude this paper with a few remarks concerning the unstable plasma. One of us has shown[2,7] that if the dielectric constant has a zero $\zeta_+ \equiv w_0 + i\gamma_0$, $(\gamma_0 > 0)$ in the upper half-plane, the method of solution of integral equations of the type (3.15) must be modified. The arguments presented in Ref. 7 for the homogeneous case can be extended immediately, without any extra difficulty, to the inhomogeneous equation (3.15); we therefore give just a few indications.

The crucial remark is that Eq. (3.15) can no longer be reduced to Eq. (5.3) by simply letting z become real; the two equations are not analytical continuations of each other. Therefore, the solution (5.5) fails for an unstable plasma.

It has been shown however in Ref. 7 that, if the solution (5.5) of (3.15) is known for the stable case, one can extend it analytically for ω in the upper half-plane. It has then been shown that this function is also a solution of (3.15) for the unstable case, provided Im $\omega > \gamma_0$. It is convenient to let ω tend towards the real axis, by performing a correct analytical continuation. It then appears that there are extra terms added to Eq. (5.5); these terms turn out to have a stabilizing action.[7,20]

ACKNOWLEDGMENTS

We wish to thank Prof. I. Prigogine for the constant interest he took in this work.

One of us (R. B.) wishes to acknowledge the partial financial support of the United States Air Force under Contract No AF 61 (052)-179, monitored by the European Office, Office of Aerospace Research. The other author (A. K) acknowledges the financial support of the Belgian Fonds de la Recherche Scientifique Fondamentale Collective.

[20] B. Abraham (to be published).

Note on the Evaluation of Some Fermi Integrals*

M. L. GLASSER
Battelle Memorial Institute, Columbus, Ohio
(Received 28 February 1964)

A method is illustrated for evaluating Fermi integrals by means of an integral representation for the derivative of the Fermi function. Several examples from free-electron diamagnetism are discussed. A convenient representation for the free energy of a fermi gas is also derived.

I. INTRODUCTION

INTEGRALS of the form

$$\int_0^\infty f(E)\,\frac{\partial f_0}{\partial E}\,dE, \tag{1}$$

where $f_0(E) = [1 + \exp\gamma(E - \zeta)]^{-1}$, $\gamma = 1/kT$, and ζ is the Fermi energy are frequently encountered in the study of thermal and magnetic properties of metals.[1] A standard method for estimating these integrals is to make the change of variable $\eta = \gamma(E - \zeta)$ and replace the lower limit, $-\zeta\gamma$, of the transformed integral by $-\infty$. The integration can then be performed by closing the contour in the upper half-plane. In the case of metals with low Fermi levels, and at intermediate temperatures, it may be of interest to have the terms neglected by this procedure explicitly displayed. Furthermore if $f(E)$ involves nonintegral powers or logarithms then this procedure is incorrect in principle because it treats inconsistently the branch point at $\eta = -\gamma\zeta$.

In this note a method is outlined by which asymptotic expansions can be obtained if $f(E)$ involves roots or logarithms of E and the integrals evaluated explicitly if this is not the case. The procedure is illustrated for several integrals of the form (1) selected from the literature.

How much of this material is actually new is not clear; certainly (12) and (14) are known results. However, it is the purpose of this note to advocate (2) as a powerful and general approach to Fermi integrals. The material included is merely suggested as representative and can easily be extended to include, e.g. transport integrals, and modified so as to apply to Bose–Einstein integrals.

II. METHOD

The formula

$$\frac{\partial f_0}{\partial E} = \tfrac{1}{4}\operatorname{sech}^2\tfrac{1}{2}\gamma(E - \zeta)$$

$$= -\frac{\gamma}{2\pi i}\int_{c-i\infty}^{c+i\infty}\pi S\csc\pi S \times \exp\{-S\gamma(E - \zeta)\}\,dS, \quad -1 < c < 1 \tag{2}$$

is derived in Appendix A. In using this integral representation the factor $\pi S \csc \pi S$ introduces a series of simple poles along the real axis. It can be shown that if $f(E)$ is single-valued the approximation discussed in the introduction is that of disregarding these poles. This point is illustrated by the first example.

We consider the integral, which occurs in the theory of electron diamagnetism,[2]

$$W = \int_0^\infty \exp[i\alpha\gamma(E - \zeta)]\,\frac{\partial f_0}{\partial E}\,dE, \quad \alpha = \frac{n\pi kT}{\mu_0 H}. \tag{3}$$

Inserting (2) into (3) gives

$$W = -\frac{\exp(-i\alpha\gamma\zeta)}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{\pi z\csc\pi z}{z - i\alpha}\,e^{z\gamma\zeta}\,dz, \quad 0 < c < 1. \tag{4}$$

The integrand of (4) has simple poles at $z = i\alpha$, ± 1, ± 2, $\cdots$; closing the contour to the left by a semicircle of infinite radius we find

$$W = -\pi\alpha\,\operatorname{csch}\pi\alpha - \exp(-i\alpha\gamma\zeta)\{[1 + \exp(-\gamma\zeta)]^{-1} - F[1, i\alpha; 1 + i\alpha; -\exp(-\gamma\zeta)]\}, \tag{5}$$

where F is Gauss' hypergeometric function. The first term, which is precisely Sondheimer and Wilson's estimate obtained by the method described in the introduction, comes from the pole at $z = i\alpha$; the remaining terms, of order $e^{-\gamma\zeta}$, are due to the poles at $z = -1, -2, \cdots$.

As a second example we consider an integral from Stephen's work on magnetic shielding,[3]

$$S_1 = \int_0^\infty dE\,E^{\frac{1}{2}}\exp(i\alpha\gamma E)\,\frac{\partial f_0}{\partial E}\,dE. \tag{6}$$

* This study was supported in part by Air Force Office of Scientific Research Grant No. AF-AFOSR-260-63.

[1] A. H. Wilson, *The Theory of Metals* (Cambridge University Press, Cambridge, England, 1954), 2nd ed.

[2] E. H. Sondheimer and A. H. Wilson, Proc. Roy. Soc. (London) **A210**, 173 (1951).

[3] M. J. Stephen, Phys. Rev. **123**, 126 (1961).

From his paper it appears that the lower limit in (6) should actually be $a = (\frac{3}{2} - m/m^*)\mu_0^* H$; we calculate the error, which is small, introduced in this way by subtracting from S_1 the quantity

$$S_2 = \int_0^a dE\, E^{\frac{1}{2}} \exp(i\alpha\gamma E) \frac{\partial f_0}{\partial E}. \tag{7}$$

Using (2) we obtain

$$S_1 = -\frac{\Gamma(\frac{3}{2})}{2\pi i \gamma^{\frac{1}{2}}} \int_{c-i\infty}^{c+i\infty} \frac{\pi z \csc \pi z}{(z - i\alpha)^{\frac{1}{2}}} \exp(z\gamma\zeta)\, dz, \qquad 0 < c < 1. \tag{8}$$

The integrand in (8) has a branch point at $z = i\alpha$ in addition to the poles at $z = \pm 1, \cdots$. We again can close the contour to the left by a semicircle of infinite radius, but an excursion must be made to pass on either side of the branch cut along $\operatorname{Im} z = \alpha$, $\operatorname{Re} z \leq 0$. The resulting contour integral gives the residues at the poles plus a counterclockwise loop integral about the branch cut. We transform the latter by making the change of variable $\xi = z - i\alpha$. The transformed contour, which we denote $[0, (\infty -)]$ passes counterclockwise about the negative real axis. [In evaluating the integral and residues we choose that branch of the integrand for which $(-1)^{-\frac{1}{2}} = i$.] Equation (8) becomes

$$S_1 = -\pi\Gamma(\tfrac{3}{2}) e^{i\alpha\gamma\zeta} [B_{\frac{1}{2}}(\gamma\zeta, \alpha) + i\alpha B_{\frac{1}{2}}(\gamma\zeta, \alpha)] + i\gamma^{-\frac{1}{2}}\Gamma(\tfrac{3}{2}) \sum_{n=1}^{\infty} (-1)^n (n + i\alpha)^{-\frac{1}{2}} n e^{-n\gamma\zeta}, \tag{9}$$

where $B_\sigma(\lambda, \mu)$ is the integral discussed in Appendix B. Then for large $\gamma\zeta$ we obtain

$$S_1 = \pi i \gamma^{-\frac{1}{2}} e^{i\alpha\gamma\zeta} \operatorname{csch}(\pi\alpha) \times \{i\alpha(\gamma\zeta)^{\frac{1}{2}} - (2\pi\alpha)^{-1} L(\pi\alpha)(\gamma\zeta)^{-\frac{1}{2}} + \cdots\} + i\gamma^{-\frac{1}{2}}\Gamma(\tfrac{3}{2}) \sum_{1}^{\infty} n(-1)^n (n + i\alpha)^{-\frac{1}{2}} e^{-n\gamma\zeta}. \tag{10}$$

The next term in brackets is of order $(\gamma\zeta)^{-\frac{1}{2}}$ and $L(x)$ is the Langevin function. The leading term in $-\operatorname{Im} S_1$ is

$$\pi\gamma^{-\frac{1}{2}} \operatorname{csch}(\pi\alpha)\alpha(\gamma\zeta)^{\frac{1}{2}} \sin(\alpha\gamma\zeta)$$

and differs from Stephen's result [Eq. (38) where $(\alpha = n\pi/\gamma\mu_0^* H)$] by a factor of (-2); also the next term, of order $(\gamma\zeta)^{-\frac{1}{2}}$ rather than $(\gamma\zeta)^{-1}$ as found by Stephen, cannot be obtained by extending his calculation.

The integral in (7) is more easily handled. Again using (2) and the formula[4]

$$\int_0^a x^{\frac{1}{2}} \exp[(i\alpha\gamma - \gamma z)x]\, dx = a^{\frac{1}{2}}\Gamma(\tfrac{3}{2})\gamma^*[\tfrac{3}{2}, a\gamma(z - i\alpha)], \qquad \operatorname{Re} z > 0,$$

where $\gamma^*(b, x)$ is a function, related to the incomplete gamma function, which is entire and single-valued, we find

$$S_2 = -\frac{a^{\frac{1}{2}}\gamma\Gamma(\frac{3}{2})}{2\pi i} \int_{c-i\infty}^{c+i\infty} \pi z \csc \pi z e^{-z\gamma\zeta} \times \gamma^*[\tfrac{3}{2}, -a\gamma(n + i\alpha)]. \tag{11}$$

The integrand of (11) has no singularities but the poles at $z = \pm 1, \pm 2, \cdots$. Closing the contour to the left we obtain

$$S_2 = \gamma a^{\frac{1}{2}}\Gamma(\tfrac{3}{2}) \sum_{n=1}^{\infty} (-1)^n e^{-n\gamma\zeta} \gamma^*[\tfrac{3}{2}, -a\gamma(n + i\alpha)],$$

a number of order $e^{-\gamma\zeta}$.

As a final application of (1) we develop the asymptotic expansions for the well-known Fermi integrals[5]

$$F_k(\gamma\zeta) = \int_0^\infty E^k f_0(E)\, dE \qquad k \geq 0.$$

Integrating by parts, and using (2), we find[6]

$$F_k(\gamma\zeta) = \frac{\pi\Gamma(k + 1)}{\gamma^{k+1}(2\pi i)} \int_{c-i\infty}^{c+i\infty} z^{-(k+1)} \csc \pi z \times \exp(z\gamma\zeta)\, dz, \qquad 0 < c < 1. \tag{12}$$

Using the expressions for the function $B_\sigma(\lambda, 0)$ derived in Appendix B we obtain the expansions

(1) k an integer:

$$F_k(\gamma\zeta) = \frac{\pi k!}{\gamma^{k+1}} \left\{ \lim_{z\to 0} \frac{d^{k+1}}{dz^{k+1}} [z \csc(\pi z) e^{z\gamma\zeta}] + \frac{(-1)^{k+1}}{\pi} \Phi(k + 1, e^{-\gamma\zeta}) \right\};$$

(2) k not an integer:

$$F_k(\gamma\zeta) = \frac{\pi\Gamma(k + 1)}{\gamma^{k+1}} \left\{ \frac{(\gamma\zeta)^{k+1}}{\pi\Gamma(k + 1)} + \frac{2}{\pi} \times \sum_{n=1}^{\infty} \frac{(2^{2n} - 1)\pi^{2n} B_n}{2n!\, \Gamma(k - 2n + 2)} (\gamma\zeta)^{k-2n+1} + \frac{(-1)^{k+1}}{\pi} \Phi(k + 1, -e^{-\gamma\zeta}) \right\},$$

where the B_n are Bernoulli's numbers and $\Phi(s, x) = \sum_{n=1}^{\infty} x^n/n^s$ is Lerch's zeta function.

In conclusion we can use (2) to write down a

[4] Bateman Manuscript Project, *Higher Transcendental Functions*, edited by A. Erdélyi (McGraw-Hill Book Company, Inc., New York, 1953), Vol. 2, Chap. IX.

[5] J. McDougal and E. C. Stoner, Phil. Trans. Roy. Soc. (London) **A237**, 67 (1938).

[6] R. Dingle, Appl. Sci. Res. **B6**, 225 (1956).

convenient expression for the free energy of an electron gas.

Beginning with Eq. (21) of Ref. 2,

$$F - n\zeta = 2\int_0^\infty z(E)\frac{\partial f_0}{\partial E}\,dE$$

(we assume there is no spin dependence and the sum over spins is represented by the factor 2), we have, using (2),

$$F - n\zeta = -\frac{\gamma}{\pi i}\int_{c-i\infty}^{c+i\infty} ds\pi s \csc \pi s e^{s\gamma t} \times \int_0^\infty z(E)e^{-s\gamma E}\,dE. \qquad (13)$$

However, since $z(E)$ is just the inverse Laplace transform of $u^{-2}Z(u)$, where Z is the partition function, we have as a general expression for a fermi system

$$F - n\zeta = \frac{i}{\gamma}\int_{c-i\infty}^{c+i\infty} ds\,\frac{\csc \pi s}{s} Z(\gamma s)e^{s\gamma t}, \quad 0 < c < 1. \qquad (14)$$

This relates the thermodynamic properties to the singularities of the partition function for the system. (14) has been used extensively in the Russian literature,[7] but no English reference seems to exist.

APPENDIX A

Consider the integral

$$I = \tfrac{1}{4}\int_{-\infty}^{\infty} \operatorname{sech}^2(\tfrac{1}{2}x)\exp(xz)\,dx.$$

By changing the variable to $u = e^x/(1 + e^x)$ the integral becomes a beta function

$$I = \int_0^1 du\, u^z(1-u)^{-z} = z\Gamma(z)\Gamma(1-z) = \pi z \csc \pi z.$$

I is the two-sided Laplace transform of the function I is the two-sided Laplace transform of the function $\frac{1}{4}\operatorname{sech}^2(\frac{1}{2}x)$ and converges in the strip $-1 <$ Re $z < 1$. Hence by the inversion theorem[8]

$$\tfrac{1}{4}\operatorname{sech}^2(\tfrac{1}{2}x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \pi z \csc \pi z \times \exp(-zx)\,dz, \quad -1 < c < 1.$$

APPENDIX B

Let

$$B_\sigma(\lambda, \mu) = \frac{1}{2\pi i}\int_0^{(\infty-)} \xi^{-\sigma}\csc\pi(\xi + i\mu) \times \exp(\lambda\xi)\,d\xi, \quad \lambda > 0,$$

where the contour begins at ∞ in the third quadrant, passes counterclockwise about the origin and proceeds to ∞ in the second quadrant in such a way as to avoid enclosing any of the poles $\xi = n\pi - i\mu$ if $\mu \neq 0$.

If σ is a positive integer, the origin is a σ-order pole for $\mu \neq 0$ and a pole of order $\sigma + 1$ if $\mu = 0$. In these cases the integral is simply the residue at the origin. If σ is a negative integer the integral vanishes if $\mu \neq 0$ and for $\mu = 0$ we have, summing the residues of the poles of csc $\pi\xi$ along the negative real axis,

$$B_\sigma(\lambda, 0) = (1/\pi)(d^\sigma/d\lambda^\sigma)[1 + \exp(-\lambda)]^{-1}.$$

If σ is not an integer then the origin is a branch point and we may consider the ξ plane to be cut along the negative real axis. In this case by Tauber's theorem[8] we may easily obtain an asymptotic expansion for B_σ, for large values of λ. If $\mu \neq 0$ we use the expansion

$$\csc\pi(\xi + i\mu) = -i\operatorname{csch}\pi\mu\sum_{k=1}^{\infty}\beta_k\xi^k,$$

where the first few coefficients are

$$\beta_0 = 1, \quad \beta_1 = \pi i\coth\mu\pi, \quad \beta_2 = \tfrac{1}{2}\pi^2(1 - 2\coth\pi\mu),$$
$$\beta_3 = \tfrac{1}{6}\pi^3 i\coth\pi\mu[5 + 6\coth^2\pi\mu].$$

Therefore

$$B_\sigma(\lambda, \mu \neq 0) \sim -i\operatorname{csch}\pi\mu\sum_{k=0}^{\infty}\frac{\beta_k}{2\pi i}\int_0^{(\infty-)}\xi^{(k-\sigma)}\exp(\lambda\xi)\,d\xi.$$

From Hankel's representation for the gamma function we find

$$\frac{1}{2\pi i}\int_0^{(\infty-)}\xi^{-\nu}\exp(\lambda\xi)\,d\xi = \frac{\lambda^{\nu-1}}{\Gamma(\nu)},$$

where the branch of the integrand, $\xi^{-\nu} = e^{-\nu\ln\xi}$, ln ξ real for $\xi > 0$, is chosen. Finally,

$$B_\sigma(\lambda, \mu \neq 0) \sim -i\operatorname{csch}\pi\mu\sum_{k=0}^{\infty}\beta_k\frac{\lambda^{\sigma-k-1}}{\Gamma(\sigma-k)}.$$

When $\mu = 0$ we use the expansion

$$\csc\pi\xi = \frac{1}{\pi\xi} + \frac{2}{\pi}\sum_{n=1}^{\infty}[(2^{2n-1}-1)B_n\pi^{2n}/(2n)!]\xi^{2n-1},$$

where the B_n are Bernoulli numbers. Thus

$$B_\sigma(\lambda, 0) \sim \frac{\lambda^\sigma}{\pi\Gamma(\sigma+1)} + \frac{2}{\pi} \times \sum_{n=1}^{\infty}\frac{[(2^{2n-1}-1)B_n\pi^{2n}/(2n)!]}{\Gamma(\sigma-2n+1)}\lambda^{\sigma-2n}.$$

[7] A. G. Samoilovich, *Thermodynamics and Statistical Physics* (in Russian) (State Press for Technical and Theoretical Literature, Moscow, 1955).

[8] B. van der Pol and H. Bremmer, *Operational Calculus* (Cambridge University Press, Cambridge, England, 1955).

Equation of State for a Gas with a Weak, Long-Range Positive Potential

DAVID S. NEWMAN
Department of Mathematics, University of Washington, Seattle, Washington
(Received 11 February 1964)

A one-dimensional fluid model in which the pair interaction potential is exponential and repulsive is considered, and the equation of state in the "long-range" limit is determined exactly. This model is complementary to one studied by Kac, in which the pair potential consists of a hard core and an exponential, attractive tail. In that model a phase transition occurs, but in the current model there is no phase transition. This fact lends support to a conjecture of Ruelle that no phase transition occurs if the potential is bounded. The possibility of applying the method of Kac to a wider class of potentials is suggested, and some of the mathematical difficulties yet to be overcome are outlined.

I. INTRODUCTION

THE purpose of this paper is to investigate the behavior of a one-dimensional fluid model which is complementary to the one studied by Kac[1], and Kac, Uhlenbeck, and Hemmer.[2,3] These models are of great interest to the theory of condensation, because the partition function and the equation of state can be treated rigorously, and exact results are obtained.

In both cases we consider a system of N particles on a line of length L, and allow L and N to $\to \infty$ while $l = L/N$ remains finite. In the Kac model, the pair interaction potential consists of a hard core of radius δ and an exponential attraction. The potential considered here is a pure exponential repulsion: $\phi(x) = +\alpha \exp(-\gamma|x|)$, $|x| \geq 0$, $\alpha, \gamma > 0$. No phase transition occurs in this case, nor does one occur in the long-range limit in which $\alpha = \alpha_0\gamma$ and γ is allowed to approach zero. This situation is in contrast to that of the Kac model in which a first-order phase transition does occur in this limit. The equation of state in this limiting case is

$$p = kT/l + \alpha_0/l^2 \tag{1}$$

which is to be compared with the van der Waals' equation occurring in the Kac model. As one might expect, the pressure is higher for given temperature and specific volume than in the ideal gas. The equation of state is derived in Sec. III.

The results proceed from an integral equation similar to that for the Kac model, but with the unfortunate difference that the kernel is not positive-definite nor even Hermitian. Complex eigenvalues may arise, but there is reason to believe that this is not the case.

The Kac integral equation is discussed in Sec. II, and some of the problems involved in obtaining deeper results from it (such as the two- and three-particle distribution functions) are considered in Sec. IV.

II. THE KAC INTEGRAL EQUATION

The partition function for this one-dimensional model is

$$\begin{aligned} Q(L, N) &= \frac{1}{\Lambda^N}\frac{1}{N!}\int_0^L \cdots \int_0^L dt_1 \cdots dt_N \\ &\quad \times \exp\left[-\frac{1}{kT}\sum_{i<j}\phi(|t_i - t_j|)\right] \\ &= \frac{1}{\Lambda^N}\frac{1}{N!}\int_0^L \cdots \int_0^L dt_1 \cdots dt_N \\ &\quad \times \exp\left[-a^2\sum_{i<j} e^{-\gamma|t_i - t_j|}\right], \end{aligned} \tag{2}$$

where we have set $\Lambda^2 = h^2/2\pi mkT$ and $a^2 = \alpha/kT$.

The derivation of the Kac integral equation is virtually the same as in Ref. 1 or 2, and it is not reproduced here. We simply list the results:

The integral equation is

$$\int_{-\infty}^{\infty} K(x, y)\psi(y)\, dy = \kappa\psi(x), \tag{3}$$

where

$$K(x, y) = \frac{W(x)p_s(x \mid y)}{[W(x)W(y)]^{\frac{1}{2}}} \exp\left[ia\left(\frac{x+y}{2}\right)\right]. \tag{4}$$

The factors in the kernel K are

$$\begin{aligned} W(x) &= (2\pi)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}x^2), \\ P(x \mid y, t) &= [2\pi(1 - e^{-2\gamma t})]^{-\frac{1}{2}} \\ &\quad \times \exp[-(y - xe^{-\gamma t})^2/2(1 - e^{-2\gamma t})], \end{aligned} \tag{5}$$

[1] M. Kac, Phys. Fluids 2, 8 (1959).
[2] M. Kac, S. E. Uhlenbeck, and P. C. Hemmer, J. Math. Phys. 4, 216 (1963).
[3] G. E. Uhlenbeck, P. C. Hemmer, and M. Kac, J. Math. Phys. 4, 229 (1963).

$$p_s(x \mid y) = \int_0^\infty dt\, e^{-st} P(x \mid y, t).$$

Then the Laplace transform of the partition function is expressible in terms of the kernel of (4):

$$\int_0^\infty dL e^{-sL} Q(L, N)$$

$$= \frac{\exp(\frac{1}{2}Na^2)}{\Lambda^N s^2} \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty dx_1 \cdots dx_N$$

$$\times e^{\frac{1}{2}ia(x_1+x_n)} [W(x_1)W(x_n)]^{\frac{1}{2}} \prod_{i=1}^{n-1} K(x_i, x_{i+1}). \qquad (6)$$

The kernel K is of type L_2 as can be readily verified; that is

$$\int_{-\infty}^\infty \int_{-\infty}^\infty |K(x, y)|^2\, dx\, dy < \infty.$$

Therefore K has at least one, and at most countably many eigenfunctions $\psi_i(x)$ and corresponding eigenvalues κ_i; and $|\kappa_0| \geq |\kappa_1| \geq \cdots$, $\kappa_n \to 0$ as $n \to \infty$. $K(x, y)$ is algebraically symmetric but not Hermitian; i.e., $K(x, y) = K(y, x)$, but $K^*(x, y) = \overline{K(y, x)} \neq K(x, y)$. Neither does K commute with its adjoint (i.e. it is not normal), and although it is symmetrized by the operator S such that $SK(x, y) = K(-x, y)$, S is not positive-definite. It is not clear, therefore, whether K has a simple expansion in terms of its own eigenfunctions ψ_i which can be conveniently iterated, or not. This difficulty precludes any consideration of the pair distribution function (as in Ref. 3) at the present time. Further remarks on the eigenfunction problem are in the Appendix.

To find the equation of state (1), it is useful to introduce the grand partition function

$$G(L, z) = \sum_{N=1}^\infty (\Lambda z)^N Q(L, N),$$

where z is the fugacity. In terms of K, the Laplace transform of $G(L, z)$ is

$$\int_0^\infty dL\, e^{-sL} G(L, z)$$

$$= \frac{1}{s^2} \sum_{N=1}^\infty \zeta^N \int_{-\infty}^\infty \int p(x) K^{(N)}(x, y) p(y)\, dx\, dy, \qquad (7)$$

where $K^{(N)}$ is the Nth iterate of the Kernel K, $p(x) = [W(x)e^{iax}]^{\frac{1}{2}}$, and $\zeta = z \exp(\frac{1}{2}a^2)$. The right-hand side of (7) is simply the Neumann–Liouville series solution for the integral equation

$$F(x) - \frac{1}{s} p(x) = \zeta \int_{-\infty}^\infty K(x, y) F(y)\, dy, \qquad (8)$$

multiplied by $p(x)/s$ and integrated with respect to x.

Now the abscissa of convergence of the Laplace transform of $G(L, z)$ is that value of s for which, when ζ is fixed and s is allowed to decrease toward zero, the series in (7) fails to converge. This occurs when s has the value for which $1/\kappa_0(s) = \zeta$, where $\kappa_0(s)$ is an eigenvalue of largest absolute value of the Kac equation (3). But this abscissa is also

$$s = \lim_{L \to \infty} \frac{1}{L} \log G(L, z) = p/kT,$$

so that $\kappa_0(p/kT) = 1/[z \exp(\frac{1}{2}a^2)]$. Finally, the equation of state is obtained from a knowledge of κ_0 by the relation between the fugacity z and the Gibbs potential per particle μ:

$$\mu = kT \log \Lambda z$$

and

$$l = (\partial\mu/\partial p)_T = -\kappa_0'(s)/\kappa_0(s), \qquad s = p/kT. \qquad (9)$$

III. THE EQUATION OF STATE IN THE LONG-RANGE LIMIT

We now consider a method for obtaining the equation of state (1) in the long-range limit discussed in the introduction. Let the Laplace transform of $G(L, z)$ be denoted by $\hat{G}(s, \zeta)$. This function is meromorphic in ζ, and has poles at $\kappa_0^{-1}(s)$, $\kappa_1^{-1}(s)$, $\cdots$, $\kappa_p^{-1}(s)$, $\cdots$, where the κ_p are the eigenvalues of the Kac equation (3). Instead of finding κ_0 directly from (7) it is more convenient to consider the traces of the iterates of K, namely

$$\sigma_N = \int_{-\infty}^\infty K^{(N)}(x, x)\, dx$$

$$= \int_{-\infty}^\infty \cdots \int dx_1 \cdots dx_N \prod_{i=1}^N K(x_i, x_{i+1})$$

with $x_{N+1} = x_1$. Now the logarithmic derivative of the Fredholm determinant $d(\zeta)$ can be expanded in a Taylor series about $\zeta = 0$, and the coefficient of ζ^N is simply $-\sigma_{N+1}$:[4]

$$\frac{d'(\zeta)}{d(\zeta)} = -\sum_{N=0}^\infty \sigma_{N+1} \zeta^N.$$

Since the roots of $d(\zeta)$ are the reciprocals of the eigenvalues κ_p, $d'(\zeta)/d(\zeta)$ is a meromorphic func-

[4] For example, see F. Smithies, *Integral Equations*, Cambridge Tracts No. 49 (Cambridge University Press, Cambridge, England, 1958), Chap. 6.

tion of ζ with simple poles at κ_0^{-1}, κ_1^{-1}, $\cdots$. Therefore the radius of convergence of this Taylor series in ζ is

$$\underline{\lim} \lim_{N\to\infty} |\sigma_N|^{-1/N} = |\kappa_0(s)|^{-1}.$$

In particular, since $\sigma_N > 0$ for all N, $|\kappa_0(s)|^{-1}$ is a singular point of $d'(\zeta)/d(\zeta)$, and therefore we may say that $\kappa_0(s)$ is positive, without loss of generality.[5] Therefore, there is an eigenvalue of maximum absolute value which is real and positive.

The integrals for σ_N can be evaluated explicitly in the limit $\alpha = \alpha_0\gamma$, $\gamma \to 0$, and the analysis is almost identical to that of Appendix II of Ref. 2. There is therefore no purpose in reproducing it here; the result is

$$\lim_{\gamma\to 0} \gamma\sigma_N = \frac{1}{2\pi} \iint_{-\infty}^{+\infty} d\xi\, d\eta \times \exp\left[-iN\eta(2\nu_0)^{\frac{1}{2}}\right][s + \tfrac{1}{2}(\xi^2 + \eta^2)]^{-N} \qquad (10)$$

which becomes, because of the symmetry of the integrand,

$$\lim_{\gamma\to 0} \gamma\sigma_N = \frac{1}{2\pi} \iint_{-\infty}^{+\infty} d\xi\, d\eta \times \cos\left[N\eta(2\nu_0)^{\frac{1}{2}}\right][s + \tfrac{1}{2}(\xi^2 + \eta^2)]^{-N}. \qquad (11)$$

Here $\nu_0 = a^2/\gamma = \alpha_0/kT$. Equation (10) should be compared to Eq. (24) of Ref. 2.

The remaining analysis of Ref. 2 does not carry through to the present case. Instead it is possible to find $\kappa_0(s)$ directly by the use of asymptotic formulas for Bessel functions.

Introducing polar coordinates in (11) by $\xi = r\cos\theta$, $\eta = r\sin\theta$,

$$\lim_{\gamma\to 0} \gamma\sigma_N = \delta_N = \frac{1}{2\pi} \int_0^\infty r\, dr \int_0^{2\pi} d\theta \times \cos\left[N(2\nu_0)^{\frac{1}{2}} r\sin\theta\right](s + \tfrac{1}{2}r^2)^{-N}.$$

By Bessel's integral formula, the integral over θ is simply $J_0(Nr(2\nu_0)^{\frac{1}{2}})$. Thus

$$\delta_N = \int_0^\infty \frac{J_0(Nr(2\nu_0)^{\frac{1}{2}})r}{(s + \frac{1}{2}r^2)^N}\, dr,$$

or by a simple change of variable ($\rho = Nr$),

$$\delta_N = (N)^{2N-2} \int_0^\infty \frac{J_0(\rho(2\nu_0)^{\frac{1}{2}})\rho\, d\rho}{(sN^2 + \frac{1}{2}\rho^2)^N}. \qquad (12)$$

[5] See E. C. Titchmarsh, *Theory of Functions* (Oxford University Press, London, 1939), 2nd ed., Chap. VIII.

There is an exact formula for this integral in Watson's treatise.[6] When applied here, the result is

$$\delta_N = [\nu_0 N^2/s]^{\frac{1}{2}(N-1)} \times [2/(N-1)!]K_{N-1}[2N(\nu_0 s)^{\frac{1}{2}}], \qquad (13)$$

where K_{N-1} is the Bessel function of pure imaginary argument of the second kind. Note that both the order and the argument are large; there is only one asymptotic expansion for $K_N(x)$ with large order and argument, as opposed to three for $J_N(x)$. This fact is directly related to the nonoccurrence of a phase transition in this model. The formula, due to Nicholson,[7] is

$$K_N(x) = \pi(2\pi x\cosh\beta_N)^{-\frac{1}{2}} \times \exp\left[-x(\cosh\beta_N - \beta_N\sinh\beta_N)\right]$$

where $N = x\sinh\beta_N$.

Thus if $\operatorname{csch}\beta_N = [(N+1)/N]\,2(\nu_0 s)^{\frac{1}{2}}$,

$$K_N(N\operatorname{csch}\beta_N) = \pi(2\pi N\coth\beta_N)^{-\frac{1}{2}} \times \exp\left[-N(\coth\beta_N - \beta_N)\right]. \qquad (14)$$

Substituting (14) into (13) and passing to the limit as $N \to \infty$,

$$\kappa_0(s) = \lim_{N\to\infty} (\delta_N)^{1/N} = \left[\frac{\nu_0}{2s}\right]^{\frac{1}{2}} \exp(1 + \beta - \coth\beta),$$

$\operatorname{csch}\beta = 2(\nu_0 s)^{\frac{1}{2}}$. Thus

$$\log\kappa_0(s) = -\tfrac{1}{2}\log s + 1 + \tfrac{1}{2}\log\tfrac{1}{2}\nu_0 + \sinh^{-1}(4\nu_0 s)^{-\frac{1}{2}} - (4\nu_0 s + 1)^{\frac{1}{2}},$$

and

$$\kappa_0'(s)/\kappa_0(s) = (1/2s)[1 + (1 + 4\nu_0 s)^{\frac{1}{2}}]. \qquad (15)$$

Introducing the thermodynamic variables

$$s = p/kT, \quad \nu_0 = \alpha_0/kT, \quad l = -\kappa_0'(s)/\kappa_0(s),$$

we obtain the equation of state (1) after first solving (15) for s.

IV. GENERAL OBSERVATIONS

(A) Ruelle[8] has proved a theorem of the van Hove type[9] concerning the existence of the free energy in the thermodynamic limit for a certain class of pair interaction potentials $\phi(x)$. He also

[6] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge University Press, Cambridge, England, 1922), Sec. 13.51.

[7] J. W. Nicholson, Phil. Mag. Ser. 6, **20**, 938–943 (1910).

[8] D. Ruelle, Helv. Phys. Acta. **36**, 183 (1963).

[9] L. van Hove, Physica **15**, 951 (1949); and **16**, 137 (1950).

proves that the pressure is a continuous decreasing function of specific volume for the case of a bounded potential. In both cases he requires the additional assumption that the potential is nonpositive for $|x| \geq R$ for some R in order to carry out the proof; but he speculates that this assumption is not essential. The results of Sec. III above add support to that speculation.

(B) It should be possible to treat now a potential of the form

$$\begin{aligned} &\phi(t) = +\infty, \qquad 0 \leq |t| < \delta, \\ &\phi(t) = \sum_{i=1}^{m} \alpha_i \gamma e^{-\gamma\beta_i|t|} - \sum_{i=1}^{m'} \alpha_i' \gamma e^{-\gamma\beta_i'|t|}, \qquad (16) \\ &|t| > \delta, \qquad \alpha_i, \alpha_i', \beta_i, \beta_i' > 0, \end{aligned}$$

at least in the limiting case $\gamma \to 0$. There remains the problem of harmonizing the different approaches taken for the two types of potentials, positive and negative. If the limit procedure of Sec. III is applied to a negative potential with hard core, the very unpleasant integral

$$\delta_N = e^{-N\epsilon\delta} N^{2N-2} \int_0^\infty \frac{\rho J_0(i\rho(2\nu_0)^{\frac{1}{2}}) \exp(-\delta\rho^2/2N)\, d\rho}{(sN^2 + \rho^2)^N}$$

replaces its counterpart (12); the added exponential in the integrand complicates an already delicate convergence as $N \to \infty$. On the other hand the methods of Kac, Uhlenbeck, and Hemmer[2] cannot be applied to the positive potential considered here. This difficulty might be resolved if more could be said about the eigenvalues and eigenfunctions of the kernel $K(x, y)$ in the positive potential case. The known properties are listed below in the Appendix.

Instead of allowing a hard core in the potential (16), one might instead consider potentials of the form

$$\phi(t) = \sum_{i=1}^{\infty} \alpha_i \gamma e^{-\gamma\beta_i t} - \sum_{i=1}^{\infty} \alpha_i' \gamma e^{-\gamma\beta_i' t},$$

$$0 < t < \infty; \qquad \lim_{t \to +0} \phi(t) = +\infty.$$

Such a "Dirichlet series" potential is appealing in that it seems more "natural" than the hard core, but to apply the entire Kac procedure to it would involve kernel $K(x, y)$ in which x, y are vectors in m space, and $m \to \infty$.

APPENDIX

The kernel $K(x, y)$ of Eq. (4) has many interesting symmetries which suggest it may have a simple expansion in terms of its own eigenvectors and eigenvalues.

It is convenient to introduce operator notation, so we write Eq. (3) as

$$K\psi = \kappa\psi. \qquad (17)$$

The complex conjugate operator $\bar{K}$ has the kernel $\overline{K(x, y)}$, and $\bar{\psi}(x) = \overline{\psi(x)}$. We also introduce the symmetry operator S, such that $S\phi(x) = \phi(-x)$. Thus $SK(x, y) = K(-x, y)$. From (4) it is clear that

$$SKS = \bar{K} = K^*,$$

and it is easily seen that if (17) holds, so does

$$K(S\bar{\psi}) = \bar{\kappa}(S\bar{\psi}). \qquad (18)$$

Thus if the spectrum of K contains any nonreal eigenvalue κ, it also contains its complex conjugate $\bar{\kappa}$. Also if κ is real we may assume without loss of generality that $S\bar{\psi} = \psi$; that is, $\psi(-x) = \overline{\psi(x)}$.

The eigenfunctions of K corresponding to different eigenvalues have two properties akin to orthogonality; however, the quadratic forms associated with the analog of the scalar product are not positive definite, so that expansions in the eigenfunctions ψ_i may not be valid. We have:

(A) If $K\psi_j = \kappa_j\psi_j$, $j = 1, 2$, and $\kappa_1 \neq \bar{\kappa}_2$, then

$$(\psi_1, S\psi_2) = 0 = \int_{-\infty}^{\infty} \psi_1(x)\overline{\psi_2(-x)}\, dx.$$

In particular if κ_1 is not real, then

$$(\psi_1, S\psi_1) = \int_{-\infty}^{\infty} \psi_1(x)\overline{\psi_1(-x)}\, dx = 0.$$

(B) If $K\psi_j = \kappa_j\psi_j$, $j = 1, 2$ and $\kappa_1 \neq \kappa_2$, then

$$(\psi_1, \bar{\psi}_2) = \int_{-\infty}^{\infty} \psi_1(x)\psi_2(x)\, dx = 0.$$

If $(\psi_i, \bar{\psi}_i) \neq 0$ for all eigenfunctions, then K possesses a simple expansion of the form

$$K(x, y) = \underset{N\to\infty}{\text{l.i.m.}} \sum_{j=1}^{N} \kappa_j \psi_j(x)\psi_j(y)$$

and one could express the resolvent of the kernel in an equally simple form. It would then be reasonable to expect that the two- and three-particle distribution functions could be expanded in terms of the ψ_j and κ_j as in Ref. 3.

Finally we note that K does have an expansion[1] in terms of Hermite polynomials:

$$K(x, y) = \underset{N\to\infty}{\text{l.i.m.}} \sum_{k=1}^{N} \frac{1}{s + k\gamma} e^{i\alpha\frac{1}{2}(x+y)} \phi_k(x)\phi_k(y).$$

Here

$$\phi_k(x) = [k!\,(2\pi)^{\frac{1}{2}}]^{-\frac{1}{2}} \exp(-\tfrac{1}{4}x^2)h_k(x),$$

$$h_k(x) = (-1)^k \exp(\tfrac{1}{2}x^2)(d^k/dx^k) \exp(-\tfrac{1}{2}x^2).$$

It is interesting that this expansion is the singular value, singular function expansion[10] of $K(x, y)$; that is,

$$\exp(-\tfrac{1}{2}iax)\phi_k(x)$$

is an eigenfunction of the symmetric, positive-definite operator $\bar{K}K$ with eigenvalue $(s + k\gamma)^{-1}$, and its complex conjugate is an eigenvector of $K\bar{K}$ with the same eigenvalue.

[10] For example, see F. Smithies, Ref. 5, Chap. 8.

The appearance of Fourier-type integrals associated with Hermite functions is suggestive, but as yet no useful results have been obtained from this association.

ACKNOWLEDGMENT

The author wishes to thank the referee for pointing out that the resulting equation of state (1) can also be derived from the virial expansion if one assumes that the physical system consists of a single-fluid phase.[11]

[11] E. H. Hauge and P. C. Hemmer, Physica **29**, 1338 (1963), Sec. 4.

Microwave Surface Impedance of a Periodic Medium

GASTON FISCHER
Laboratories RCA Ltd., Zurich, Switzerland
(Received 3 March 1964)

The surface impedance of a medium is calculated, the complex but scalar conductivity of which is a periodic function in a direction parallel to the surface. Radiation is assumed to impinge vertically onto the surface. It is shown that there are two independent field configurations according to whether the induced microwave current flows in a direction parallel to the conductivity pattern or perpendicular to it. These two configurations lead to different surface impedances, the one obtaining in the perpendicular configuration being generally larger. The calculation is carried through when the conductivity is of the form $\sigma(y) = \sigma_0 + 2\sigma_1 \cos(ky)$, and a local relation between current and fields is assumed. Equations are also derived for the more general problem when the periodic conductivity is an arbitrary Fourier series and a nonlocal relation between current and fields is assumed. These problems are of interest in measurements of the microwave surface impedance of hard superconductors.

I. INTRODUCTION

THE microwave surface impedance of a metal is defined in terms of a semi-infinite medium (e.g., $z > 0$) on which a polarized electromagnetic plane wave is impinging vertically from the vacuum ($z \leq 0$). A system of electric and magnetic fields is set up in the metal in the vicinity of the surface, but because of skin effect the fields decay rapidly as one moves deeper into the metal. The surface impedance Z is given by the ratio of the electric field E at the surface $z = 0$ to the projection $J_\parallel$ of the total induced current density $\mathbf{j}$, integrated from $z = 0$ to $z = \infty$. The determination of $\mathbf{j}$ requires the knowledge of σ, the electric conductivity. A great deal of attention has been concentrated on such situations where σ is of tensorial nature,[1] or where the relation between $\mathbf{j}$ and $\mathbf{E}$ is nonlocal.[2] So far, however, all theories have assumed σ to be independent of position.

In this paper we attempt to calculate the surface impedance of a hypothetical metal, of which the scalar conductivity σ is periodic in a direction parallel to the surface. Taking this direction as y axis, we assume that $\sigma(y)$ can be expanded in a Fourier series:

$$\sigma(y) = \sum_n \sigma_n e^{inky}, \tag{1}$$

with generally complex σ_n and

$$k = 2\pi/d, \tag{2}$$

d being the period of the conductivity pattern.

We have thus defined a Cartesian coordinate system with x and y axis parallel to the metal–vacuum boundary, and positive z axis pointing into the metal. Our calculation will show that the two field configurations, where one of the two components E_x or E_y of the electric field vector $\mathbf{E}$ vanishes, are independent. We also find that the surface impedance derived for each of these two configurations is different.

We shall carry out the calculation in the limit of a local ($\mathbf{j} - \mathbf{E}$) relation and a complex conductivity $\sigma(y)$ restricted to a constant and only one oscillatory term, writing then

$$\sigma(y) = \sigma_0 + \sigma_1(e^{iky} + e^{-iky}) = \sigma_0 + 2\sigma_1 \cos(ky). \tag{3}$$

We note the even symmetry of $\sigma(y)$ as $\sigma_1 = \sigma_{-1}$. In Sec. VI, however, the nonlocal problem with a conductivity of the form (1) will be discussed and the basic equations of this much more general problem will be derived.

From an intuitive point of view, one may imagine a periodic conductivity pattern produced by the juxtaposition in parallel of alternate laminae of two different conductivities. In one field configuration ($E_y = 0$), the currents are flowing *parallel* to the laminae, and if the thickness of the laminae is not too large compared with their skin depths it would seem that the regions of high conductivity will tend to short out the regions of low conductivity. As a result the current will concentrate in the regions of low resistivity. In the other configuration ($E_x = 0$), the currents are flowing *perpendicularly* to the laminae and, because of continuity requirements, one expects that the same current is forced to flow through both regions of low and high resistivity.

[1] A. B. Pippard, Proc. Roy. Soc. (London) **A203**, 98 (1950); **A224**, 273 (1954); E. H. Sondheimer, *ibid.* **A224**, 260 (1954); G. E. Smith, Phys. Rev. 115, 1561 (1959).

[2] G. E. H. Reuter and E. H. Sondheimer, Proc. Roy. Soc. (London) **A195**, 336 (1948); R. B. Dingle, Physica 19, 311 (1953).

By this intuitive argument, then, one expects the average surface resistance to be higher in the perpendicular than in the parallel configuration. It is also clear that the two configurations are expected to yield the same average resistance when the laminae thicknesses become very large against their skin depths, since the fields at one point inside a metal cannot influence the fields at points distant by much more than one skin depth.

We should like to emphasize the coarseness of the above arguments particularly because at high frequencies the conductivity of a metal is complex. The fields in adjacent laminae, therefore, need not be in phase and it has little meaning to say, in the parallel configuration, that one region tends to short out another. It turns out, however, that the more careful analysis which follows essentially confirms the intuitive viewpoint.

The problem which we set out to analyze is of interest in measurements of the microwave surface impedance of hard superconductors[3] where the generation and parallel alignment of so-called "fluxoids" in an external dc magnetic field results in a structure with a conductivity modulated in space. Since we intend to discuss this question elsewhere[4] we do not discuss it further here.

II. FORMULATION OF THE PROBLEM

We write Maxwell's equations omitting the displacement currents, as these are negligibly small here,[5] and we assume for all field quantities a time dependence of the form $\exp(i\omega t)$;

$$\nabla \times \mathbf{B} = (4\pi/c)\mathbf{j}, \tag{4}$$

$$\nabla \times \mathbf{E} = -(i\omega/c)\mathbf{B}. \tag{5}$$

$\mathbf{j}$ is the current induced by the microwave radiation and $\mathbf{B}$ the magnetic induction. Elimination of $\mathbf{B}$ at once yields the differential equation

$$\nabla \times \nabla \times \mathbf{E} = -(4\pi i\omega/c^2)\mathbf{j}. \tag{6}$$

This is the basic equation of all surface impedance problems. Any particular problem is characterized by a second independent relation between $\mathbf{j}$ and $\mathbf{E}$ and by specific boundary conditions at the metal surface. We have already discussed our choice of relation between $\mathbf{j}$ and $\mathbf{E}$, and now turn our attention to the boundary conditions.

A. Boundary Conditions

At the boundary $z = 0$ between metal and vacuum we assume perfectly specular reflection. The calculations of Reuter and Sondheimer[2] have shown that when the mean free path l does not exceed the classical penetration depth δ_{cl} the assumption of specular or diffuse scattering is completely irrelevant. This condition strongly applies in our limit when a local $(\mathbf{j}-\mathbf{E})$ relation is valid. At the same time the assumption of specular reflection considerably simplifies the calculations. To approximate the condition of specular reflection at the surface we follow the method of Klein.[6] If the surface $z = 0$ is a perfect reflector, an electron reflected by the surface has its wave-vector components k_x and k_y unaltered, whereas its component k_z reverses sign. After reflection the electron is thus exactly in the situation it would be if it came from the other side of the surface, provided the components of electric and magnetic fields on both sides of the surface satisfied the following conditions (see Appendix II),

$$E_{\parallel}(-z) = +E_{\parallel}(+z), \quad E_z(-z) = -E_z(+z), \tag{7}$$

$$B_{\parallel}(-z) = -B_{\parallel}(+z), \quad B_z(-z) = +B_z(+z). \tag{8}$$

We may therefore replace the actual problem by one in which an infinite sample of metal is subjected to fields having the properties (7) and (8) in addition to obeying Maxwell's equations (4) and (5). Such fields may be set up by a current sheet confined to the plane $z = 0$. Instead of the impinging microwaves, it is this sheet of *applied* currents which will be regarded as the source of the *induced* fields $\mathbf{E}$ and $\mathbf{B}$ in the metal. In our equivalent problem the current $\mathbf{j}$ appearing in Eqs. (4) and (6) is therefore the sum of the *applied* and *induced* currents. Calling I the sheet current per unit width, we write for the applied current $\mathbf{j}_0$

$$\mathbf{j}_0 = \mathbf{I}\delta(z), \tag{9}$$

and for the induced current $\mathbf{j}$ we have of course

$$\mathbf{j} = \sigma(y)\mathbf{E}(y, z). \tag{10}$$

Since we are concerned only with such solutions where $\mathbf{E}$ and $\mathbf{B}$ decrease when $|z|$ increases, integration of Eqs. (4) or (6) from $z = -\infty$ to $z = +\infty$ together with conditions (7) and (8) requires that the total integrated current component $j_{\parallel}$ parallel

[3] M. Cardona, G. Fischer and B. Rosenblum, Phys. Rev. Letters **12**, 101 (1964); M. Cardona and B. Rosenblum, Phys. Letters **8**, 308 (1964); B. Rosenblum and M. Cardona, *ibid.*, **9**, 220 (1964).

[4] G. Fischer, Phys. Rev. (to be published).

[5] Displacement currents can be neglected up to frequencies of 100 Gc/sec when the magnitude of the conductivity is everywhere larger than about one $(\Omega\ \mathrm{cm})^{-1}$.

[6] O. Klein, Ark. Mat. Astr. Fys. **A31**, No. 12 (1945). Our presentation of this point is essentially taken from A. B. Pippard, Rep. Progr. Phys. **23**, 218, 219 (1960).

to the surface $z = 0$ vanish:

$$I = -\int_{-\infty}^{+\infty} j_{\parallel}(z)\, dz, \tag{11}$$

or

$$\int_{0}^{+\infty} j_{\parallel}(z)\, dz = -\tfrac{1}{2}I. \tag{12}$$

We should like to point out that in spite of having assumed a periodic medium we have not allowed the sheet current to be periodic. The correctness of this assumption can easily be proved and follows from the symmetry properties (7) and (8) of the **E** and **B** fields, combined with Eqs. (4) and (5) (see Appendix II). This result arises because a metal is, for all purposes, a short circuit to the impinging microwave beam so that a predetermined total current is induced in the metal, irrespective of the conductivity.

B. Mathematical Method

The mathematical method which we shall use is that of Fourier transforms, although it is not possible because of the spatial variation of the conductivity σ to reduce the equations to the usual simple form that they assume in the transformed formulation. We express the z dependence by a Fourier integral, whereas we describe the y dependence by a Fourier series since we expect the fields **E** and **B** to be periodic with the same period as σ. The electric field is, for example, written in the form

$$\mathbf{E}(y, z) = \int_{-\infty}^{+\infty} dq \sum_{m=-\infty}^{+\infty} \mathbf{E}_{mq} e^{i\mathbf{Q}_m \cdot \mathbf{r}}. \tag{13}$$

$\mathbf{E}_{mq}$ will be called the transform of $\mathbf{E}(y, z)$, and actually stands for a function of q, viz., $\mathbf{E}_m(q)$. The vector $\mathbf{Q}_m$ has components

$$\mathbf{Q}_m = (0, mk, q), \tag{14}$$

with the same meaning for k as in Eqs. (1)–(3). The transform $\mathbf{E}_{mq}$ is then

$$\mathbf{E}_{mq} = \frac{k}{4\pi^2} \int_{-\pi/k}^{+\pi/k} dy \int_{-\infty}^{+\infty} dz\, \mathbf{E}(y, z) e^{-i\mathbf{Q}_m \cdot \mathbf{r}}. \tag{15}$$

From now onward all limits of integration over y, z and q and of summation over m (or l) will be the same as in (13) and (15), unless other limits are specifically indicated.

The transform $\mathbf{I}_{lq}$ of the applied current $\mathbf{j}_0$ is obtained easily if one remembers that the Fourier transform of the delta function is a constant $1/2\pi$:

$$\mathbf{I}_{lq} = \frac{1}{2\pi} \mathbf{I}\, \delta(l). \tag{16}$$

In this equation $\delta(l)$ signifies that $\mathbf{I}_{lq}$ is always zero except when $l = 0$. For the induced current $\mathbf{j}(y, z)$, the transform is derived from (10) with (3) and (13). One obtains first

$$\int dq \sum_{l} \mathbf{j}_{lq} e^{i\mathbf{Q}_l \cdot \mathbf{r}} = \{\sigma_0 + \sigma_1(e^{ikv} + e^{-ikv})\} \int dq \sum_{m} \mathbf{E}_{mq} e^{i\mathbf{Q}_m \cdot \mathbf{r}}, \tag{17}$$

which simplifies to

$$\mathbf{j}_{lq} = \sigma_0 \mathbf{E}_{lq} + \sigma_1(\mathbf{E}_{(l-1)q} + \mathbf{E}_{(l+1)q}). \tag{18}$$

Since the transform of $\nabla \times \mathbf{E}(y, z)$ is simply $i\mathbf{Q}_l \times \mathbf{E}_{lq}$, we can now write the transform of our basic equation (6),

$$\mathbf{Q}_l \times \mathbf{Q}_l \times \mathbf{E}_{lq} = \frac{2i\omega}{c^2} \mathbf{I}\delta(l) + \frac{4\pi i\omega}{c^2} \cdot \{\sigma_0 \mathbf{E}_{lq} + \sigma_1(\mathbf{E}_{(l-1)q} + \mathbf{E}_{(l+1)q})\}. \tag{19}$$

Remembering the definition (14) of $\mathbf{Q}_l$ we can write (19) in components, obtaining three infinite sets of equations. Before doing so we shall simplify the notation by the introduction of parameters δ_n having properties and dimensions of skin depths:

$$1/\delta_n^2 = (4\pi i\omega/c^2)\sigma_n. \tag{20}$$

We then get for the x component

$$\left(l^2k^2 + q^2 + \frac{1}{\delta_0^2}\right)E_{lqx} + \frac{1}{\delta_1^2}(E_{(l-1)qx} + E_{(l+1)qx}) = -\frac{2i\omega}{c^2} I_x \delta(l); \tag{21}$$

for the y component

$$\left(q^2 + \frac{1}{\delta_0^2}\right)E_{lqy} - lkqE_{lqz} + \frac{1}{\delta_1^2}(E_{(l-1)qy} + E_{(l+1)qy}) = -\frac{2i\omega}{c^2} I_y \delta(l); \tag{22}$$

and for the z component

$$\left(l^2k^2 + \frac{1}{\delta_0^2}\right)E_{lqz} - lkqE_{lqy} + \frac{1}{\delta_1^2}(E_{(l-1)qz} + E_{(l+1)qz}) = 0. \tag{23}$$

Equations (21) to (23) are the basic formulas of our particular problem. It is clearly evident from these equations that there are two independent field configurations; Eq. (21) only contains the x component of the electric field, whereas y and z components are interconnected in Eqs. (22) and (23). As was mentioned in Sec. I, the two field configurations are expected to yield different surface impedances

and we shall label $Z_\parallel$ the one obtaining when $E_y = E_z = 0$, and $Z_\perp$ the one resulting when $E_x = 0$. It can easily be verified that when the conductivity is a constant σ_0, both configurations lead to the same well-known formulas for the surface impedance Z_0 of a homogeneous isotropic metal, in the limit of a local relation. By definition,

$$Z = \frac{E(z=0)}{\int_0^\infty j_\parallel(z)\,dz} = \frac{E(z=0)}{(-\frac{1}{2}I)}. \tag{24}$$

When $\sigma = \sigma_0$, all terms with $l \neq 0$ vanish in Eqs. (21) to (23) and one finds successively

$$E(z=0) = -\frac{2i\omega}{c^2} I \int_{-\infty}^{+\infty} \frac{dq}{q^2 + 1/\delta_0^2} = -\frac{2i\omega}{c^2} I\pi\delta_0, \tag{25}$$

and

$$Z_0 = \frac{4\pi i\omega\delta_0}{c^2} = \left(\frac{4\pi i\omega}{c^2\sigma_0}\right)^{\frac{1}{2}}. \tag{26}$$

III. THE PARALLEL CONFIGURATION ($E_y = E_z = 0$)

A. Convergence Considerations and Principle of the Solution

We look now at the set of Eqs. (21) only, and, for simplicity, shall drop the indices x and q. Because of the even symmetry of $\sigma(y)$ we also require even symmetry for E,

$$E_l = E_{-l}, \tag{27}$$

and restrict ourselves to positive integers. Starting with $l = 0$ we have the system

$$\left(q^2 + \frac{1}{\delta_0^2}\right)E_0 + \frac{2}{\delta_1^2}E_1 = -\frac{2i\omega}{c^2} I, \quad \text{for } l = 0, \tag{28a}$$

$$\left(l^2k^2 + q^2 + \frac{1}{\delta_0^2}\right)E_l + \frac{1}{\delta_1^2}(E_{l-1} + E_{l+1}) = 0, \quad \text{for } l \geq 1. \tag{28b}$$

We note that (28a) is the only inhomogeneous equation of the system. It is therefore possible to calculate all terms E_l, and the electric field $E_x(y, z)$ is then

$$E_x(y, z) = \int dq\, e^{iqz} E_q, \tag{29}$$

where E_q is the sum

$$E_q = E_0 + 2\sum_1^\infty E_l \cos(lky). \tag{30}$$

We shall call "extremal values" of E_q the sums E_{ex} occurring when the conductivity is extremal, i.e., when $\sigma = \sigma_{ex} = \sigma_0 \pm 2\sigma_1$, and call y_{ex} the corresponding values of y,

$$y_{ex} = n\pi/k = nd/2. \tag{31}$$

We have then

$$E_{ex} = E_0 \pm 2\sum_1^\infty E_l. \tag{32}$$

From the formula that applies to homogeneous metals [Eq. (25)], it is easy to deduce the value of E_{ex} in the limit when $d^2 \gg |\delta_0^2|$, or in other words, when $|k^2\delta_0^2| \ll 1$, and no particular restriction is imposed on δ_1^2,

$$\lim_{k\to 0}(E_{ex}) = -\frac{2i\omega}{c^2} I\left(q^2 + \frac{1}{\delta_0^2} \pm \frac{2}{\delta_1^2}\right)^{-1}. \tag{33}$$

Let us return to Eq. (28b). We can rewrite it in the form

$$\frac{E_l}{E_{l-1}} = \frac{-1}{\delta_1^2(l^2k^2 + q^2 + 1/\delta_0^2) + E_{l+1}/E_l}. \tag{34}$$

When l becomes very large this reduces to $-1/l^2k^2\delta_1^2$, which guarantees absolute convergence of the sums (30) and (32). The form of (34) is that of a "continued fraction"; if we write

$$a_l = \delta_1^2(l^2k^2 + q^2 + 1/\delta_0^2), \tag{35}$$

we get

$$\frac{E_l}{E_{l-1}} = \cfrac{-1}{a_l + \cfrac{-1}{a_{l+1} + \cfrac{-1}{a_{l+2} + \cfrac{-1}{\ddots}}}}, \tag{36}$$

which is usually written in the following way:

$$\frac{E_l}{E_{l-1}} = \left|\frac{-1}{a_l}\right| \left|\frac{-1}{a_{l+1}}\right| \left|\frac{-1}{a_{l+2}}\right| - \cdots, \quad \text{for } l \geq 1. \tag{37}$$

Since we know that this continued fraction converges we have a means to derive E_l/E_{l-1}. In particular, we can derive E_1/E_0. But Eq. (28a) also relates E_1 to E_0, and since this second equation is inhomogeneous it is possible to determine E_0 and E_1 separately, and then all other E_l.

B. Approximate Solution

It is clear from Eqs. (34)–(37) that if a_1 is finite all a_l and all E_l are finite. However, in the sum (30) all oscillatory terms have common extrema at $y = y_{ex}$ given by (31). Let us then ask under what

conditions E_q can be satisfactorily approximated by an expression containing only one oscillatory term, i.e.,

$$E_q^1 = E_0 + 2E_1 \cos(ky). \tag{38}$$

To this end let us calculate the ratio E_2/E_1 with the assumption that

$$|a_5| = |\delta_1^2(25k^2 + q^2 + 1/\delta_0^2)| \gg 1. \tag{39}$$

Since k^2 and q^2 are both always real and positive, condition (39) will generally be satisfied if

$$|25\delta_1^2 k^2| \gg 1. \tag{40}$$

Under these conditions we find that

$$\left|\frac{E_2}{E_1}\right| \cong \left| \frac{-1}{\left|\frac{-1}{a_1}\right| \left|\frac{-1}{a_2}\right| \left|\frac{-1}{a_3}\right|} \right| = \left| \frac{-1}{\delta_1^2\left(4k^2+q^2+\frac{1}{\delta_0^2}\right)} \times \left\{1 + \frac{1}{\delta_1^4\left(4k^2 + q^2 + \frac{1}{\delta_0^2}\right)\left(9k^2 + q^2 + \frac{1}{\delta_0^2}\right)}\right\}\right|. \tag{41}$$

For large q^2 this expression is $\ll 1$. For small q^2 we must require, if this ratio is to remain $\ll 1$, that

$$|4\delta_1^2 k^2| \gg 1. \tag{42}$$

Condition (42) is compatible with (40) but even more restrictive. Since we have, so far, not imposed limitations to σ_1 and d, we now require them to be such that condition (42) obtains. In terms of σ_1 and d this condition states

$$|\sigma_1| \ll (4\pi\sigma^2/\omega)(1/d^2). \tag{43}$$

With the relation between σ_0 and δ_0, (43) can also be written

$$\frac{d^2}{|\delta_0^2|} \ll (4\pi)^2 \left|\frac{\sigma_0}{\sigma_1}\right| \cong 160 \left|\frac{\sigma_0}{\sigma_1}\right|. \tag{44}$$

When $|\sigma_1| \ll |\sigma_0|$ the above condition is hardly a restriction at all; The critical limit is rather that of large variations of the conductivity. If σ_1 and σ_0 have same argument, their ratio is never[5] to exceed $\frac{1}{2}$. d is then allowed to become as large as about ten times the average penetration depth before higher-order terms become important. When they have different arguments, the ratio of σ_1 to σ_0 may perhaps exhibit a module much larger than unity without violation of the condition requested in footnote 5. The largest permissible d would then be accordingly reduced.

We shall not discuss further in this paper the physical implications of condition (44) but shall assume that it is satisfied.[4] Setting $E_l = 0$ for $l \geq 2$ we then solve equations (28a) and (28b), obtaining

$$E_0 = -\frac{2i\omega}{c^2} I \frac{k^2 + q^2 + 1/\delta_0^2}{(q^2 + 1/\delta_0^2)(k^2 + q^2 + 1/\delta_0^2) - 2/\delta_1^4} = F_{kq} I, \tag{45}$$

$$E_1 = \frac{2i\omega}{c^2} I \frac{1/\delta_1^2}{(q^2 + 1/\delta_0^2)(k^2 + q^2 + 1/\delta_0^2) - 2/\delta_1^4} = G_{kq} I. \tag{46}$$

These two expressions are not fully consistent with the limit described by (33), but this is not surprising since (33) eventually violates condition (42) when $k \to 0$.

C. The Fields and Surface Impedance

We are now able to write down an explicit expression for the electric field $E_x(y, z)$ in the parallel configuration. With Eq. (5) we also derive the other nonvanishing field components of this configuration:

$$E_x(y, z) = I \int dq\, e^{iqz}[F_{kq} + 2G_{kq} \cos(ky)], \tag{47}$$

$$B_y(y, z) = \frac{c}{\omega} I \int dq\, e^{iqz} q[F_{kq} + 2G_{kq} \cos(ky)], \tag{48}$$

$$B_z(y, z) = -\frac{2ck}{i\omega} I \sin(ky) \int dq\, e^{iqz} G_{kq}. \tag{49}$$

The integrals in the above expressions can all be calculated in closed form, without approximations, by the method of residues (see Appendix I).

It is worth noticing that B_z does not have a constant term and that its oscillatory term is 90° out of phase with those of E_x and B_y. At the metal surface $z = 0$, one also finds B_y to be a constant independent of y, since, as is proven (see Appendices I and II),

$$\int dq\, qG_{kq} \equiv 0. \tag{50}$$

From Eq. (47) the surface impedance is obtained at once,

$$Z_\parallel(y) = \frac{E_x(z = 0)}{(-\frac{1}{2}I)} = -2 \int dq\, [F_{kq} + 2G_{kq} \cos(ky)]. \tag{51}$$

The integrals in this expression can also be calculated without approximations (see Appendix I). However, discussion of the final result is postponed to Sec. V, after we have derived the corresponding expression for $Z_\perp$.

IV. THE PERPENDICULAR CONFIGURATION ($E_z = 0$)

Deriving $Z_\perp$ leads through similar steps as deriving $Z_\parallel$. Each step is more complicated since we have

two infinite sets of coupled equations (22) and (23). A discussion of the convergence criteria would be very cumbersome, perhaps even impossible, if it had to be carried out with the electric field components E_y and E_z. It is possible, however, to return to a single set of equations if one works with the magnetic field component B_x. From Maxwell's equations (4) and (5) and the boundary conditions discussed in Sec. III a) one can derive a differential equation for B_x:

$$-\left[\frac{\partial^2 B_x}{\partial z^2}+\frac{\partial^2 B_x}{\partial y^2}\right]+\frac{4\pi i\omega}{c^2}\sigma(y)B_x = \frac{4\pi}{c}I_y\delta'(z)+\frac{\sigma'(y)}{\sigma(y)}\frac{\partial B_x}{\partial y}. \tag{52}$$

The left-hand side of this equation is identical with that which obtains for E_x in the parallel configuration where the right-hand side reduces to $(4\pi i\omega/c^2)I_x\,\delta(z)$. The right-hand side of (52) is more complicated, essentially because of the last term which embodies the whole difference between $\parallel$ and $\perp$ configurations. The Fourier transform of this equation can still be used to show that the expansion in Fourier series generally converges.

We shall not discuss the convergence criteria further. In spite of the greater complication inherent in this configuration, one finds the same results as for the parallel configuration. In particular it is found that the fields can be satisfactorily approximated with a single oscillatory term when condition (44) is satisfied.

We proceed now to solve the equations of this configuration. We could solve Eq. (52) to find B_x and then derive E_y and E_z with (4). Alternately we can solve Eqs. (22) and (23) to find first E_y and E_z, and derive B_x with (5). The two methods are of course equivalent; we choose the second one here because we have already written the basic equations in Fourier components.

The fields E_y and E_z cannot, in this configuration, both be taken to have the even symmetry of $\sigma(y)$. Since the current at $z = 0$ flows parallel to the y axis, E_y will have even symmetry too. Then Maxwell's equation div $\mathbf{E} = 0$ requires E_z to be odd, which means

$$E_{-lqz} = -E_{lqz}. \tag{53}$$

When (53) is introduced into the z-component equation with $l = 0$, one sees at once that

$$E_{0qz} \equiv 0, \tag{54}$$

i.e., the E_z component has no constant term. The electric field in the perpendicular configuration has, therefore, the same symmetry properties as the magnetic field in the parallel configuration.

Setting $E_l = 0$ for $l \geq 2$ we have to solve the three equations derived from (22) and (23):

$$\left(q^2+\frac{1}{\delta_0^2}\right)E_{0qy}+\frac{2}{\delta_1^2}E_{1qy} = -\frac{2i\omega}{c^2}I_y, \tag{55}$$

$$\left(q^2+\frac{1}{\delta_0^2}\right)E_{1qy}-kqE_{1qz}+\frac{1}{\delta_1^2}E_{0qy} = 0, \tag{56}$$

$$\left(k^2+\frac{1}{\delta_0^2}\right)E_{1qz}-kqE_{1qy} = 0. \tag{57}$$

The solution can be written in the form

$$\frac{E_{0qy}}{A_{kq}} = \frac{E_{1qy}}{B_{kq}} = \frac{E_{1qz}}{C_{kq}} = I_y, \tag{58}$$

with

$$A_{kq} = -\frac{2i\omega}{c^2} \times \frac{k^2+q^2+1/\delta_0^2}{(q^2+1/\delta_0^2)(k^2+q^2+1/\delta_0^2)-(2/\delta_1^4)(\delta_0^2k^2+1)}, \tag{59}$$

$$B_{kq} = -A_{kq}\frac{\delta_0^2k^2+1}{\delta_1^2(k^2+q^2+1/\delta_0^2)}, \tag{60}$$

$$C_{kq} = A_{kq}\frac{\delta_0^2kq}{\delta_1^2(k^2+q^2+1/\delta_0^2)}. \tag{61}$$

The nonvanishing field components of this configuration thus are

$$E_y(y,z) = I_y\int dq\, e^{iqz}[A_{kq}+2B_{kq}\cos(ky)], \tag{62}$$

$$E_z(y,z) = 2iI_y\sin(ky)\int dq\, e^{iqz}C_{kq}, \tag{63}$$

$$B_x(y,z) = -\frac{c}{\omega}I_y\int dq\, e^{iqz}q \times\left[A_{kq}+2\left(B_{kq}-\frac{k}{q}C_{kq}\right)\cos(ky)\right]. \tag{64}$$

One can again show that the oscillatory part of B_x vanishes for $z = 0$ (see Appendices I and II).

The surface impedance $Z_\perp$ of this configuration is derived from Eq. (62),

$$Z_\perp = \frac{E_y(z=0)}{(-\frac{1}{2}I_y)} = -2\int dq[A_{kq}+2B_{kq}\cos(ky)]. \tag{65}$$

The integrals of this expression can be calculated is closed form (see Appendix I) and the result is given in Sec. V.

TABLE I. Symbols for Eq. (66).

	For $Z_\parallel$	For $Z_\perp$
Z_{00}	$4\pi i\omega\delta_0/c^2$	*idem*
Γ	$1+\frac{1}{2}b$	*idem*
b	$\delta_0^2k^2=4\pi^2(\delta_0^2/d^2)$	*idem*
B	$(\frac{1}{4}b^2+2a^2)^{\frac{1}{2}}$	$\{\frac{1}{4}b^2+2a^2(1+b)\}^{\frac{1}{2}}$
a	$\delta_0^2/\delta_1^2=\sigma_1/\sigma_0$	*idem*

V. RESULTS AND DISCUSSION

For both parallel and perpendicular configuration, the *average* surface impedance $Z_{0\parallel}$ or $Z_{0\perp}$ can conveniently be expressed by the following formula:

$$Z_0=Z_{00}\frac{(\Gamma+B-1)(\Gamma+B)^{\frac{1}{2}}-(\Gamma-B-1)(\Gamma-B)^{\frac{1}{2}}}{2B(\Gamma^2-B^2)^{\frac{1}{2}}}, \quad (66)$$

The meaning of the symbols being given in Table I. The whole difference between the two impedances is concentrated in B.

The spatial dependence of $Z_\parallel$ and $Z_\perp$ is described by

$$Z(y)=Z_0+2Z_1\cos(ky), \quad (67)$$

where

$$Z_{1\parallel}=-Z_{00}a\frac{(\Gamma+B_\parallel)^{\frac{1}{2}}-(\Gamma-B_\parallel)^{\frac{1}{2}}}{2B_\parallel(\Gamma^2-B_\parallel^2)^{\frac{1}{2}}}, \quad (68)$$

and

$$Z_{1\perp}=-Z_{00}a(1+b)\frac{(\Gamma+B_\perp)^{\frac{1}{2}}-(\Gamma-B_\perp)^{\frac{1}{2}}}{2B_\perp(\Gamma^2-B_\perp^2)^{\frac{1}{2}}}. \quad (69)$$

The parameters of these two formulas are also given in Table I. Formulas (66)–(69) are valid in the range where condition (44) applies, that is for periods d of the conductivity pattern which are shorter than a certain upper limit. For very long periods d, however, it is easy to write down the correct expression for the surface impedance. When $d>|\delta_0|$, both configurations lead to identical results, the surface impedance being then uniquely determined by the *local* value $\sigma(y)$ of the conductivity [compare with (26)],

$$Z(y)=\frac{4\pi i\omega\delta(y)}{c^2}=\left[\frac{4\pi i\omega}{c^2\sigma(y)}\right]^{\frac{1}{2}}. \quad (70)$$

The average value Z_0 is then obtained by integrating $Z(y)/d$ over a period d.

When $a\to 0$, that is when the oscillatory part of $\sigma(y)$ disappears, both $Z_\parallel$ and $Z_\perp$ tend toward Z_{00}. This consistency check brings out an important point: In spite of assuming a local "current–field" (**j**–**E**) relation we do not end up with a local "surface impedance–conductivity" (Z–σ) relation. There is a smearing-out effect which results here not from a long mean free path l but because of field penetration effects. The fields cannot change much over distances smaller than a skin depth, this being true whether one moves perpendicularly to the periodic conductivity pattern ($\parallel$ y axis) or parallel to it ($\parallel$ x or z axis). But the periodic pattern acts differently depending on the direction of the electric field vector, in a manner reminiscent of the polarizing action of a metal grating on a microwave beam.

The difference between $Z_\parallel$ and $Z_\perp$ is difficult to analyze in general because the formulas are complex. σ_0 and σ_1 themselves are generally complex. Since the limiting behavior when $|\delta_0|<d$ and therefore $|\delta_0k|<2\pi$ is adequately described with formula (70), we shall consider the other limit

$$|\delta_0^2k^2|=|b|\gg 1 \quad \text{and} \quad |b|\gg|a| \quad \text{or} \quad |a^2|. \quad (71)$$

In this second limit, $Z_{0\parallel}$ and $Z_{0\perp}$ assume very simple forms

$$Z_{0\parallel}\to Z_{00}, \quad (72)$$

$$Z_{0\perp}\to Z_{00}(1-2a^2)^{-\frac{1}{2}}. \quad (73)$$

We shall not discuss here the values that a may take from a physical point of view since we intend to do so elsewhere[4]; but if we assume for a moment that σ_1 and σ_0 have identical arguments, their ratio a becomes real and positive. By virtue of Eq. (3) the ratio $2\sigma_1/\sigma_0=2a$ then gives the degree of modulation of the periodic conductivity and obviously a cannot become larger than $\frac{1}{2}$. When $a=\frac{1}{2}$ there are planes

$$y=[(2n+1)/k]\pi=(n+\tfrac{1}{2})d, \quad (74)$$

where the conductivity becomes zero. Our whole analysis then breaks down since by neglecting displacement currents[5] we have specifically implied that the conductivity is always larger than a certain minimum of the order of one $(\Omega\,\text{cm})^{-1}$. It is nevertheless interesting to compare the behavior of $Z_{0\parallel}$ and $Z_{0\perp}$ when a increases toward the ratio $\frac{1}{2}$, and conditions (71) also hold,

$$Z_{0\parallel}\to Z_{00}, \quad (75)$$

$$Z_{0\perp}\to\sqrt{2}Z_{00}. \quad (76)$$

VI. BASIC EQUATIONS OF A MORE GENERAL PROBLEM

The generalization which we intend to perform in this section is to assume a periodic conductivity of the general form (1) instead of form (3). In addition we choose a more general nonlocal rela-

tion between current and electric field, of the form

$$\mathbf{j}(\mathbf{r}) = \int d^3r'\sigma(r')K(\mathbf{r} - \mathbf{r}')\mathbf{E}(\mathbf{r}'), \tag{77}$$

the integrals extending over the entire space. (77) is not the most general form of (**j**–**E**) relation that can be suggested, but it is interesting because it combines position-dependent and nonlocal features and leads to equations sufficiently simple that solution along the lines proposed in the present paper appears possible.

Since there is no x dependence in σ and $\mathbf{E}$, the integration over this coordinate can be effected at once. In the y direction, σ and $\mathbf{E}$ are periodic so that (77) can now be rewritten in less general form:

$$\mathbf{j}(y, z) = \int_{-\pi/k}^{+\pi/k} dy' \int_{-\infty}^{+\infty} dz'\sigma(y') \times K_p(y - y', z - z')\mathbf{E}(y', z'), \tag{78}$$

where K_p is an *equivalent periodic* kernel, independent of x and related to the original nonperiodic kernel K by the sum and integral

$$K_p(y - y', z - z') = \int_{-\infty}^{+\infty} dx' \sum_{n=-\infty}^{+\infty} K(x - x', y - y' - nd, z - z'). \tag{79}$$

All quantities under the integral signs of (78) are known and therefore also their Fourier transforms. For example, one has

$$K_{lq} = \frac{k}{4\pi^2} \int dy \int dz K_p(\mathbf{r} - \mathbf{r}')e^{-i\mathbf{Q}_l(\mathbf{r}-\mathbf{r}')}. \tag{80}$$

Our next step is to find the transform $\mathbf{j}_{lq}$ of $\mathbf{j}(\mathbf{r})$,

$$\mathbf{j}_{lq} = \frac{k}{4\pi} \int dy \int dz \int dy' \int dz'\sigma(\mathbf{r}') \times K_p(\mathbf{r} - \mathbf{r}')\mathbf{E}(\mathbf{r}')e^{-i\mathbf{Q}_l\cdot\mathbf{r}}. \tag{81}$$

In spite of writing for simplicity the quantities under the integrals as functions of $\mathbf{r}$ and $\mathbf{r}'$, we remember that $\mathbf{E}$ and K_p are independent of x and x', and σ also of z and z'. Now (81) can be at once be transformed into

$$j_{lq} = K_{lq} \int dy' \int dz'\sigma(\mathbf{r}')E(\mathbf{r}')e^{-i\mathbf{Q}_l\cdot\mathbf{r}'}. \tag{82}$$

With (1) and (13) the above equation can in turn be transformed to eliminate all integrations, finally obtaining the form

$$\mathbf{j}_{lq} = K_{lq}\frac{4\pi^2}{k} \sum \sigma_n \mathbf{E}_{(l-n)q}. \tag{83}$$

This expression, the generalization of (18), is essentially a nonlocal (**j**–**E**) relation in phase-space but has the advantage over (78) that the kernel part is not under summation. The nonlocal character of (83) arises because of the periodic nature of σ; if we had a uniform conductivity σ_0, Eq. (77) would transform into the usual relation

$$\mathbf{j}(q) = \sigma_0 K(q)\mathbf{E}(q). \tag{84}$$

The spatial variation of σ prevents us from deriving such a simple relationship.

We are now capable of writing the generalized form of Eqs. (21)–(23). We have for the x component

$$(l^2k^2 + q^2)E_{lqx} + \frac{4\pi^2}{k} K_{lq} \sum_n \frac{1}{\delta_n^2} E_{(l-n)qx} = -\frac{2i\omega}{c^2} I_x \delta(l); \tag{85}$$

for the y component

$$q^2E_{lqy} - lkqE_{lqz} + \frac{4\pi^2}{k} K_{lq} \sum_n \frac{1}{\delta_n^2} E_{(l-n)qy} = -\frac{2i\omega}{c^2} I_y \delta(l); \tag{86}$$

and for the z component

$$l^2k^2E_{lqz} - lkqE_{lqy} + \frac{4\pi^2}{k} K_{lq} \sum_n \frac{1}{\delta_n^2} E_{(l-n)qz} = 0. \tag{87}$$

Equations (85)–(87) are the basic formulas of the more general surface impedance problem when the scalar conductivity is periodic in a direction parallel to the surface. Any particular problem will have to specify the components σ_n and K_{lq} of the conductivity and kernel expansions. The existence of two independent configurations is again clearly evident from these three sets of equations, since the first set only contains x components whereas y and z components are interconnected in the other two sets.

ACKNOWLEDGMENTS

We would like to thank Dr. P. Erdös, Dr. J. J. Quinn and Dr. W. Rehwald for many helpful discussions.

APPENDIX I

In the calculation of the surface impedance via formulas (51) and (65) one has to calculate integrals of the following type

$$\int_{-\infty}^{+\infty} dqF_{kq}, \quad \text{or} \quad \int_{-\infty}^{+\infty} dqG_{kq}, \tag{88}$$

where F_{kq} and G_{kq} are defined by (45) and (46). Except for a few poles, F_{kq} and G_{kq} are analytic functions for any complex q. We can therefore apply Cauchy's theorem:

$$\oint_C drF(\mathbf{r}) = 2\pi i \sum_C \text{Res}\,\{F(\mathbf{r})\}. \tag{89}$$

The integration is along a closed curve C in the positive sense and the summation is over the residues of the poles of $F(\mathbf{r})$ inside the domain enclosed by C. We then choose as curve C the whole real axis and the infinite semicircle of the positive half-plane. Our functions F_{kq} and G_{kq} vanish everywhere on the semicircle and we can write

$$\int_{-\infty}^{+\infty} dqF_{kq} = 2\pi i \sum \text{Res}\,(F_{kq}), \tag{90}$$

where the sum extends over the entire positive half-plane.

F_{kq} and G_{kq}, and also A_{kq} and B_{kq} defined by (59) and (60), all have biquadratic denominators in q. For each root in the positive half-plane there is an equal but negative root in the negative half-plane. Thus our functions have two poles in the positive half-planes and their residues can easily be found. Let us, as an example, consider F_{kq} in some detail,

$$F_{kq} = -\frac{2i\omega}{c^2}\frac{k^2 + q^2 + 1/\delta_0^2}{(q^2 + 1/\delta_0^2)(k^2 + q^2 + 1/\delta_0^2) - 2/\delta_1^4}$$

$$= -\frac{2i\omega}{c^2} f(q). \tag{91}$$

We transform $f(q)$ with some evident substitutions:

$$f(q) = \frac{k^2 + q^2 + 1/\delta_0^2}{(q^2 + 1/\delta_0^2 + \frac{1}{2}k^2)^2 - (2/\delta_1^4 + \frac{1}{4}k^4)}$$

$$= \frac{q^2 + 2\gamma - 1/\delta_0^2}{(q^2 + \gamma)^2 - \beta^2}, \tag{92}$$

or

$$f(q) = \frac{q^2 + 2\gamma - 1/\delta_0^2}{(q^2 + \gamma + \beta)(q^2 + \gamma - \beta)}. \tag{93}$$

The poles of $f(q)$ are

$$q_{1,2} = \pm i(\gamma + \beta)^{\frac{1}{2}} \quad \text{and} \quad q_{3,4} = \pm i(\gamma - \beta)^{\frac{1}{2}}. \tag{94}$$

One root of each $\pm$ set must be in the positive half-plane and the other in the negative half. Let us assume that the $+$ sign corresponds to the roots in the positive half. This choice is not arbitrary; it has to be such that in limiting cases the correct result is obtained, for example when σ_1, and thus also $1/\delta_1^2$, vanishes. With this choice of signs we find the residues

$$\text{Res}\,(q_1) = \frac{\gamma - \beta - 1/\delta_0^2}{[2i(\gamma + \beta)^{\frac{1}{2}}](-2\beta)}, \tag{95}$$

and

$$\text{Res}\,(q_3) = \frac{\gamma + \beta - 1/\delta_0^2}{(2\beta)[2i(\gamma - \beta)^{\frac{1}{2}}]}. \tag{96}$$

We then obtain for the integral (90):

$$\int_{-\infty}^{+\infty} dqF_{kq} = 2\pi i\left(-\frac{2i\omega}{c^2}\right)$$

$$\cdot\frac{(\gamma + \beta - 1/\delta_0^2)(\gamma + \beta)^{\frac{1}{2}} - (\gamma - \beta - 1/\delta_0^2)(\gamma - \beta)^{\frac{1}{2}}}{4i\beta(\gamma^2 - \beta^2)^{\frac{1}{2}}}. \tag{97}$$

With (51) and Table I, (97) leads at once to the surface impedance formula (66). The other integrals are evaluated in similar fashion; in particular one sees that if the numerator of $f(q)$ had simply been q, the two residues would have been of equal magnitude but opposite signs. This proves (50).

APPENDIX II

We want to prove, as a consequence of specular reflection, the symmetry relations (7) and (8) and the resulting independence in expression (12) of the sheet current I vs coordinate y.

The force $\mathbf{F}$ acting on an electron with wave vector $\mathbf{k}$ is

$$\mathbf{F} = -e\mathbf{E} - (e\hbar/mc)(\mathbf{k}\times\mathbf{B}) = -e\mathbf{E} - 2\mu_B(\mathbf{k}\times\mathbf{B}). \tag{99}$$

In components, (99) reads

$$\begin{aligned} F_x &= -eE_x - 2\mu_B(k_yB_z - k_zB_y), \\ F_y &= -eE_y - 2\mu_B(k_zB_x - k_xB_z), \\ F_z &= -eE_z - 2\mu_B(k_xB_y - k_yB_x). \end{aligned} \tag{100}$$

Under reflection at $z = 0$, all z components of $\mathbf{r}$, $\mathbf{k}$, and $\mathbf{F}$ change sign whereas x and y components remain unchanged. Changing k_z without changing F_x and F_y requires changing B_x and B_y. B_z, on the other hand, remains unchanged so that (8) is proved. The contribution of $\mathbf{E}$ to $\mathbf{F}$ is parallel to $\mathbf{E}$, which leads to (7). The remaining part of the proof is different in the two configurations.

A. Perpendicular Configuration

In this configuration one has

$$B_y = B_z = E_x = j_z \equiv 0. \tag{101}$$

Taking the divergence of (4) we find $\nabla\cdot\mathbf{j} = 0$, which means in the integral form

$$\int dfj_n = 0. \tag{102}$$

This integral is over the entire surface of a singly connected volume and j_n means the component of the current density normal to the surface. Choosing as volume a rectangular prism with edges parallel to the coordinate axis and a body diagonal represented by $\mathbf{r} = (1, y, \infty)$, we find

$$\int_0^y [j_z(y, 0) - j_z(y, \infty)]\, dy + \int_0^\infty [j_y(0, z) - j_y(y, z)]\, dz = 0. \quad (103)$$

But $j_z(y, \infty) = 0$ and $j_z(y, 0) = \sigma(y)E_z(y, 0)$ also vanishes because E_z, according to (7), is antisymmetric with respect to coordinate z. It then follows that

$$\int_0^\infty j_y(y, z)\, dz = -\tfrac{1}{2}I_y \quad (104)$$

is independent of coordinate y. With the integral form of (4) and the same volume of integration, one also finds that

$$B_z(y, 0) = (2\pi/c)I_y \quad (105)$$

is independent of y.

B. Parallel Configuration

In this configuration one has

$$B_z = E_y = j_y = E_z = j_z \equiv 0, \quad (106)$$

and from (8) we know that B_z is symmetric with respect to coordinate z,

$$(\partial B_z/\partial z)_{z=0} = 0. \quad (107)$$

But from (5) we derive $\nabla \cdot \mathbf{B} = 0$, and therefore

$$(\partial B_y/\partial y)_{z=0} = 0. \quad (108)$$

In this configuration we thus find again

$$\int_0^\infty j_z(y, z)\, dz = -\tfrac{1}{2}I_z = -\frac{c}{4\pi} B_y(y, 0), \quad (109)$$

with I_z and $B_y(y, 0)$ independent of y.

On the Integrodifferential Equations of the Nonlocal Theory of Superconductivity*

FAROUK M. ODEH
International Business Machines Corporation, Thomas J. Watson Research Center, Yorktown Heights, New York
(Received 1 October 1962; final manuscript received 28 February 1964)

The nonlocal model of superconductivity has, in recent years, received strong support on both microscopic theoretical and experimental grounds. In the present report two of the mathematical aspects of the theory are discussed. One concerns the question of well-posedness of the equations and boundary conditions. The other concerns some methods of obtaining approximate solutions. Existence and uniqueness theorems are given for the solutions of the integrodifferential equations governing the vector potential, under suitable restrictions on the kernels involved. Both diffuse and specular reflection types of boundary conditions are considered in both scalar (one-dimensional) and vector (three-dimensional) problems. An existence theorem is also given for the time-dependent equations. Approximate solutions for the problem of a film in a parallel magnetic field are given in two limiting cases. In one the small-coherence-length limit in the Pippard model under diffuse scattering boundary conditions is considered, and a boundary-layer-type solution is given. In the other a thin-film approximation to the diffuse problem is found. An exact series representation is also given for the solution of the problem of a circular cylinder in a parallel magnetic field with specular reflection boundary conditions.

INTRODUCTION

THE London phenomenological model of superconductivity, as described by London's well-known equations, has proved to be extremely useful in describing the electromagnetic behavior of superconducting bulk materials. In the last decade, however, several experimental and theoretical investigations indicated that generalizations of this model should be considered. On the basis of various properties of superconductors, mainly the variation of penetration depth with impurity contents and the sharpness of the transition in zero field, Pippard[1] concluded that there is a coherence effect in superconductors which extends over regions comparable to the penetration depth. Pippard then proposed a nonlocal generalization of London's first equation which relates the superconducting current density and the magnetic vector potential via a linear integral transformation. A similar result was also derived from the Bardeen–Cooper–Schrieffer (BSC) microscopic theory.[2] A recent experiment by Sommerhalder and Drangeid[3] on the reversal of sign of a magnetic field inside a superconductor supports the nonlocal over the local theory.

The time-dependent (second) London equation was also generalized by London himself in 1953.[4] These generalizations provide a nonlocal electrodynamics which should well account for the behavior of superconductors in weak fields.

Several calculations[5,6] employing the nonlocal theories were made a few years ago. The theoretical and experimental support which the theory received later led to several other recent ones.[7,8] However, the formal mathematical aspects of these nonlocal theories, e.g., formulation, existence, and uniqueness questions, etc., received little attention.[9] The present report is, in part, concerned with some of these aspects. Some methods of finding approximate solutions in certain limiting cases are also considered. For convenience, the plan of the report is outlined.

In Sec. 1, which is largely expository, the nonlocal equations are introduced. In order to formulate the problems of this nonlocal theory as boundary-value problems, a discussion is given of the notions of "diffuse scattering" and "generalized specular reflection" on the boundaries of a superconductor. The specular-reflection discussion proceeds along the same lines as in Schafroth.[9] However, the discussion in Ref. 9 is incomplete as far as the relationship

* The results reported in this paper were obtained in the course of research jointly sponsored by the Mathematics Branch of the Office of Naval Research [Contract Nonr-3504(00)] and IBM.

[1] A. B. Pippard, Proc. Roy. Soc. (London) **A216**, 547 (1953).

[2] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **108**, 3, 1175 (1957).

[3] R. Sommerhalder and K. Drangeid, Phys. Rev. Letters **8**, 467 (1962).

[4] J. Bardeen, "Theory of Superconductivity," *Encyclopedia of Physics*, edited by S. Flügge (Springer-Verlag, Berlin, 1956), Vol. XV, No. II, pp. 274–369.

[5] J. R. Schrieffer, Phys. Rev. **106**, 47 (1957).

[6] M. Peter, Phys. Rev. **109**, 1856 (1958).

[7] R. Sommerhalder and H. Thomas, Helv. Phys. Acta **34**, 29 (1961).

[8] A. M. Toxen, "Critical Fields of Thin Superconducting Films," NC-42, IBM Research Center, Yorktown Heights, New York (1962).

[9] M. R. Schafroth, "Theoretical Aspects of Superconductivity," *Advances in Solid-State Physics* (Academic Press Inc., New York, 1960), Vol. 10.

between the current and vector potential is concerned. Moreover, as Schafroth points out, the assumptions in Ref. 9 do not specify the specular-reflection kernels uniquely. An additional, rather natural, condition is given to complete Schafroth's discussion.

In Sec. 2 the existence and uniqueness of the solution to the nonlocal equations in both the diffuse and specular reflection cases is proved. An existence theorem is also given for the time-dependent case. For such theorems to hold it was found necessary to restrict the class of the kernels of the integral transformations which relate the current and vector potential. The main restriction may be stated as requiring a negative susceptibility at all wavelengths in a super conductor and therefore corresponds to a real physical limitation. The theorems are valid in any number of dimensions and arbitrary smooth shapes of superconductors although not all of the boundary conditions (e.g., diffuse scattering) are physically meaningful in more than one dimension.

In Sec. 3 various approximations to the solution of the equations are discussed since exact solutions seem to be known for one case only, namely the half-space (one-dimensional) geometry.[10] First the small-coherence-length case in the Pippard picture is discussed. The solution in this case approximates the London solution, but the higher-order terms change rapidly near the boundary, i.e., the solution is of a boundary-layer type. The thin-film approximation is also discussed. This case was treated by Peter.[6] However, a simple convergent series representation for the solution which seems to be different from Peter's is given. The generalized specular reflection type of boundary condition is shown to give the same solution as that of Schrieffer[5] in the particular case of the plane-film geometry. The general specular reflection problem is then solved for the case of a circular cylinder in an axial field.

1. PRELIMINARIES AND FORMULATION

A. Infinite Media

The diamagnetic properties of a superconductor are usually described by introducing (the notion of) a superconducting current density, $\mathbf{J}^s$, which is governed, in the London model, by the equations field, and $\mathbf{A}$ is the vector potential in the London

$$\mathbf{J}^s(\mathbf{r}) = -(1/c\Lambda)\mathbf{A}(\mathbf{r}), \tag{1.1}$$

$$(\partial/\partial t)\mathbf{J}^s = (1/\Lambda)\mathbf{E}, \tag{1.2}$$

[10] E. H. Sondheimer and G. E. H. Reuter, Proc. Roy. Soc. (London) **A195**, 336 (1948).

where Λ is a material constant, $\mathbf{E}$ is the electric gauge uniquely defined in simply connected specimens by div $\mathbf{A} = 0$ and $\mathbf{A}\cdot\mathbf{n} = 0$ on the boundaries of a superconductor, where $\mathbf{n}$ is the unit (outward) normal vector. The second equation (1.2) is, in infinite media,[4] a consequence of (1.1) and Maxwell's equations. Hence, it is sufficient to consider the first equation. To fix ideas assume that we are dealing with an infinite three-dimensional space. Then the most general linear continuous (the notion of continuity and the proof of these statements will be given later in the section) transformation between $\mathbf{J}$ and $\mathbf{A}$ has, in terms of Cartesian components, the form

$$J_l^s(\mathbf{r}) = \sum_{m=1}^{3} \int K_{lm}^0(\mathbf{r}, \mathbf{r}') A_m(\mathbf{r}')\, d\mathbf{r}', \qquad l, m = 1, 2, 3, \tag{1.3}$$

where the kernel K_{lm}^0 depends on the material and the integration is carried over the whole space. For example, the London model is described by $K_{lm}^0(\mathbf{r}, \mathbf{r}') = -(c\Lambda)^{-1}\, \delta_{lm}\, \delta(\mathbf{r} - \mathbf{r}')$, where δ_{lm} is the Kronecker delta. In the Pippard model, the relation (1.3) reads

$$\mathbf{J}^s(\mathbf{r}) = -\frac{3}{4\pi\xi_0 c\Lambda} \int \frac{(\mathbf{r} - \mathbf{r}')[(\mathbf{r} - \mathbf{r}')\cdot\mathbf{A}(\mathbf{r}')]e^{-R/\xi}}{R^4}\, d\mathbf{r}', \tag{1.4}$$

where $R = |\mathbf{r} - \mathbf{r}'|$; ξ is the coherence length, ξ_0 is the coherence length in pure materials.

To obtain an equation for the vector potential, one combines the linear relation (1.3) with Maxwell's equations. In the static case $J = J^s$, and therefore

$$\nabla^2 A_l(\mathbf{r}) = -\frac{4\pi}{c} J_l^s(\mathbf{r}) = -\frac{4\pi}{c} \sum_{m=1}^{3} \int K_{lm}^0(\mathbf{r}, \mathbf{r}') \times A_m(\mathbf{r}')\, d\mathbf{r}', \qquad l = 1, 2, 3. \tag{1.5}$$

The second London law (1.2) is generalized (Ref. 4, p. 320) by assuming that the relation (1.3) holds at all times and that the kernel K^0 does not depend on time. (These assumptions should be true only if the frequencies involved are small with respect to a certain critical frequency; see Ref. 4.) Since, in bulk materials, $\mathbf{E} = -(1/c)\, \partial\mathbf{A}/\partial t$, the time derivative of (1.3) gives

$$\frac{\partial \mathbf{J}_l^s}{\partial t} = -c \sum_{m=1}^{3} \int K_{lm}^0(\mathbf{r}, \mathbf{r}') E_m(\mathbf{r}', t)\, d\mathbf{r}', \qquad l, m = 1, 2, 3. \tag{1.6}$$

Equations (1.3), (1.6) replace (1.1), (1.2) in de-

fining the new electrodynamics. To obtain a time-dependent equation for $\mathbf{A}$ one can assume a two-fluid model, i.e., that the total current density $\mathbf{J}$ is composed of the supercurrent plus a normal current which is equal to $\sigma\mathbf{E}$ where σ is the conductivity. Maxwell's equations then give

$$\nabla \times \nabla \times \mathbf{A} = (4\pi/c)[\mathbf{J}^s + \sigma\mathbf{E}] + (1/c)\,\partial E/\partial t. \quad (1.7)$$

Substituting (1.3) into (1.7),

$$\frac{1}{c^2}\mathbf{A}_{tt} = \nabla^2\mathbf{A} + \frac{4\pi}{c}\int K^0\mathbf{A}\,d\mathbf{r} - \frac{4\pi\sigma}{c^2}\mathbf{A}_t. \quad (1.8)$$

In an infinite isotropic medium, the kernel K^0 becomes a difference kernel (see Ref. 9, p. 317), and (1.8) reduces to

$$-\nabla^2\mathbf{A} = \frac{4\pi}{c}\mathbf{J}^s = \frac{4\pi}{c}\int K^0(|\mathbf{r}-\mathbf{r}'|)\mathbf{A}(\mathbf{r}')\,d\mathbf{r}'. \quad (1.9)$$

If the displacement current is neglected in (1.8), then the time-dependent equation becomes

$$\frac{4\pi\sigma}{c^2}\frac{\partial\mathbf{A}}{\partial t} = \nabla^2\mathbf{A} + \frac{4\pi}{c}\int K^0(|\mathbf{r}-\mathbf{r}'|)\mathbf{A}(\mathbf{r}')\,d\mathbf{r}'. \quad (1.10)$$

In an infinite medium, only the normalized forms (1.9), (1.10) will be discussed.

B. Finite Regions

Special forms of the kernel K^0 which relate the vector potential with the supercurrent [Eq. (1.3)] have been derived—or suggested—for infinite media only. Although it may be possible to derive the effects of boundaries on such kernels from the fundamental microscopic theories, there exist, on the macroscopic level, two ways of taking these effects into account:

1. "Diffuse Scattering"

Consider a bounded volume Ω of superconducting material. Let us assume the extreme case when the superconducting electrons lose all memory of their past history once they hit the boundary $\partial\Omega$. Since the nonlocal kernel measures, in some sense, the interaction in a superconductor, it seems plausible to assume that the new "interaction" K, appropriate to the region Ω, can be derived from the original kernel K^0 by restricting the latter to the superconducting volume only, without any other change in form or magnitude. Hence, the "diffuse scattering" case may be defined by introducing a new kernel $K(\mathbf{r}, \mathbf{r}')$ where

$$K(\mathbf{r}, \mathbf{r}') = \begin{cases} K^0(\mathbf{r}, \mathbf{r}') & \text{on} \quad \Omega \times \Omega, \\ 0 & \text{otherwise.} \end{cases} \quad (1.11)$$

However, as Bardeen notes in Ref. 4, this definition may violate, in general, the physical requirement that $\mathbf{J}\cdot\mathbf{n} = 0$ on the boundary of the super region. In the special cases of a plane-surface boundary, with parallel magnetic field, or an infinite cylinder with arbitrary cross section in a uniform axial field, it is clear that the normal component of the current on the boundary vanishes because it is equal to the tangential derivative of the field. Then, as stated in Ref. 4, the diffuse scattering case can be defined consistently.

2. Generalized Specular Reflection

In connection with a discussion of the relation between the flux B and magnetization M in a superconductor, Schafroth[9] proposed another method for constructing the kernel appropriate to a finite region Ω. He also indicated the necessary changes required to obtain the kernel relating J and A. A similar discussion will be given here of the same question for two reasons. First, the changes indicated in Ref. 9 are not sufficient to define a complete problem. Second, and more important, a new assumption will be added to those of Schafroth. This assumption removes the nonuniqueness in the choice of Schafroth's specular reflection kernel.

Let $K^0(\mathbf{r}, \mathbf{s})$ be the kernel relating J and A in an infinite isotropic medium in the London gauge. Then K^0 should be a difference kernel whose Fourier transform $\chi(\mathbf{q}) = \chi(|q|)$ will be assumed to have the form

$$\chi(q) = P(q^2)/Q(q^2), \quad (1.12)$$

where P, Q are polynomials of degrees n, m, respectively. (For greater detail, see Ref. 9.) Following the same arguments as in Ref. 9, one finds:

(i) The assumption of the existence of a linear continuous relation between J and A leads, by the Riesz representation theorem (Reference 11, p. 517, exercise 46) to the existence of a kernel $K_{lm}(\mathbf{r}, \mathbf{s})$ which is integrable and of bounded variation such that

$$J_l(\mathbf{r}) = \int_\Omega \sum_{m=1}^{3} K_{lm}(\mathbf{r}, \mathbf{s})A_m(\mathbf{s})\,d\mathbf{s}, \qquad l, m = 1, 2, 3. \quad (1.13)$$

(ii) The principle that superconductivity is a bulk property[9] (i.e., that point relations are preserved) requires that

$$Q(-\nabla^2)\mathbf{J}(\mathbf{r}) = P(-\nabla^2)\mathbf{A}(\mathbf{r}). \quad (1.14)$$

[11] N. Dunford and J. T. Schwartz, *Linear Operators* (Interscience Publishers, Inc., New York, 1958).

A special kernel which satisfies (1.14) may be constructed as follows: Let $\{\mathbf{u}_n, \lambda_n\}$ denote the complete [in $L_2(\Omega)$] orthonormal set of eigenfunctions and the associated set of eigenvalues defined by the following problem:

$$\text{curl curl } \mathbf{u}_n = \lambda_n^2 \mathbf{u}_n \quad \text{in} \quad \Omega, \tag{1.15}$$

$$\text{div } \mathbf{u}_n = 0 \quad \text{in} \quad \Omega, \tag{1.16}$$

$$\mathbf{u}_n \cdot \boldsymbol{\nu} = 0 \quad \text{on} \quad \partial\Omega, \tag{1.17}$$

$$\boldsymbol{\nu} \cdot \text{curl } \mathbf{u}_n = 0 \quad \text{on} \quad \partial\Omega, \tag{1.18}$$

where $\boldsymbol{\nu}$ is the normal to $\partial\Omega$. Conditions (1.16), (1.17) are imposed because both $\mathbf{J}$ and $\mathbf{A}$ must satisfy them also. Condition (1.18), which is missing in Ref. 9, then *has* to be imposed to make the problem self-adjoint (Reference 12, Sec. 13.1), and thus assures us of the existence of the eigenfunctions. To prove that the problem is self-adjoint, one uses Green's second vector identity in a standard manner. Now let

$$\mathbf{J} = \sum J_n \mathbf{u}_n, \tag{1.19}$$

$$\mathbf{A} = \sum A_n \mathbf{u}_n. \tag{1.20}$$

Then an easy calculation shows that the choice

$$J_n = \chi(\lambda_n) A_n \tag{1.21}$$

does satisfy the differential relation (1.14). This choice is equivalent to the following choice of kernel in relation (1.13):

$$K_{lm}(\mathbf{r}, \mathbf{s}) = \sum_n [\mathbf{u}_n(\mathbf{r})]_l \chi(\lambda_n) [\mathbf{u}_n(\mathbf{s})]_m. \tag{1.22}$$

Notice that the summation converges, in the mean, if one assumes that K^0 dies at infinity, so that the sequence $\{\chi(\lambda_n)\} \in l_2$ where l_2 is the space of sequences with the property that the sum of the squares of the terms is convergent. However, it should be noted, in answer to a question raised by Schafroth, that (1.22) is *not* the most general kernel which makes the local relationship (1.14) hold true and at the same time defines a continuous linear relationship between $\mathbf{J}$ and $\mathbf{A}$, except in one very special case. For example, consider the one-dimensional case in which both A, J have one component only, in the z direction for example, and the region Ω is the plane film $-\pi \leq x \leq \pi$. In this case, the conditions (1.16), (1.17) are automatically satisfied for J and A, and (1.18) reduces to $\partial u/\partial x = 0$ at $x = \pm\pi$. Let $f(x)$ be a solution to $Q(-d^2/dx^2)\, f(x) = 0$ with the boundary conditions $df/dx = 0$ at $x = \pm\pi$. There is always a nontrivial solution if Q is of a degree greater than one. Define $J(x)$ now by

$$J(x) = \int_{-\pi}^{+\pi} [K(x, \xi) + f(x + \xi)] A(\xi)\, d\xi, \tag{1.23}$$

where K is given by (1.22); then the differential relation (1.14) is again satisfied. It can also be shown that (1.23) is the most general relation which satisfies all the assumptions imposed, i.e., that the symmetric kernel K is unique, modulo functions f which satisfy $Q(-\nabla^2)f = 0$ and the boundary conditions (1.16), (1.17), and (1.18). Only in the particular case $Q(z) = z + \alpha$, $\alpha \geq 0$, do the functions f have to vanish. This nonuniqueness may be avoided by adding the following condition.

(iii) Let Ω_N denote an expanding sequence of regions in three dimensions which converge to the entire space as $N \to \infty$. Then the specular reflection kernel for Ω_N should approach, in a suitable sense, the original kernel K^0.

Without specifying the exact limiting operation mentioned above, we remark here that the kernels (1.22) satisfy condition (iii) above. If a nontrivial solution of $Q(-\nabla^2)f = 0$ is added—as in Eq. (1.23)—to the kernel, then the condition will be violated. These statements are clear in one-dimensional cases. They follow in the general case from the results of Levitan[13] about the spectral function of Laplace's operator. Therefore, conditions (i), (ii), and (iii) specify uniquely a kernel which will be called a specular reflection kernel. The justification for the name will become apparent in Sec. 3 where it is shown that it leads to the same solution as given by specular reflection boundary conditions in plane-geometry problems.

2. UNIQUENESS AND EXISTENCE THEOREMS

A. Diffuse Scattering

Consider first the one-dimensional problem of a superconducting film, of thickness $2d$ in the x direction and infinite in both the y, z directions. If magnetic fields of magnitudes H_1, H_2 are applied parallel to the film, in the y direction, then both $\mathbf{A}$, $\mathbf{J}$ have only a z component which depends only on x. In this case (1.9) reduces to

$$\frac{d^2A}{dx^2} = \int_{-d}^{+d} K(|x - \xi|) A(\xi)\, d\xi, \quad -d \leq x \leq d, \tag{2.1}$$

[12] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953).

[13] B. M. Levitan, "On the Asymptotic Behavior of the Spectral Function and Expansion in Eigenfunctions of $\Delta u + (\lambda - q)\, u = 0$," Amer. Math. Soc. Transl. Ser. 2, **20**, 1 (1962).

where $K = -(4\pi/c)K^0$ together with the boundary conditions

$$dA/dx = H_{1,2} \quad \text{at} \quad x = \pm d. \tag{2.2}$$

In order to discuss the properties of the solution to the problem (2.1), (2.2), one has to consider the bilinear form $\mathfrak{K}(f, g)$ defined over the real continuous functions on $(-d, d)$ by

$$\mathfrak{K}(f, g) = \langle f, Kg \rangle \equiv \int_{-d}^{+d} f(x)\,dx \int_{-d}^{+d} K(x - \xi)g(\xi)\,d\xi. \tag{2.3}$$

To this effect one has

Lemma 1. Let $K(t)$ be an even, integrable and square-integrable real function of t on $-\infty < t < \infty$. Also let K be of bounded variation in a neighborhood of almost every t. Let $\bar{K}(z)$, the Fourier transform of K, be positive for all real z. Then the form $\mathfrak{K}(f, f)$ is a positive-definite form, i.e., the right side of (2.3) with $f = g$ is positive unless $f \equiv 0$.

Proof: From the hypotheses of the lemma one has a.e.

$$K(x - \xi) = \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{+\infty} e^{iz(x-\xi)} \bar{K}(z)\,dz,$$

so that

$$2\pi\mathfrak{K}(f, f) = \int_{-d}^{+d} \int_{-d}^{+d} f(x)f(\xi)\,dx\,d\xi \times \int_{-\infty}^{+\infty} e^{iz(x-\xi)} \bar{K}(z)\,dz,$$

and by Fubini's theorem

$$= \int_{-\infty}^{+\infty} \bar{K}(z)\,dz \int_{-d}^{+d} e^{izx} f(x)\,dx \int_{-d}^{+d} e^{-iz\xi} f(\xi)\,d\xi$$

$$= \int_{-\infty}^{+\infty} \bar{K}(z)\,dz \left| \int_{-d}^{+d} e^{izx} f(x)\,dx \right|^2 \geq 0.$$

Moreover, if f is not identically zero, then $\mathfrak{K} > 0$, since otherwise $\mathfrak{K} = 0$ implies

$$\int_{-d}^{+d} e^{izx} f(x)\,dx = 0, \quad \text{a.e.},$$

which implies that f vanishes, a.e., and hence identically because of continuity.

Notice here that Bochner's well-known theorem on positive functions provides a generalization of the above lemma. We state here Bochner's theorem.

Lemma 2. The form $\mathfrak{K}(f, f)$ is nonnegative-definite if and only if $K(t)$ is a characteristic function, i.e., it is the Fourier–Stieljes transform of a (probability) distribution function.

For proof, see Loeve,[14] p. 207.

Corollary. Since the values of $K(t)$ outside $(-2d, 2d)$ are immaterial, we have that $\mathfrak{K} \geq 0$ if and only if $K(t)$ coincides on $(-2d, 2d)$ with a characteristic function.

Examples:

(1) Let

$$K(t) = e^{-|t|^\nu}, \qquad 0 < \nu \leq \infty; \tag{2.4}$$

then we have

(a) $0 < \nu \leq 2$ implies that $\mathfrak{K} > 0$ for all $f \not\equiv 0$;
(b) $\nu > 2$, then $\mathfrak{K}$ is not a positive form.

For proof, see Bochner,[15] p. 96.

(2) Let

$$K(t) = \sum_{n=1}^{N} \alpha_n e^{-\beta_n |t|}, \tag{2.5}$$

where $N < \infty$, $\alpha_n, \beta_n > 0$. Then the Fourier transform

$$\bar{K}(z) = \sum_{n=1}^{N} \frac{2\alpha_n\beta_n}{1 + z^2} > 0.$$

Therefore, Lemma 1 applies and we have $\mathfrak{K}(f, f) > 0$ unless $f \not\equiv 0$.

(3) Let $K(t)$ be an even function which is bounded, convex, monotonically decreasing, and tends to a nonnegative constant K_∞ as $t \to \infty$; then $\mathfrak{K} > 0$.

The proof rests on the fact that the form (2.3) defined by the kernel $K(t) - K_\infty$ is positive-definite (Titchmarsh,[16] Theorem 124, p. 170), and that the contribution of K_∞ to the form $\mathfrak{K}$ is again nonnegative.

(4) A kernel of special importance is the logarithmic kernel

$$K(t) = \begin{cases} \alpha \log |t|, \ \alpha < 0 & -1 \leq t \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2.6}$$

This kernel approximates some physically important kernels, e.g., the Pippard one in certain cases. The fact that the form $\mathfrak{K}$ defined by the kernel (2.6) is positive-definite may be deduced from the corollary to Lemma 2. The following is a direct proof.

[14] M. Loeve, *Probability Theory* (D. Van Nostrand, Inc., New York, 1955).
[15] S. Bochner, *Lectures on Fourier Integrals* (Princeton University Press, Princeton, New Jersey, 1959).
[16] E. C. Titchmarsh, *Fourier Integrals* (Oxford University Press, London, 1948).

Let

$$I = \alpha \int_{-1}^{+1} f(x)\,dx \int_{-1}^{+1} \log|x-\xi|\, f(\xi)\,d\xi, \qquad (2.7)$$

and expand $\log|t|$ into a cosine series,

$$\log|t| = \sum_0^\infty c_n \cos\frac{n\pi t}{2}, \qquad |t| \leq 2. \qquad (2.8)$$

It is easily proved that $c_n < 0$ for all n. Substituting from (2.8) into (2.7), the following is obtained:

$$I = \alpha \sum_0^\infty c_n(a_n^2 + b_n^2),$$

where a_n, b_n are the Fourier components of f. Since $\alpha < 0$, we have $I > 0$ unless $a_n = b_n = 0$, i.e., unless $f \equiv 0$.

(5) Let $K(t) = e^{-\frac{1}{2}t^2}(\sin t/t)$; then $\mathfrak{K}$ is positive-definite, as is easily seen. This example shows that the kernel can change its sign and yet define a positive-definite form.

Before proving the first uniqueness theorem, we give a definition of "regular" kernels.

Definition: A kernel $K(t)$ is called regular on $(-d, +d)$ if it satisfies either (i) the hypotheses of Lemma 1 or (ii) is nonnegative on $(-2d, 2d)$, positive on a subset of positive measure, and is a characteristic function.

Theorem 1 (Uniqueness of diffuse-scattering solution).

If $K(t)$ is a regular kernel, then the solution to the problem defined by (2.1), (2.2), if it exists, is also unique.

Proof: Since K is real, it is sufficient to consider real solutions A of (2.1), (2.2). Let there be two solutions $A_{1,2}$ and set $u(x) = A_1 - A_2$; then

$$\int_{-d}^{d} u\frac{d^2u}{dx^2}\,dx = u\frac{du}{dx}\Big|_{-d}^{d} - \int_{-d}^{d}\left(\frac{du}{dx}\right)^2 dx.$$

Since du/dx vanishes on the boundary, we get by substituting from (2.1)

$$\int_{-d}^{d} u(x)\,dx \int_{-d}^{d} K(x-\xi)u(\xi)\,d\xi = -\int_{-d}^{d}\left(\frac{du}{dx}\right)^2 dx. \qquad (2.9)$$

The left side of (2.9) is nonnegative and therefore both sides are equal to zero. Now if $K(t)$ satisfies property (i) of regular kernels, then the vanishing of the left side of (2.9) implies that $u \equiv 0$. If K satisfies property (ii), then u is a constant, which must vanish because the integral of K is positive. Hence $u \equiv 0$ and $A_1 = A_2$.

Note: The main property of regular kernels which was used in proving the uniqueness is the positivity of its Fourier transform. This is equivalent to having $\chi(q)$ of Sec. 1 negative (and relatively smooth). But this fact just states that the susceptibility of a superconductor is negative at all wavelengths, which expresses the diamagnetic property of superconductors.

To prove the existence of the solution to (2.1), (2.2), one makes use of the following two results from the theory of compact operators.

Lemma 3. Let $K(t)$ belong to $L_2(-d, d)$; then the operator

$$Qf(x) = \int_{-d}^{d} K(x-\xi)f(\xi)\,d\xi$$

is completely continuous on L_2.

For proof, see Ref. 17, p. 179.

Lemma 4. Let T denote the operator $-d^2/dx^2 + Q$ together with the boundary conditions $du/dx = 0$ [more precisely, the self-adjoint extension of the operator in $L_2(-d, d)$]. Then the spectrum of T is a pure point spectrum with no limit points.

The lemma follows from Weyl's classical theorem (Ref. 13, p. 367) which states the invariance of the limit points of a self-adjoint operator under the perturbation by a compact operator.

Theorem 2 (Existence of diffuse scattering solution). If $K(t)$ is regular, then there exists a unique solution to (2.1), (2.2).

Proof: Let $v(x)$ be any fixed smooth function which satisfies the conditions $dv/dx|_{x=\pm d} = H_{1,2}$. Set $w(x) = A(x) - v(x)$; then w satisfies

$$Tw(x) = (-\nabla^2 + Q)w(x) = (\nabla^2 - Q)v(x) \equiv f(x), \qquad (2.10)$$

and

$$\frac{dw}{dx}\Big|_{x=\pm d} = 0.$$

The preceding lemma shows that the resolvent operator $(T - \lambda I)^{-1}$ exists except if λ belongs to the point spectrum of T. The uniqueness theorem shows that $\lambda = 0$ is not an eigenvalue and hence

[17] F. Riesz and B. Nagy, *Functional Analysis* (Frederic Ungar Publishing Company, New York, 1955).

the solution of (2.1), (2.2) exists and is given by $v(x) + T^{-1}f(x)$.

Alternative proof: It is possible to prove the existence of A without appealing to Weyl's theorem, as follows:

Let $G(x, \xi)$ be the generalized Green's function which satisfies

$$(d^2/dx^2)G(x, \xi) = \delta(x - \xi) - 1/2d, \tag{2.11a}$$

$$dG/dx|_{x=\pm d} = 0, \tag{2.11b}$$

and

$$\int_{-d}^{+d} G(x, \xi)\, dx = 0. \tag{2.11c}$$

Then, using Green's formula, one gets the following integral equation for A:

$$A(x) = A_0(x) + \int_{-d}^{d} P(x, \xi)A(\xi)\, d\xi, \tag{2.12}$$

where $A_0(x) = H_1G(x, -d) - H_2G(x, +d)$, and

$$P(x, \xi) = \int_{-d}^{d} \left[G(x, z)K(z, \xi) + \frac{1}{2d} \right] dz.$$

It is easy to show that the Fredholm alternative holds for Eq. (2.12) because of the smoothness of P. On the other hand, the homogeneous version of (2.12) is equivalent to the differential problem (2.1) with zero boundary conditions, and therefore has no nonzero solutions by Theorem 1. Thus, there exists a unique solution to (2.12) which, again because of the smoothness of P, satisfies the problem (2.1), (2.2).

Remarks on higher-dimensional problems: The preceding discussion can be immediately generalized to more dimensions. In such a case one is concerned with equations of the form

$$\nabla^2 A(\mathbf{r}) = \int_\Omega K(\mathbf{r} - \mathbf{s})A(\mathbf{s})\, d\mathbf{s},$$

with boundary condition $\partial A/\partial n = f$, given on $\partial\Omega$. If the kernel K satisfies the conditions of Lemma 1 (positive Fourier transform and integrability and smoothness restrictions), then Lemma 1 holds in any dimension and the uniqueness theorem is again true. Both procedures for proving the existence of the solution in one dimension give similar results in higher dimensions. All details are omitted because, as mentioned in Sec. 1, diffuse scattering may not be physically meaningful in such generality.

B. Generalized Specular Reflection

Using the same notation as in Sec. 1, we now consider the uniqueness and existence questions for the vector problem.

$$[\text{curl curl } \mathbf{A}(\mathbf{r})]_l = \frac{4\pi}{c} \mathbf{J}_l(\mathbf{r}), \quad \text{and by (1.13)}$$

$$= \frac{4\pi}{c} \int_\Omega \sum_{m=1}^{3} K_{lm}(\mathbf{r}, \mathbf{s})A_m(\mathbf{s})\, d\mathbf{s}, \tag{2.13}$$

$$\text{div } \mathbf{A} = 0, \tag{2.14}$$

$$\nu \cdot \mathbf{A} = 0, \quad \text{on} \quad \partial\Omega, \tag{2.15}$$

$$\text{curl } \mathbf{A} = \mathbf{H}(\mathbf{t}) \quad \text{given on} \quad \partial\Omega, \tag{2.16}$$

where ν is the unit outward normal to $\partial\Omega$, and $\mathbf{H}(\mathbf{t})$ is assumed to be continuous. The specular reflection kernel is defined by [see Eq. (1.22)]

$$K_{lm}(\mathbf{r}, \mathbf{s}) = \sum_n [\mathbf{u}_n(r)]_l \chi(\lambda_n)[\mathbf{u}_n(\mathbf{s})]_m. \tag{2.17}$$

Equation (2.13) can be written in the abbreviated form

$$\text{curl curl } \mathbf{A} = \mathfrak{K}\mathbf{A}, \tag{2.13a}$$

where $\mathfrak{K}$ is a dyadic (integral) operator with a kernel K defined by (2.17).

The uniqueness and existence proofs for the problem (2.13)–(2.17) are given below.

Lemma 5. If the vector field $\mathbf{A} \in C^{(2)}(\Omega)$ and satisfies the problem (2.13)–(2.17) with $\mathbf{H}(\mathbf{t})$ being equal to zero, then the assumption $\chi < 0$ implies that $\mathbf{A} \equiv 0$.

Proof: It is sufficient to consider real $\mathbf{A}$. Using Green's vector identity one gets

$$\int_\Omega \mathbf{A} \cdot (\text{curl curl } \mathbf{A})\, d\Omega$$

$$= \int_\Omega |\text{curl } \mathbf{A}|^2\, d\Omega + \int_{\partial\Omega} (\text{curl } \mathbf{A} \times \mathbf{A}) \cdot \nu\, dS, \tag{2.18}$$

where dS is a surface element.

If $\mathbf{A}$ satisfies the hypotheses of the lemma, then (2.18) yields

$$\int_\Omega |\text{curl } \mathbf{A}|^2\, d\Omega = \int_\Omega \mathbf{A} \cdot (\mathfrak{K}\mathbf{A})\, d\Omega. \tag{2.19}$$

Expand $\mathbf{A}$ in the complete set of eigenfunctions $\mathbf{u}_n(\mathbf{r})$ defined by (1.15)–(1.18),

$$\mathbf{A} = \sum_n A_n \mathbf{u}_n(\mathbf{r}). \tag{2.20}$$

Let I denote the right side of (2.19); then

$$\frac{c}{4\pi} I = \int_r \sum_{i=1}^{3} \left\{ \sum_{m=1}^{\infty} A_m u_m^i(\mathbf{r}) \right.$$

$$\left. \times \int_s \sum_{j=1}^{3} K_{ij}(\mathbf{r}, \mathbf{s}) \sum_{n=1}^{\infty} A_n u_n^j(\mathbf{s})\, d\mathbf{s} \right\} d\mathbf{r}$$

$$= \int_r \sum_i \sum_m A_m u_m^i(\mathbf{r}) \int_s \sum_j \Big(\sum_p \chi(\lambda_p) u_p^j(\mathbf{r}) u_p^j(\mathbf{s})\Big) \times \sum_n A_n u_n^j(\mathbf{s})\, ds\, dr.$$

Using the orthogonality of the eigenfunctions u_n, we get

$$\frac{c}{4\pi} I = \sum_{n=1}^{\infty} A_n^2 \chi(\lambda_n).$$

Hence, substituting in (2.19) and noting that χ is negative, one has $A_n = 0$ for all n, so A vanishes identically.

Notice that the above proof of uniqueness is very similar to that of Lemma 1 and Theorem 1 except that we use a "generalized Fourier representation" for both the kernel and the solution **A**. Similar to Lemma 3, one has

Lemma 6. The dyadic operator $\mathfrak{K}\mathbf{A} = \int KA$ is completely continuous on the Hilbert space H of vector-valued functions which belong to $L_2(\Omega)$.

The proof may be sketched as follows: Take $u_n(\mathbf{r})$ as an orthonormal basis of H. If $\mathbf{A} = \sum A_n \mathbf{u}_n(\mathbf{r})$; then the correspondence $\mathbf{A} \sim \{A_n\}$ defines an isometry between $L_2(\Omega)$ and l_2. To the operator $\mathfrak{K}\mathbf{A}$ then corresponds the series $\{\chi(\lambda_n)A_n\}$. Since $\{\chi(\lambda_n)\} \in l_2$ [see (1.22)], the map: $\{A_n\} \to \{\chi(_n)A_n\}$ is completely continuous and therefore the dyadic $\mathfrak{K}$ defines a compact operator on L_2 by isometry.

Theorem 3 (Existence and uniqueness of specular reflection solution). There exists a unique solution for the vector problem (2.13)–(2.17).

Sketch of the proof: Let $\mathbf{v}$ be a fixed smooth function with zero divergence which satisfies the conditions

$$\nu\cdot\mathbf{v} = 0, \qquad \operatorname{curl} \mathbf{v} = \mathbf{H}(t) \quad \text{on} \quad \partial\Omega,$$

and let $\mathbf{w} = \mathbf{A} - \mathbf{v}$. If T denotes the operator

$$T\mathbf{w} = [\operatorname{curl}\operatorname{curl} - \mathfrak{K}]\mathbf{w}$$

together with the conditions

$$\operatorname{div} \mathbf{w} = 0,$$

$$\nu\cdot\mathbf{w} = 0 \quad \text{and} \quad \nu\cdot\operatorname{curl}\mathbf{w} = 0 \quad \text{on} \quad \partial\Omega,$$

then T defines a self-adjoint operator on L_2 with pure point spectrum because $\mathfrak{K}$ is completely continuous. The vector problem becomes then equivalent to $T\mathbf{w} = \mathbf{f}$, where $\mathbf{f}$ is a known smooth function derived from $\mathfrak{K}$ and $\mathbf{v}$. The uniqueness property (Lemma 5) then proves the existence of $\mathbf{w}$, and hence of $\mathbf{A}$ in a similar fashion to Theorem 2.

Remark: Both Theorems 2 and 3 prove the existence of the solution when the magnetic field is known on the boundary. In practice, only the applied field, which is distorted by the superconducting body, is known. The complete problem reduces then to an exterior–interior problem for the vector potential such that the asymptotic behavior of the potential is known at infinity. However, only minor changes in the methods used above, similar to those of Odeh,[18] are needed for the more general case.

C. The Time-Dependent Problem

The time-dependent problem, discussed in Sec. 1, may be summarized as follows: To find a smooth vector-valued function $\mathbf{A}(\mathbf{r}, t)$ defined over $\Omega \times [0, \infty)$ where Ω has a smooth boundary $\partial\Omega$ such that $\mathbf{A}$ satisfies

$$\frac{\partial \mathbf{A}}{\partial t} = -\operatorname{curl}\operatorname{curl}\mathbf{A} + \int_\Omega \mathfrak{K}(\mathbf{r}, \mathbf{s})\mathbf{A}(\mathbf{s}, t)\, d\mathbf{s}, \qquad r \in \Omega, \quad t > 0, \tag{2.21}$$

$$\operatorname{div}\mathbf{A} = 0, \tag{2.22}$$

$$\nu\cdot\mathbf{A} = 0 \quad \text{on} \quad \partial\Omega, \tag{2.23}$$

$$\operatorname{curl}\mathbf{A} = \mathbf{H}(\mathbf{r}, t) \quad \text{on} \quad \partial\Omega, \tag{2.24}$$

$$\mathbf{A}(\mathbf{r}, 0) = \mathbf{f}(\mathbf{r}) \quad \text{on} \quad \Omega, \tag{2.25}$$

$$|\mathbf{A}(\mathbf{r}, t)| \leq M < \infty \quad \text{for all time}, \tag{2.26}$$

where $\mathbf{H}$, $\mathbf{f}$ are given smooth functions. For simplicity in notation, we consider the diffuse scattering case in one dimension. Then the dyadic $\mathfrak{K}$ reduces to a multiplication by a scalar function $K^0(x - \xi)$ and the region Ω reduces to an interval I of the x axis. The generalizations to the vector problem are easily made as in part B of this section. We have now to consider the problem

$$\frac{4\pi\sigma}{c^2}\frac{\partial A}{\partial t} = \frac{d^2A}{dx^2} + \frac{4\pi}{c}\int_{-d}^{+d} K(x - \xi)A(\xi, t)\, d\xi, \qquad x, \xi \in I, \tag{2.27}$$

$$\partial A/\partial x|_{x=\pm d} = H_{1,2}(t), \qquad t \geq 0, \tag{2.28}$$

$$A(x, 0) = f(x), \qquad x \in I, \tag{2.29}$$

$$|A(x, t)| \leq M < \infty, \qquad t \geq 0. \tag{2.30}$$

Theorem 4. The initial boundary-value problem (2.27)–(2.30) possesses a unique solution for all continuous data $H_{1,2}$ and f.

Proof: Let Q denote the integral operator with kernel K defined on $L_2(I)$ and let $T = d^2/dx^2 + Q$.

[18] F. Odeh, J. Math. Phys. **4**, 1, 141 (1963).

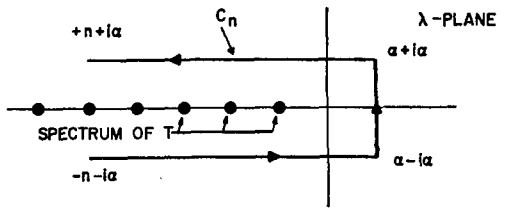


FIG. 1. The contour for the integral representation of $U_n(t)$.

Consider the problem (2.27)–(2.30) defined on $L_2(I) \times C^{(1)}[0, \infty)$. By subtracting an appropriate solution of the heat equation from A, one can reduce the problem to finding a function u such that

$$\partial u/\partial t = Tu + g(x, t), \qquad x \in I, \tag{2.31}$$

$$\partial u/\partial x|_{x=\pm d} = 0, \tag{2.32}$$

$$u(x, 0) = f(x), \tag{2.33}$$

where g is a known function. To prove the existence of u, notice first that $(T - \lambda I)^{-1}$ exists for all $Rl\ \lambda \geq 0$ and, in fact, that the whole spectrum of T lies on the negative real axis since the operators d^2/dx^2 under the restriction (2.32) and Q are both negative (see Fig. 1). We distinguish now two cases:

(1) $g(x, t) = 0$. In this case one has a homogeneous "diffusion" equation and the following result holds: Let C_n denote the contour $(-n - i\alpha, \alpha - i\alpha, \alpha + i\alpha, -n + i\alpha)$ where $\alpha > 0$, and let

$$U_n(t) = \frac{1}{2\pi i} \int_{C_n} (\lambda I - T)^{-1} e^{\lambda t}\, d\lambda.$$

Then the operators $U_n(t)$ converge in the uniform operator topology to an operator $U(t)$ for all $t > 0$, and $U_n(0)$ converge strongly to the identity operator.

This result, which is a particular case of Theorem 12.5.1 in Ref. 19 (see also Ref. 20, p. 136), shows that the solution to (2.31), (2.32), (2.33) is given by

$$u(x, t) = U(t)f(x).$$

(2) $g(x, t) \not\equiv 0$. Since g is smooth and bounded, in the L_2 norm, uniformly in t, the variation of parameters method shows that

$$u(x, t) = \lim_{n\to\infty} U_n(t)\left[f(x) + \int_0^t U_n(-\tau)g(x, \tau)\, d\tau\right]$$

$$= U(t)f(x) + \int_0^t U(t - \tau)g(x, \tau)\, d\tau.$$

This concludes the proof.

[19] F. Hille, *Functional Analysis and Semi-Groups* (American Mathematical Society Colloquium Publications, Providence, Rhode Island, 1948), Vol. 31.

[20] G. Birkhoff and E. P. Wigner, editors, "Nuclear Reactor Theory," Proc. Symp. Appl. Math., 11, (1961).

Notes: (i) It is possible to prove the above theorem by use of Laplace transforms, but the above proof was adopted because it generalizes immediately to the vector (and specular reflection) problem in higher dimensions.

(ii) An expansion of the solution to (2.27) in terms of the eigenfunctions of T, and exponential functions in time, can be proved. Such an expansion would be helpful in studying the effect of nonlocality on transition problems.

3. APPROXIMATE AND SPECIAL SOLUTIONS

In this section some approximate solutions to the static equations (2.1), (2.2) will be considered in the diffuse scattering case. The Pippard kernel is used in two cases, small coherence lengths and thin films. This latter case was considered by Peter,[6] but a simpler convergent representation for the solution will be given. It is also shown that the generalized specular reflection reduces to Schrieffer's solution[5] in the special case of the slab which he considered. A series solution is then given for the case of general specular reflection on the boundary of a circular cylinder in an axial field.

A. Diffuse Scattering Approximations

1. *The Boundary-Layer Solution of Pippard's Model with Small Coherence Length*

Consider a superconducting film, infinite in the y, z directions, which lies between $-d$, $+d$ in the x direction. Using the Pippard kernel (1.4), Eq. (2.1) reduces to

$$\frac{d^2\mathfrak{a}}{dt^2} = \frac{3}{4\lambda^2\xi_0} \int_{-d}^{+d} K\left(\frac{|t - t'|}{\xi}\right)\mathfrak{a}(t')\, dt', \tag{3.1}$$

where ξ, ξ_0 are defined in (1.4), λ is the London penetration depth, and K is defined by

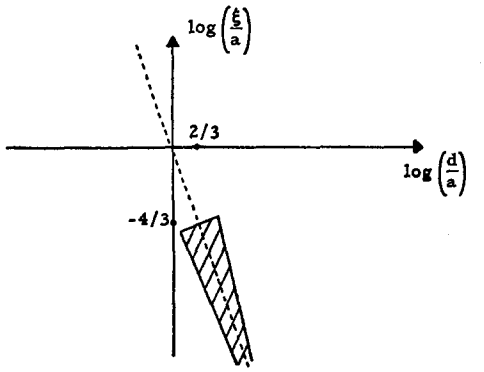


FIG. 2. Region of validity for the asymptotic expansion.

$$K(s) = \int_1^\infty \left(\frac{1}{t} - \frac{1}{t^3}\right) e^{-|s|t}\, dt = \mathrm{Ei}_1(s) - \mathrm{Ei}_3(s), \quad (3.2)$$

where the Ei's are the usual exponential integrals.

The applied magnetic field(s) are assumed known on the boundary, so that

$$d\mathfrak{A}/dt|_{t=\pm d} = H_{1,2}. \quad (3.3)$$

For the sake of simplicity in the notation which follows, we consider the symmetric case $H_1 = H_2 = H$; the general case may be treated similarly. Introduce the dimensionless variables

$$x = t/d, \qquad A(x) = (1/Hd)\mathfrak{A}(t). \quad (3.4)$$

Then (3.1) reduces to

$$\frac{d^2A}{dx^2}(x) = \frac{3}{4}\frac{d^2}{\lambda^2}\frac{\xi}{\xi_0}\frac{d}{\xi}\int_{-1}^{+1} K\left\{\frac{x-x'}{(\xi/d)}\right\}A(x')\,dx', \qquad dA/dx|_{x=\pm 1} = 1. \quad (3.5)$$

Let $\xi/d = \epsilon$ and $\frac{3}{4}(d^2/\lambda^2)(\xi/\xi_0) = c$. Consider the case when $\epsilon \to 0$ while c remains finite and positive, $0 < c < \infty$. This means that $d \gg \xi$ while $d^2 = O[\lambda^2(\xi_0/\xi)]$. If we set $\lambda^2\xi_0 = a^3$, then the analysis will cover these cases where both the conditions $d/a = O(\epsilon^{-\frac{1}{3}}) \gg 1$ and $(\xi/a) = O(\epsilon^{\frac{1}{3}}) \ll 1$ are satisfied. The shaded area in Fig. 2 shows the region of validity for $\epsilon \leq 10^{-2}$.

Equation (3.5) may now be written as

$$\frac{d^2A}{dx^2} = c\int_{-1}^{+1}\frac{1}{\epsilon}K\left\{\frac{x-x'}{\epsilon}\right\}A(x')\,dx'. \quad (3.5a)$$

To investigate the solution to (3.5a) as $\epsilon \to 0$, one notices that, in this limit, the kernel $(1/\epsilon)K\{(x-x')/\epsilon\}$ becomes proportional to a delta function in the interior of $(-1, +1)$. However, if x approaches the boundary, a major part of the area under the kernel lies outside the interval of integration (see Fig. 3). Hence, the right side of (3.5a) will clearly change appreciably when x approaches the boundary. This is seen more clearly from Lemma 7 below. In order to find this "edge effect" one can add a boundary-layer solution obtained from the solution to a half-infinite problem by a Wiener–Hopf technique, as in Ref. 10, to the solution valid in the interior. Such a procedure is suggested by Carrier.[21] However, the degree of approximation thus obtained is not clear and, moreover, the half-infinite solution is rather complicated. Here we use instead a technique, systematized by Miranker,[22] of singular perturbations by a method of undetermined coefficients. First an asymptotic expansion for the right side of (3.5a) in terms of powers of ϵ will be given.

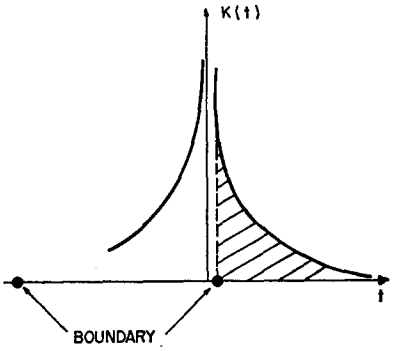


FIG. 3. Behavior of the kernel near the boundary.

Lemma 7. Let $A(x) \in C^\infty(x)$ on $[-1, +1]$; then the right side, R, of (3.5a) has the following asymptotic expansion:

$$R = c\sum_{m=0}^{\infty}\epsilon^{2m}a_{2m}A^{(2m)}(x) + ce^{-(1-x)/\epsilon}\sum_{m=0}^{\infty}\epsilon^m f_m(x)A^{(m)}(1) + ce^{-(x+1)/\epsilon}\sum_{m=0}^{\infty}\epsilon^m g_m(x)A^{(m)}(-1), \quad (3.6)$$

where the a_{2m} are known constants and $f_m(x)$, $g_m(x)$ are known smooth bounded functions.

Proof: Consider

$$\frac{R}{c} = \frac{1}{\epsilon}\int_{-1}^{+1}K\left(\frac{|x-z|}{\epsilon}\right)A(z)\,dz = \frac{1}{\epsilon}\int_{-1}^{+1}A(z)\,dz\int_1^\infty\left(\frac{1}{t}-\frac{1}{t^3}\right)e^{-|x-z|t/\epsilon}\,dt.$$

Interchanging order of integration, by Tonelli's theorem, one obtains

$$\frac{R}{c} = \frac{1}{\epsilon}\int_1^\infty\left(\frac{1}{t}-\frac{1}{t^3}\right)dt\int_{-1}^{+1}e^{-|x-z|t/\epsilon}A(z)\,dz. \quad (3.7)$$

Let

$$I(x, t; \epsilon) = \int_{-1}^{+1}e^{-|x-z|t/\epsilon}A(z)\,dz. \quad (3.8)$$

The asymptotic behavior of $I(\epsilon)$ is obtained by Laplace's method,

$$I = \int_{-1}^{x}e^{-(x-z)t/\epsilon}A(z)\,dz + \int_x^1 e^{-(z-x)t/\epsilon}A(z)\,dz = I_1 + I_2,$$

[21] G. F. Carrier, "Integral Equation Boundary Layer Problems," *50 Jahre Grenzschichtforschung* (Verlag Friedr. Vieweg und Sohn, Braunschweig, Germany, 1954).

[22] W. Miranker, J. Math. and Phys. 42, 47 (1963).

$$I_1 = \sum_{n=0}^{\infty} \left(\frac{\epsilon}{t}\right)^{n+1} (-1)^n A^{(n)}(z) e^{-(x-z)t/\epsilon} \Big|_{-1}^{z}$$
$$= \sum_{n=0}^{\infty} \left(\frac{\epsilon}{t}\right)^{n+1} (-1)^n A^{(n)}(x) - \left[\sum_{n=0}^{\infty} \left(\frac{\epsilon}{t}\right)^{n+1} (-1)^n A^{(n)}(-1)\right] e^{-(x+1)t/\epsilon},$$

where $A^{(n)}$ denotes the nth derivative of A. Similarly,

$$I_2 = \sum_{n=0}^{\infty} \left(\frac{\epsilon}{t}\right)^{n+1} A^{(n)}(x) - \left[\sum_{n=0}^{\infty} \left(\frac{\epsilon}{t}\right)^{n+1} A^{(n)}(+1)\right] e^{-(1-x)t/\epsilon}.$$

Adding and substituting in (3.7), we get

$$\frac{R}{c} = \frac{1}{\epsilon} \int_1^{\infty} \left(\frac{1}{t} - \frac{1}{t^3}\right) dt \left\{ \sum_{m=0}^{\infty} 2\left(\frac{\epsilon}{t}\right)^{2m+1} A^{(2m)}(x) - \left[\sum_{m=0}^{\infty} \left(\frac{\epsilon}{t}\right)^{m+1} A^{(m)}(1)\right] e^{-(1-x)t/\epsilon} - \left[\sum_{m=0}^{\infty} \left(\frac{\epsilon}{t}\right)^{m+1} (-1)^m A^{(m)}(-1)\right] e^{-(1+x)t/\epsilon} \right\}. \tag{3.9}$$

Comparing (3.9) with (3.6), we have

$$a_{2m} = 2 \int_1^{\infty} \left(\frac{1}{t} - \frac{1}{t^3}\right)\left(\frac{1}{t}\right)^{2m+1} = 2\left\{\frac{1}{2m+1} - \frac{1}{2m+3}\right\}. \tag{3.10}$$

The coefficients of the boundary layer $e^{-(1-x)/\epsilon}$ are obtained by integrating the second term in (3.9) to give

$$f_m(x) = -e^{(1-x)/\epsilon} \times \left\{ \mathrm{Ei}_{m+2}\left(\frac{1-x}{\epsilon}\right) - \mathrm{Ei}_{m+4}\left(\frac{1-x}{\epsilon}\right) \right\}, \tag{3.10a}$$

which is, by the properties of exponential integrals,

$$\left\{ \frac{\epsilon}{(1-x)} \int_0^{\infty} e^{-(1-x)u/\epsilon} \times \left[\frac{m+2}{(1+u)^{m+3}} - \frac{(m+4)}{(1+u)^{m+5}}\right] du \right\}. \tag{3.11}$$

Notice that $f_m(1) = -2/(m+3)(m+1)$. Hence, $f_m(x)$ is a function which is $O(\epsilon)$ if x is not near the boundary but falls smoothly to a negative value as $x \to 1$. Similarly,

$$g_m(x) = (-1)^{m+1} e^{-(1+x)/\epsilon} \times \left\{ \mathrm{Ei}_{m+2}\left(\frac{1+x}{\epsilon}\right) - \mathrm{Ei}_{m+4}\left(\frac{1+x}{\epsilon}\right) \right\}. \tag{3.12}$$

The function g_m exhibits a similar behavior to f_m; in fact, $g_m(z-1) = (-1)^m f_m(1-z)$. Equations (3.6), (3.10), (3.11), (3.12) define the asymptotic expansion completely. From the form of the expansion, it is clear that the solution has a boundary layer behavior which we assume to be of the form (suggested by W. Miranker)

$$A(x) \sim U(x) + V(x) + W(x) \tag{3.13}$$

where

$$U(x) \sim \sum_{n=0}^{\infty} \epsilon^n u_n(x), \tag{3.14}$$

$$V(x) \sim e^{-(1-x)/\epsilon} \sum_{n=0}^{\infty} \epsilon^n v_n(x), \tag{3.15}$$

$$W(x) \sim e^{-(x+1)/\epsilon} \sum_{n=0}^{\infty} \epsilon^n w_n(x). \tag{3.16}$$

All terms which are $\leq o(e^{-1/\epsilon})$ are neglected. To determine the coefficients $\{u_n, v_n, w_n\}$ one substitutes (3.13)–(3.16) into (3.5a). Use is made of Lemma 7 in determining the right side of (3.5a). Only one of the two boundary layers, U, will be discussed since W can be found similarly. In fact, in the symmetric case, since A must be odd, W is found from U by inspection. The orders are matched on both sides of (3.5a) as follows:

Step −1: $O[(1/\epsilon^2)e^{-(1-x)/\epsilon}]$ gives

$$v_0(x) = 0; \quad \text{similarly } w_0(x) = 0. \tag{3.17}$$

Step 0: $O[(1/\epsilon)e^{-(1-x)/\epsilon}]$, $O(\epsilon^0)$ give

$$v_1(x) = 0, \qquad w_1(x) = 0, \tag{3.18}$$

and

$$(d^2u_0/dx^2)(x) = ca_0u_0, \tag{3.19}$$

with

$$du_0/dx|_{x=\pm 1} = 1,$$

where a_0 is given by (3.10). Hence,

$$u_0(x) = \frac{\sinh (a_0c)^{\frac{1}{2}}x}{(a_0c)^{\frac{1}{2}} \cosh (a_0c)^{\frac{1}{2}}}. \tag{3.19a}$$

Step 1: $O(e^{-(1-x)/\epsilon})$, $O(\epsilon)$. Now the first boundary layer appears, namely $V_2(x) = \epsilon^2 e^{-(1-x)/\epsilon} v_2(x)$ and a similar $W_2(x)$. If V_2, W_2 are added to the solution $u_0(x)$, then the boundary conditions in (3.19) are spoiled by an amount equal to the derivative of the boundary layers, which is $O(\epsilon)$. Hence, by adding also the interior solution ϵu_1 whose boundary values exactly cancel those arising from V_2, W_2, the boundary conditions are satisfied exactly and Eq. (3.5a)

is satisfied up to, and including, $O(\epsilon)$. To determine $v_2(x)$ we equate coefficients of $e^{-(1-x)/\epsilon}$ to get, after some calculation,

$$v_2(x) = cf_0(x)u_0(1), \tag{3.20}$$

where f_0 is given by (3.10a) and

$$dV_2/dx|_{x=1} = \epsilon v_2(1) + O(\epsilon^2). \tag{3.20a}$$

Similarly,

$$dW_2/dx|_{x=-1} = -\epsilon w_2(1) = \epsilon v_2(1).$$

[From symmetry $v_n(x) = -w_n(x)$; this can be checked directly.] Now $u_1(x)$ is obtained by equating $O(\epsilon)$ and using the negative of the boundary conditions (3.20a), i.e.,

$$\begin{aligned} d^2u_1/dx^2(x) &= ca_0u_1, \\ du_1/dx|_{x=\pm 1} &= -v_2(1) = b_1. \end{aligned} \tag{3.21}$$

Hence

$$u_1(x) = b_1 \sinh (a_0c)^{\frac{1}{2}}x/(a_0c)^{\frac{1}{2}} \cosh (a_0c)^{\frac{1}{2}}. \tag{3.21a}$$

Step 2: $O(\epsilon e^{-(1-x)/\epsilon})$, $O(\epsilon^2)$. Repeating the same arguments as in Step 1, one gets

$$v_3(x) + 2v_2'(x) = f_0(x)u_1^{(0)}(1) + f_1(x)u_0^{(1)}(1), \tag{3.22}$$

which gives

$$v_3(x) = f_0(x)u_1(1) + f_1(x)u_0^{(1)}(x) - 2cu_0(1)f_0'(x). \tag{3.22a}$$

The function $u_2(x)$ is found by equating $O(\epsilon^2)$,

$$\begin{aligned} d^2u_2/dx^2 &= c[a_0u_2(x) + a_2u_0^{(2)}(x)], \\ du_2/dx|_{x=\pm 1} &= -(v_3(x) + v_2'(x))|_{x=1} = b_2, \end{aligned} \tag{3.23}$$

which determines u_2.

An induction argument, similar to the one used in Ref. 21 now proves that $\{v_{n+1}, w_{n+1}, u_n\}$ can be determined for all n, and the errors involved are then of $O(\epsilon^n)$. Now a few remarks about the solution:

(i) The first approximation, $u_0(x)$, defined by (3.19), is just the London solution, as expected.

(ii) Apart from the boundary-layer terms, $\{v_n, w_n\}$, the asymptotic solution considered above is a superposition of London-type solutions. This fact partially justifies the suggestions of Tinkham[23] and Ittner[24] concerning the variation of the parameters in London's theory to account for the nonlocal effects. It is not clear, however, that such a procedure would yield a good approximation when ϵ becomes large, as may be seen from the next paragraph.

(iii) The solution $A(x)$, correct to $O(\epsilon^2)$ is

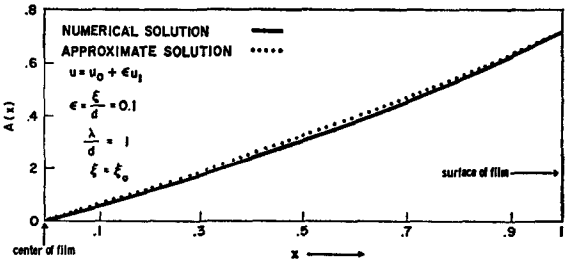


FIG. 4. Comparison of the small-coherence-length approximation with the numerical calculation of Odeh and Liniger.

$$\begin{aligned} A(x) &= \frac{\sinh (a_0c)^{\frac{1}{2}}x}{(a_0c)^{\frac{1}{2}} \cosh (a_0c)^{\frac{1}{2}}} \\ &+ \epsilon \frac{b_1 \sinh (a_0c)^{\frac{1}{2}}x}{(a_0c)^{\frac{1}{2}} \cosh (a_0c)^{\frac{1}{2}}} + \epsilon^2 \Bigg\{ \Bigg[\frac{b_2 \sinh (a_0c)^{\frac{1}{2}}x}{(a_0c)^{\frac{1}{2}} \cosh (a_0c)^{\frac{1}{2}}} \\ &- \frac{\cosh (a_0c)^{\frac{1}{2}}(x-1)}{(a_0c)^{\frac{1}{2}} \sinh 2(a_0c)^{\frac{1}{2}}} \int_{-1}^{x} (a_2c) \cosh (1+\xi)^{\frac{1}{2}} \\ &\times u_0^{(2)}(\xi)\, d\xi - \frac{\cosh (a_0c)^{\frac{1}{2}}(x+1)}{(a_0c)^{\frac{1}{2}} \sinh 2(a_0c)^{\frac{1}{2}}} \\ &\times \int_{x}^{1} (a_2c) \cosh (\xi-1)^{\frac{1}{2}} u_0^{(2)}(\xi)\, d\xi \Bigg] \\ &+ ce^{-(1-x)/\epsilon} f_0(x)u_0(1) - ce^{-(1-x)/\epsilon} f_0(x)u_0(1) \Bigg\} + O(\epsilon^3), \end{aligned}$$

where $u_0(x)$ is given by (3.19a) and b_2 is defined by (3.23), (3.22a). The approximation $u_0(x) + \epsilon u_1(x)$ is plotted in Fig. 4, where it is compared with the exact numerical solution obtained by Liniger and Odeh.[25]

2. *The Thin-Film Approximation*

Consider now the case when the thickness d is small with respect to either the coherence length or the penetration depth. If $\xi/d = \alpha$, then Eq. (3.5) may be written as

$$\frac{d^2A}{dx^2} = c\frac{1}{\alpha}\int_{-1}^{+1} K\left\{\frac{|x-x'|}{\alpha}\right\} A(x')\, dx', \tag{3.24}$$

$$dA/dx|_{x=\pm 1} = h_{1,2}, \tag{3.25}$$

where h_1 and h_2 are normalized external fields. If c/α is small, which means that $d^3 \ll \lambda^2\xi$, then the solution is easily found by converting (3.24), (3.25) into a pure integral equation and then using a Neumann expansion for the solution. Let $G(x, \xi)$ be the generalized Green's function for $-d^2/dx^2$, with a vanishing derivative on the boundary, namely

$$G(x, \xi) = \tfrac{1}{2}|x-\xi| - \tfrac{1}{4}(x^2+\xi^2) - \tfrac{1}{6}. \tag{3.26}$$

By the use of Green's theorem and the facts that

[23] M. Tinkham, Phys. Rev. 110, 26 (1958).
[24] W. B. Ittner, Phys. Rev. Suppl. 111, 148 (1958).
[25] W. Liniger and F. Odeh, Phys. Rev. 132, 5, 1934 (1963).

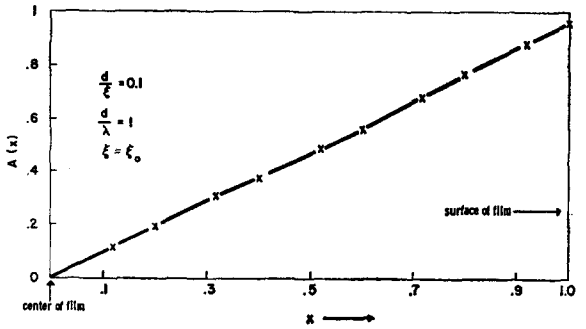


FIG. 5. Comparison of the thin-film approximation with the numerical calculation of Odeh and Liniger.

$d^2G/dx^2 = \delta(x - \xi) - \frac{1}{2}$ and $dG/dx|_{x=\pm 1} = 0$, one finds that Eqs. (3.24), (3.25) are equivalent to

$$A(x) = f(x) + m + \frac{c}{\alpha}\int K^{(1)}(x, x')A(x')\,dx', \quad (3.27)$$

where $f(x) = h_1G(x, -1) - h_2G(x, 1)$, m is the average of A, and

$$K^{(1)}(x, x') = \int_{-1}^{+1} G(x, \xi)K\left(\frac{\xi - x'}{\alpha}\right)d\xi. \quad (3.28)$$

If c/α is small enough, then (3.27) may be solved by a direct interation with m being treated as a parameter. The solution of (3.27) will then be a function of x and m which must satisfy

$$\frac{1}{2}\int_{-1}^{+1} A(x, m)\,dx = m. \quad (3.29)$$

Equation (3.29) defines m and then A is uniquely defined by (3.27). We apply this iteration technique to the simple case when $h_1 = h_2 = 1$. It is more convenient now to subtract the linear part of A. Let $w(x) = A(x) - x$; then w satisfies

$$w(x) = w_0(x) + \frac{c}{\alpha}\int_{-1}^{+1} K^{(1)}(x, x')w(x')\,dx',$$

where

$$w_0(x) = \frac{c}{\alpha}\int_{-1}^{+1} G(x, x')\,dx' \int_{-1}^{+1} K\left(\frac{x' - z}{\alpha}\right)z\,dz. \quad (3.30)$$

If c is finite, i.e., $d = O(\lambda)$, while $\xi \gg d$, one can replace K by its asymptotic value

$$K(z/\alpha) \sim -(\gamma + \tfrac{1}{2}) + \log\alpha - \log|z|, \quad |z/\alpha| \ll 1, \quad (3.31)$$

where γ is Euler's constant. Hence, one gets

$$w_0(x) \sim -\frac{c}{\alpha}\int_{-1}^{+1} G(x, x')\,dx' \times \int_{-1}^{+1} \log|x' - z|\,z\,dz. \quad (3.32)$$

The constant terms, such as $\log\alpha$, do not contribute because G is orthogonal to unity and hence to every constant.

The approximate solution $A_0(x) = x + w_0(x)$ was calculated for the case $\alpha = 10$, $d = \lambda$, $\xi = \xi_0$; hence, $c/\alpha = 0.075$. The result, given in Fig. 5, differs by less than 0.5% from the numerical solution, and the difference is not shown in Fig. 5. A similar problem was treated by Peter,[6] but it is not easy to compare this solution with Peter's because of the difference in the boundary condition used in the two cases.

3. *Remarks on the Intermediate Case*

When the lengths ξ, λ are comparable with the thickness d, there does not seem to exist an easy analytical method for approximating the solution. However, if the kernel $K(t)$ can be approximated by a polynomial in $|t|$, then the resulting equation can be solved exactly. Suppose that $K(t) = \sum_{n=0}^{N} \alpha_n |t|^n$ and substitute in (3.5). Since both the odd and even parts of A are solutions of the integrodifferential equation, each could be considered separately. Consider the odd part for example. Then, by successive differentiation, one can reduce (3.5) to a pure differential equation of order $2(N + 1)$. The assumption that this differential equation should be equivalent to (3.5) provides us with $2N$ additional boundary conditions which determine the solution uniquely. For example, let $N = 1$ and $K(t) = \alpha_0 - \alpha_1|t|$, $\alpha_1 > 0$; then (3.5) may be shown to be equivalent to

$$d^4A/dx^4 + 2\alpha_1 A(x) = 0,$$

$$dA/dx|_{x=\pm 1} = 1, \qquad d^3A/dx^3|_{x=\pm 1} = 0,$$

and A is given by

$$A(x) = L_1 \cos kx \sinh kx + L_2 \sin kx \cosh kx,$$

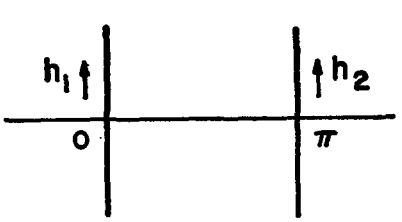


FIG. 6. Specular reflection for film in parallel field.

where $2k^4 = \alpha_1$,

$$L_1 = \frac{(c - s)}{2k(c^2 + s^2)}, \qquad L_2 = \frac{(c - s)}{2k(c^2 + s^2)}$$

and $c = \cos k \cosh k$, $s = \sin k \sinh k$.

B. Specular Reflection

1. The Plane Film

Schrieffer's calculation[5] will now be redone to show how the generalized specular-reflection definition [Eqs. (1.14), (1.22)] yields the same solution obtained by Schrieffer's current-sheets method.

Consider a superconducting film $0 \leq x \leq \pi$ with a parallel external magnetic field in the y direction of magnitude h_1 at $x = 0$, h_2 at $x = \pi$ (Fig. 6). Then there is one component of the vector potential $A_s(x)$, and it satisfies

$$\frac{d^2A}{dx^2} = \int_0^\pi K(x, \xi)A(\xi)\, d\xi \tag{3.33}$$

and

$$dA/dx|_{0,\pi} = h_{1,2}. \tag{3.34}$$

The "specular reflection" kernel $K(x, \xi)$ is defined by (1.22). The eigenfunctions $u_n(x)$ in this case are just $\cos nx$, $n \geq 0$. Hence,

$$K(x, \xi) = -\sum_n \chi(n) \cos nx \cos n\xi \equiv \sum_n k_n \cos nx \cos n\xi. \tag{3.35}$$

Let

$$A(x) = \sum_{n=0}^{\infty} A_n \cos nx, \tag{3.36}$$

where the equality is taken in the L_2 sense. Substituting into (3.33), the right side of that equation reduces to $\sum A_n k_n \cos nx$. However, the series (3.36) cannot be differentiated term by term. In fact, the expansion (3.36) corresponds to extending A to be even and periodic with period 2π. Hence, dA/dx is odd with the same period and possesses jumps of magnitude $2h_1$, $-2h_2$ at $x = 0$ and $x = \pi$, respectively. The second derivative d^2A/dx^2 has to be interpreted then as a symbolic derivative.[26] Then, equation (3.33) yields

$$\Sigma - n^2 A_n \cos nx + 2h_1\delta_p(x) - 2h_2\delta_p(x - \pi) = \Sigma k_n A_n \cos nx, \tag{3.37}$$

[26] B. Friedman, *Principles and Techniques of Applied Mathematics* (John Wiley & Sons, Inc., New York, 1956), Chap. 3.

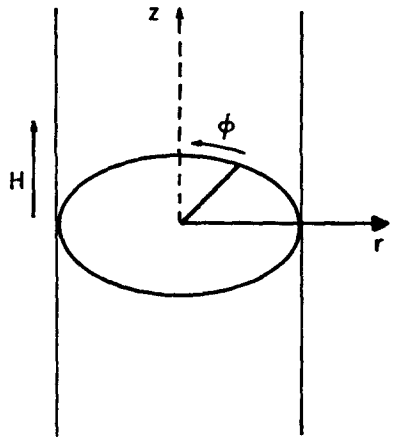


FIG. 7. Specular reflection for circular cylinder in parallel field.

where $\delta_p(x)$ represents a periodic array of δ functions with period 2π.

The above delta functions (which do not appear in the proof of Theorem 3 because the boundary conditions were made homogeneous) can be Fourier-analyzed, and (3.37) then gives

$$k_0 A_0 = (h_1 - h_2)/\pi,$$

$$A_n(n^2 + k_n) = 2[h_1 - h_2(-1)^n]/\pi,$$

which is the same result as in Refs. 5 and 7.

Note: The main characteristic of the specular reflection kernels is the fact that the integral operators defined by them commute with the operator ∇^2, or more generally, with the curl curl operator. Hence, there should exist a common set of eigenfunctions which diagonalize both operators, and the integrodifferential equation (3.33) reduces, in this eigenfunction representation, to an algebraic relation between K and A. The procedure of defining K as in (1.22) is just a particular way of producing such a commuting kernel.

2. The Circular Cylinder

Consider a circular cylinder, of radius a and axis in the z direction (Fig. 7). Let a uniform field H be applied along the axis of the cylinder. Because of symmetry, there is only one component of $\mathbf{A}$, namely A_ϕ, which depends only on r. In order to obtain the specular kernel $K(r, r')$ for this problem, one has to find the normalized eigenfunctions u_n which satisfy

$$\begin{aligned} &\text{curl curl } \mathbf{u}_n = \lambda_n^2 u_n, \qquad \text{div } \mathbf{u}_n = 0, \\ &\hat{e}_r \cdot \mathbf{u}_n = 0 \quad \text{on} \quad r = a, \\ &\hat{e}_r \cdot \text{curl } \mathbf{u}_n = 0 \quad \text{on} \quad r = a, \end{aligned} \tag{3.38}$$

where $\hat{e}_r$, $\hat{e}_\phi$, $\hat{e}_z$ are the usual unit vectors. If we seek solutions $\mathbf{u}_n = (0, u_n(r), 0)$, then Eqs. (3.38) reduce to

$$\nabla^2(\hat{e}_\phi u_n(r)) = -(\lambda_n^2 u_n)\hat{e}_\phi,$$

and

$$(1/r)(\partial/\partial r)(ru_n) = 0 \quad \text{at} \quad r = a. \tag{3.39}$$

Since $\nabla^2(\hat{e}_\phi u_n(r)) = \hat{e}_\phi(\nabla^2 u_n - u_n/r^2)$, we have

$$u_n = \alpha_n J_1(\lambda_n r) \tag{3.40}$$

where the α_n's are normalizing factors, and the λ_n's satisfy the equation

$$\partial/\partial r[rJ_1(\lambda_n r)] = 0 \quad \text{at} \quad r = a.$$

The eigenvalues λ_n are infinite in number and the eigenfunctions u_n are complete in L_2 as may be seen from the Sturm–Liouville theory. The specular reflection now is defined by

$$K(r, r') = \sum_{\lambda_n} \chi(\lambda_n) u_n(r) u_n(r').$$

Now the problem for the vector potential A_ϕ is

$$(\nabla^2 - 1/r^2)A = -2\pi \int_0^a K(r, r')A(r')r'\, dr' \tag{3.41}$$

with

$$(1/a)(\partial/\partial r)(rA) = H \quad \text{at} \quad r = a. \tag{3.42}$$

With the aid of the eigenfunctions $\{u_n\}$, one now diagonalizes Eq. (3.41). Let

$$A(r) \sim \sum A_n u_n(r), \tag{3.43}$$

and substitute into (3.41). Again the differentiation operator brings out a generalized function (a δ function) so that

$$\sum_{\lambda_n} -\lambda_n^2 A_n u_n(r) + 2H\delta(r - a) = \sum k_n A_n u_n(r)$$

where $k_n = -2\pi\chi(\lambda_n)$, and the factor $2H$ comes from the fact that the $\{u_n\}$ expansion of a discontinuous function again converges to the mean value. Expanding $\delta(r - a)$ in terms of $\{u_n\}$, one gets an expression for A_n. The answer is given in terms of the unnormalized eigenfunctions $J_1(\lambda_n r)$. Let

$$A(r) = \sum b_n J_1(\lambda_n r),$$

$$K(r, r') = \sum \tilde{k}_n J_1(\lambda_n r) J_1(\lambda_n r');$$

then

$$b_n = 4H_0/a(\lambda_n^2 + \tilde{k}_n)J_1(\lambda_n a). \tag{3.44}$$

Notice that

$$\tilde{k}_n = \frac{k_n}{\int_0^a |J_1(\lambda_n r)|^2 r\, dr} = \frac{-4\pi\chi(\lambda_n)}{a^2[J_2(\lambda_n a)]^2}.$$

ACKNOWLEDGMENTS

The author wishes to acknowledge many helpful discussions with Professor J. Moser, Dr. P. Marcus, Dr. H. Cohen, Dr. W. Liniger, and, in particular, with Dr. W. Miranker. He is also indebted to the referee for numerous suggestions.

Asymptotic Expansion of Certain Integrals Containing the Bessel Function $J_0(x)$*

PAUL W. SCHMIDT
Physics Department, University of Missouri, Columbia, Missouri
(Received 5 November 1963)

An asymptotic expansion is calculated for the integral

$$\frac{2}{\pi}\int_0^1 \frac{dxf(x)J_0(kx)}{(1-x^2)^{\frac{1}{2}}}$$

where $J_0(x)$ is the zero-order Bessel function of the first kind, and the function $f(x)$ and all its derivatives are continuous for $0 \leq x \leq 1$. To find the asymptotic expansion, the integral is converted to a Fourier integral, to which asymptotic expansion techniques are applied. The method can be generalized to apply to related integrals.

ALTHOUGH methods for calculating asymptotic expansions of Fourier integrals have been developed.[1,2] these techniques cannot be directly applied to a number of integrals which occur in small angle x-ray scattering theory.[3] Certain of these integrals have the form

$$I(k) = \frac{2}{\pi}\int_0^1 \frac{dxf(x)J_0(kx)}{(1-x^2)^{\frac{1}{2}}}$$

where $J_0(x)$ is the zero-order Bessel function of the first kind, and the function $f(x)$ and all its derivatives are continuous in the interval $0 \leq x \leq 1$.

By use of an integral representation for $J_0(x)$, the integral $I(k)$ can be expressed as a double Fourier integral. While the asymptotic expansion of this double integral could be evaluated by the method of Jones and Kline,[2] the expansion can be calculated more conveniently by writing $I(k)$ in the form

$$I(k) = \left(\frac{2}{\pi}\right)^2 \int_0^1 dy\, g(y) \cos ky$$

where

$$g(y) = \int_y^1 \frac{f(x)\,dx}{(1-x^2)^{\frac{1}{2}}(x^2-y^2)^{\frac{1}{2}}}.$$

The asymptotic expansion will be determined by the behavior of $g(y)$ and its derivatives at the end points of the interval of integration and at any interior points at which $g(y)$ or any of its derivatives have discontinuities.[1] Since by hypothesis $f(x)$ and all its derivatives are continuous throughout the interval of integration, $g(y)$ will have no discontinuities at interior points of the interval,[4] and therefore the behavior of $g(y)$ need be investigated only in the neighborhood of $x = 0$ and $x = 1$.

The technique of computation of an asymptotic expansion by representing an integral as a Fourier integral can be applied to other integrals besides $I(k)$. In particular, the method can be used when $J_0(x)$ is replaced by other Bessel functions.

To study the properties of $g(y)$ in the neighborhood of $y = 0$, one can write

$$g(y) = g_1(y) + g_2(y),$$

where

$$g_1(y) = \int_y^\epsilon \frac{f(x)\,dx}{(1-x^2)^{\frac{1}{2}}(x^2-y^2)^{\frac{1}{2}}},$$

$$g_2(y) = \int_\epsilon^1 \frac{f(x)\,dx}{(1-x^2)^{\frac{1}{2}}(x^2-y^2)^{\frac{1}{2}}},$$

with $1 \geq \epsilon > 0$ and $\epsilon > y > 0$. Then at $y = 0$ all derivatives of $g_2(y)$ are continuous. The odd-order derivatives vanish, while the even-order derivatives give no contribution to the asymptotic expansion. Therefore $g_2(y)$ need not be considered in evaluating the contribution of the point $y = 0$ to the asymptotic expansion.[5]

When $f(x)$ is expanded to a Taylor series about $x = 0$, $g_1(y)$ can be put in the form

* Work supported by the National Science Foundation.

[1] A. Erdélyi, *Asymptotic Expansions* (Dover Publications, Inc., New York, 1956), p. 49.

[2] D. S. Jones and M. Kline, J. Math. & Phys. **37**, 1 (1958).

[3] A. Miller and P. W. Schmidt, J. Math. Phys. **3**, 92 (1962); P. W. Schmidt (unpublished research).

[4] The method of calculation can be generalized to allow $f(x)$ to have discontinuities in the interval $0 < x < 1$, without affecting the terms in the expansion contributed by the points $x = 0$ and $x = 1$.

[5] When all derivatives of $f(x)$ are continuous for $0 \leq x \leq 1$, one can let $\epsilon = 1$. Then $g_2(y)$ is identically zero. The number ϵ is introduced to permit generalization to the case where $f(x)$ or some of its derivatives are discontinuous, since when $g(y)$ is broken into $g_1(y)$ and $g_2(y)$, one need assume that the Taylor expansion of $f(x)$ converges only for $0 \leq x \leq \epsilon$, and not for the entire interval $0 \leq x \leq 1$.

$$g_1(y) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} \int_y^c \frac{x^n\,dx}{(1-x^2)^{\frac{1}{2}}(x^2-y^2)^{\frac{1}{2}}}$$

$$= \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \frac{f^{(n)}(0)}{n!\,m!} \frac{\Gamma(m+\frac{1}{2})}{\Gamma(\frac{1}{2})} G_{n+2m}(y),$$

where

$$f^{(n)}(0) = \frac{d^n f}{dx^n}\bigg|_{x=0}, \qquad G_i(y) = \int_y^c \frac{x^i\,dx}{(x^2-y^2)^{\frac{1}{2}}}.$$

The $G_i(y)$ can be evaluated by partial integration. Then, for the same reasons that $g_2(y)$ can be neglected in calculating the contribution of the point $y = 0$, $g_1(y)$ can be written

$$g_1(y) = g_3(y) - \log y \sum_{n=0}^{\infty} A_n y^{2n},$$

where $g_3(y)$ is a function which makes no contribution to the asymptotic expansion, and

$$A_n = \frac{\Gamma(n+\frac{1}{2})}{\Gamma(\frac{1}{2})n!} \sum_{m=0}^{n} \frac{\Gamma(m+\frac{1}{2})}{\Gamma(\frac{1}{2})m!} \frac{f^{(2n-2m)}(0)}{(2n-2m)!}.$$

To investigate the derivatives of $g(y)$ at $y = 1$, let $x = [1-(1-y^2)u^2]^{\frac{1}{2}}$. Then

$$g(y) = \int_0^1 \frac{du f([1-(1-y^2)u^2]^{\frac{1}{2}})}{(1-u^2)^{\frac{1}{2}}[1-(1-y^2)u^2]^{\frac{1}{2}}}$$

$$= \sum_{n=0}^{\infty} (-1)^n f^{(n)}(1) \sum_{k=0}^{n} \frac{(-1)^k}{k!\,(n-k)!} \times \int_0^1 \frac{du}{(1-u^2)^{\frac{1}{2}}} [1-(1-y^2)u^2]^{\frac{1}{2}(k-1)}$$

$$= \frac{\pi}{2} \sum_{n=0}^{\infty} \frac{(-1)^n \Gamma(n+\frac{1}{2})(1-y^2)^n}{\Gamma(\frac{1}{2})(n!)^2} \sum_{l=0}^{n} (-1)^l \times f^{(l)}(1) \sum_{k=0}^{l} \frac{(-1)^k \Gamma(\frac{1}{2}(k+1))}{k!\,(l-k)!\,\Gamma(\frac{1}{2}(k+1)-n)}$$

since

$$\sum_{k=0}^{n} \frac{(-1)^k \Gamma(\frac{1}{2}(k+1))}{k!\,(n-k)!\,\Gamma(\frac{1}{2}(k+1)-l)} = 0$$

for $n > l$, while for $n \leq l$,

$$\sum_{k=0}^{n} \frac{(-1)^k \Gamma(\frac{1}{2}(k+1))}{k!\,(n-k)!\,\Gamma(\frac{1}{2}(k+1)-l)} = \frac{(-1)^l (2l-n)!}{2^{2l-n} n!\,(l-n)!}.$$

Let

$$B_n = \frac{\pi}{2} \sum_{l=0}^{n} \frac{(-1)^l f^{(l)}(1)}{2^{n-l} l!} \times \sum_{\substack{j=0 \\ j \leq \frac{1}{2}n}}^{n-l} \frac{(-1)^j (2n-2j-l)!\,\Gamma(n-j+\frac{1}{2})}{j!\,(n-l-j)!\,\Gamma(\frac{1}{2})(n-j)!\,(n-2j)!}.$$

Then

$$g(y) = \sum_{n=0}^{\infty} B_n (1-y)^n.$$

The contribution of $g_1(y)$ to the asymptotic expansion can be calculated from the theorem of Jones and Kline[6] for asymptotic expansion of Fourier integrals with logarithmic discontinuities, while the contribution of $g(y)$ for the point $y = 1$ can be found by use of Erdélyi's theorem.[1] The asymptotic expansion of $I(k)$ thus can be written

$$I(k) \sim \frac{2}{\pi} \sum_{n=0}^{N-1} \frac{(-1)^n (2n)!\,A_n}{k^{2n+1}} + \left(\frac{2}{\pi}\right)^2 \sum_{n=0}^{M-1} \frac{(-1)^n n!\,B_n}{k^{n+1}} \sin(k + \tfrac{1}{2}n\pi).$$

The asymptotic expansion for $I(k)$ can be simplified for certain forms of $f(x)$. For example,[7] when $f(x)$ can be represented by the expansion

$$f(x) = \sum_{n=0}^{\infty} a_n x(1-x^2)^n$$

for the entire interval of integration, all the $f^{(2n)}(0)$ vanish, and evaluation of the B_n gives

$$I(k) = \frac{2}{\pi} \sum_{n=0}^{M} \frac{\sin(k - \frac{1}{2}n\pi)}{k^{n+1}} \times \sum_{i=0}^{i \leq \frac{1}{2}n} \frac{(-1)^i 2^{n-2i} n!\,\Gamma(n-j+\frac{1}{2})}{j!\,(n-2j)!\,\Gamma(\frac{1}{2})} a_{n-i}.$$

This same expansion of $I(k)$ can be obtained by writing

$$I(k) = \frac{2}{\pi} \sum_{n=0}^{\infty} a_n \int_0^1 dx x(1-x^2)^{n-\frac{1}{2}} J_0(kx)$$

and employing the relation[8]

$$\int_0^1 dx x(1-x^2)^{n-\frac{1}{2}} J_0(kx) = \frac{2^{n-\frac{1}{2}} \Gamma(n+\frac{1}{2}) J_{n+\frac{1}{2}}(k)}{k^{n+\frac{1}{2}}}.$$

The previously calculated asymptotic expansion for $I(k)$ is then obtained by expressing the $J_{n+\frac{1}{2}}(k)$ in terms of sines, cosines, and negative powers of k.

---

[6] Reference 2, p. 26.

[7] The author would like to thank Dr. George H. Weiss for suggesting consideration of this form of $f(x)$.

[8] W. Magnus and F. Oberhettinger, *Formulas and Theorems for the Functions of Mathematical Physics* (Chelsea Publishing Company, New York, 1954), p. 30.