JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 7, NUMBER 8 AUGUST 1966

Laws like Newton's*

RICHARD ARENS
Department of Mathematics, University of California, Los Angeles, California
(Received 25 August 1964)

Instantaneous interaction for n particles is defined in terms of kinematic concepts only. A set of "laws," reminiscent of, but much weaker than, Newton's three laws, is formulated in solely kinematic terms. Invariance under the Euclidean group, the Galilean group, and the special Galilean group is defined, and the most general interactions satisfying the laws, and invariant under these groups, are found. It is shown that they all satisfy Newton's laws. It is shown that the interaction between moving charges cannot be instantaneous and Galilean-invariant.

1. INTRODUCTION

A SET of laws is presented involving only kinematic concepts (such as velocity and acceleration, and not involving concepts such as force and mass) which are much weaker than Newton's three laws of motion. We then show that if the interactions of the particles are required to be invariant under Euclidean motions, then Newton's laws can be deduced from ours.

Having thus brought transformation-invariance into the discussion, we discover the most general interactions allowed by the Euclidean group, the Galilean group, and the special Galilean group.

The word "interaction" is not used here in an informal and general sense, but in a very carefully defined sense (see the definition in Sec. 2). The type of interaction we consider is that in which the acceleration of each particle at a given time t_0 depends only on the positions and velocities of all the particles at *that same time* t_0.

As an application, we prove that the familiar interaction between moving charged particles cannot be of the instantaneous type invariant under the special Galilean group.

* Prepared with assistance of the National Science Foundation, Grant G-2045.

2. THE FOUR LAWS GOVERNING INSTANTANEOUS INTERACTIONS

A *kind of instantaneous*[1] *interaction for n particles in* $\mathbf{R}^3$ is a class K of ordered n-tuplets $(P_1, \cdots, P_n)$, where each P_i is a C^∞ curve in $\mathbf{R}^3$ [that is, having as domain an open set $\Omega(P_i)$ on the real line $\mathbf{R}$, and such that for $t_0 \in \Omega(P_i)$, $P_i(t_0) \in \mathbf{R}^3$, the components P_i^1, P_i^2, P_i^3 being C^∞ functions] subject to the following conditions, traditionally called *laws*:

Law 0: Given n distinct points $p_1, \cdots, p_n$ *in* $\mathbf{R}^3$, *n vectors* $v_1, \cdots, v_n$ (v_i *being a vector at* p_i), *and a real number* t_0, *then there is one and only one* $P = (P_1, \cdots, P_n)$ *in* K *such that* $P_i(t_0) = p_i$ *and* $P_i'(t_0) = v_i$, $i = 1, 2, \cdots, n$.

This law not only reflects the deterministic nature of such interactions but in fact shows that the motions of the individual particles are governed by a system of second-order differential equations—the system depending only on the K kind of interaction.

Before proceeding further, we should identify the objects before us to which the traditional terms

[1] This word will henceforth be omitted for the sake of brevity. In H. van Dam and E. P. Wigner, Phys. Rev. **138**, B1576 (1965), there are shown to exist Lorentz invariant types of interactions. These are not a special case of those treated here.

of kinematics should be applied. Given $P = (P_1, \cdots, P_n) \in K$, we call $P_i(t_0)$ the *position* of the ith particle at time t_0 and $P_i'(t_0)$ [the vector at $P_i(t_0)$ with Cartesian components $P_i^{1\prime}(t_0)$, $P_i^{2\prime}(t_0)$, $P_i^{3\prime}(t_0)$] the *velocity* of the ith particle at time t_0. What name, however, should one apply to the P_i, or to the P itself? The P itself is a set of world lines belonging to or allowed by the interaction. Each P_i is the *world line* of the ith particle. The name "world line" is usually considered to apply to the curve in $\mathbf{R}^4$, which is the *graph* of P_i, but after all, nowadays a function is usually identified with its graph. The set of points $P_i(\tau)$, τ running through all values for which $P_i(\tau)$ is defined, is the *path* of the ith particle and is not to be confused with P_i.

P might aptly be called a *history*, because it assigns to each particle a definite spatio-temporal behavior (viz., its world line). Thus Law 0 says that, for a specified interaction K, one infinitesimally long glimpse at the activity serves to identify the entire history and that, however the particles be placed and instructed to commence their motion, there is a future behavior possible for each, compatible with the kind of interaction embodied by K. Using these terms, we can say that $P_i''(\tau)$ is the acceleration of the ith particle at time τ, in the history P.

Theorem 2.1: For each kind of interaction K there are vector-valued functions $A_1, \cdots, A_n$ of $6n + 1$ variables such that for each P in K the acceleration of the ith particle is given by

$$P_i''(t) = A_i(P(t), P'(t), t), \tag{2.1}$$

where $P(t)$ stands for the list of $3n$ components of the positions, and $P'(t)$ for those of the velocities.

Proof: Given $p_1, \cdots, p_n, v_1, \cdots, v_n$, and τ, we select the (unique) P such that $P_i(\tau) = p_i$ and $P_i'(\tau) = v_i$. We define $A_i(p_1, \cdots, p_n, v_1, \cdots, v_n, \tau)$ as $P_i''(\tau)$. Then, of course $A_i(P(\tau), P'(\tau), \tau) = P_i''(\tau)$, which is all the theorem asserts.

An obvious consequence is that the set of world lines satisfies a second-order system of differential equations.

The remaining laws will be formulated in terms of these *acceleration functions* A_i.

Law 1: Select an index i, and select values for $v_1, \cdots, v_n$ and t. Let $p_1, \cdots, p_n$ vary in any manner such that the distance $||p_i - p_j||$ from p_j to p_i tends to infinity for each j different from i. Then $A_i(p_1, \cdots, p_n, v_1, \cdots, v_n, t)$ tends to 0.

This law implies that if all the particles except one (the ith, in the formulation) are removed to infinity, then the remaining one must move in a straight line with constant speed. This is surely a reasonable way to construe Newton's first law of motion.[2]

The next law has two parts. One is concerned with limits of the sort just considered, and the other is vector-algebraic.

Law 2: Select $v_1, \cdots, v_n$ and t, and two[3] *indices i, j. Let p_i and p_j be distinct fixed points, but let the other p_k tend to infinity. Then $A_i(p_1, \cdots, p_n, v_1, \cdots, v_n, t)$ tends to a limit, to be denoted by $A_{ij}(p_i, p_j, v_i, v_j, t)$. Moreover,*

$$\sum_j A_{ij}(p_i, p_j, v_i, v_j, t) = A_i(p_1, \cdots, p_n, v_1, \cdots, v_n, t) \tag{2.2}$$

for any distinct $p_1, \cdots, p_n$.

Evidently, A_{ij} yields the acceleration of the ith particle when all but the jth particle have been removed. The existence of such "binary interaction accelerations" is surely just as reasonable to assume as the limit in Law 1. In fact, one could reasonably ask also that $A_1(p_1, p_2, p_3, \cdots)$ have a limit as $p_4, p_5, \cdots, p_n$ tend to infinity, etc. However, it is not necessary to assume this, because it follows from the vector-algebraic dissection into binary accelerations given by the second item of Law 2.

Which of Newton's laws makes an assertion of this sort? We must regard it as an "unwritten" part of Newton's second law. After all, in every application of Newton's law, one writes down "mass times acceleration equals" and then considers the other particles, putting down a force contributed by each. The same dissection also applies to the acceleration.

We come now to the third law. Newton's third law implies that there is a linear relation

$$m_1A_1 + \cdots + m_nA_n = 0, \tag{2.3}$$

where these coefficients are constants. According to

---

[2] Newton's first law, in its familiar form, may be construed as saying nothing at all [see R. B. Lindsay and H. Margenau, *Foundations of Physics* (John Wiley & Sons, Inc., New York, 1936), p. 87]. However, Galileo and Newton probably stated it in order to refute the older idea that a free particle would tend to come to rest. To give their law more content, we ask that a particle, far away from all others, moves almost uniformly. We use this assumption later. The following reference, kindly supplied by the referee, provides an instructive criticism of Mach's ideas, as well as further references: C. G. Pendse, Phil. Mag. 29, 477 (1940).

[3] These indices are supposed distinct.

the ideas of Mach (cf., Lindsay and Margeneau, Ref. 2, p. 93) these coefficients are then defined as the mass.

We want to postulate a weaker relation, namely that if all the accelerations except one are known, then the missing one can be found by solving linear equations. This seems to be the essence of the idea of reaction, simplified by a condition of linearity. So we postulate a relation of the form

$$A_1M_1 + \cdots + A_nM_n = 0, \tag{2.4}$$

where the M_i are constant 3×3 matrices written on the right of A_i because A_i is most naturally regarded as a 1×3 matrix or row-matrix.

Before (2.3) or (2.4) can lead to a definition of mass, the relative uniqueness of the M_i must be assured. If all the A_i were 0 ("noninteracting particles"), both Laws 1 and 2, as well as (2.3) would hold, and yet the M_i would remain completely arbitrary. The difficulty cannot be overcome by merely excluding this zero interaction. Looking ahead, we see that a unit of mass shall be chosen. This shall be a single particle, and clearly has to be dynamically comparable to each of the other particles. By this we mean roughly that the interaction be such that each other particle reacts to this particle in a suitable way. More precisely we will require that there be at least one particle α such that given any other particle β, then there exist three sets of initial conditions,

$$\begin{gathered}(p_\alpha, p_\beta, v_\alpha, v_\beta, t), \quad (p'_\alpha, p'_\beta, v'_\alpha, v'_\beta, t'), \\ (p''_\alpha, p''_\beta, v''_\alpha, v''_\beta, t''),\end{gathered} \tag{2.5}$$

for these two, such that if the other particles are removed to infinity, then the three accelerations which these conditions assign to β are linearly independent. This requirement will be recognized as very weak when it is considered that Newton's law of gravitation implies that, among the particles of nonzero mass, any particle can be used to play the role of the above α.

In formulating Law 3, we prefer to avoid the word "particle," but the reader will see that we have elected our "first" particle to be the one capable of influencing the others to the required degree.

Law 3: $A_{21}, A_{31}, \cdots, A_{n1}$ are each capable of three linearly independent values. Each of the components of A_1 is a linear combination with constant coefficients of the $3n-3$ components of $A_2, A_3, \cdots, A_n$.

Theorem 2.4: There is a unique set of matrices $M_2, M_3, \cdots, M_n$ such that

$$A_1 + A_2M_2 + \cdots + A_nM_n = 0. \tag{2.6}$$

Proof: Law 3 ensures that one such relation (2.6) exists. The uniqueness has to be shown. Pick $p_1, p_2, v_1, v_2, v_3, \cdots, v_n$ and t, and let $p_3, p_4, \cdots, p_n$ tend to infinity in such a way that $||p_j - p_k|| \to \infty$ whenever j, k are distinct and greater than 2. It follows from Law 1 that $A_i \to 0$ for $i > 2$. On the other hand, $A_1 \to A_{12}$ and $A_2 \to A_{21}$, so that

$$A_{12} + A_{21}M_2 = 0. \tag{2.7}$$

Suppose there was a relation

$$A_1 + A_2N_2 + \cdots + A_nN_n = 0. \tag{2.8}$$

Then, in the same way, one obtains

$$A_{12} + A_{21}N_2 = 0. \tag{2.9}$$

Therefore, $A_{21}(M_2 - N_2) = 0$. Since Law 3 provides that A_{21} is capable of three linearly independent values, we select p_1, p_2, v_1, v_2, t in three ways [compare (2.5)] to make A_{21} take on three linearly independent values. It follows that $M_2 - N_2 = 0$. In the same way, it follows that $M_3 = N_3, \cdots, M_n = N_n$.

The matrices $M_1, M_2, \cdots, M_n$, where M_1 is the identity and $M_2, \cdots, M_n$ satisfy (2.6), are now called the *masses*. M_i is the *mass of the ith particle*. The vector (or rather row-matrix) A_iM_i shall be called the *force on the ith particle*.

Hence we have

$$\begin{gathered}\text{acceleration of } i\text{th particle} \times \text{mass of} \\ i\text{th particle} = \text{force on } i\text{th particle.}\end{gathered} \tag{2.10}$$

By the limit process used several times, we obtain from (2.6)

$$A_iM_i + A_jM_j = 0, \tag{2.11}$$

or

$$\begin{gathered}\text{force on } i\text{th particle} + \text{force on} \\ j\text{th particle} = 0 \text{ when there are} \\ \text{no other particles around.}\end{gathered} \tag{2.12}$$

Thus Laws 1, 2, and 3 together imply Newton's first and second laws, plus part of the third (see Remark 4.3 below).

However, the masses are matrices. They would be scalars if we deliberately required in Law 3 that the *accelerations* were linearly dependent. This would make a stronger law, but still acceptable to Newton, so to speak, because he does require (2.6) with scalars (*positive* scalars indeed).

We prefer to arrive at scalar matrices through the symmetry considerations of the next section.

3. INVARIANCE OF INTERACTIONS

Let T be a mapping of $\mathbf{R}^3$ onto itself. Let K be a kind of interaction for n particles in the technical sense defined in the last section. Let $P = (P_1, \cdots, P_n)$ be a member of K. For each curve P_i here, one can form the transformed curve $T \circ P_i$ where $(T \circ P_i)(t) = T(P_i(t))$.

The class of curves $(T \circ P_1, \cdots, T \circ P_n)$, one may reasonably denote by $T \circ P$. Here $T \circ P$ may or may not be again a member of K.

Definition 3.1: K is called *T-invariant* if $T \circ P$ belongs to K for each P in K.

Definition 3.2: If K is T-invariant for every translation T or every orthogonal transformation T, then K is called *translation-invariant* or *orthogonal-invariant*, respectively. If K is both translation-invariant and orthogonal-invariant, then K is called a *Euclidean interaction.*

An orthogonal transformation with positive determinant shall be called a *rotation.* If we replace "orthogonal" by "rotation" in Definition 3.2, we obtain the definitions of *rotation-invariant* and *special Euclidean* interaction. Euclidean implies special Euclidean, but not conversely.

Proposition 3.3: K is T-invariant if and only if for the $A_1, \cdots, A_n$,

$$A_i(T(p_1), \cdots, T(p_n), T(v_1), \cdots, T(v_n), t) = T[A_i(p_1, \cdots, p_n, v_1, \cdots, v_n, t)],$$

for all $p_1, \cdots, p_n, v_1, \cdots, v_n, t$.

By Tv_j we mean the vector at Tp_j into which the vector v_j at p_j is carried by T. If $x^iT = \bar{x}^i$ and the components of v are $v^1, \cdots, v^n$, then the components of $\bar{v} = T(v)$ are $\bar{v}^i = v^k(\partial \bar{x}^i / \partial x^k)$, where summation on repeated indices is understood. With this explanation, the reader can easily convince himself of Proposition 3.3.

Suppose T is a linear homogeneous transformation of $\mathbf{R}^3$ into itself. Then corresponding to the two-time choice of the Cartesian basis, there is a matrix $[T]$ assigned to T. Let $[T]$ have T_{ij} in the ith row and jth column. Then $\bar{x}^i = x^iT = T_{ik}x^k$, where x^j is the jth Cartesian coordinate. Moreover, if $[p]$ is the column matrix whose entries are the Cartesian coordinates of the point p, then

$$[T(p)] = [T][p] \text{ (matrix product on the right)}, \quad (3.1)$$

and if for a vector v we let $[v]$ be the column matrix formed from its Cartesian components, then

$$[T(v)] = [T][v]. \quad (3.2)$$

In this notation, the invariance condition of Proposition 3.3 can be written

$$[A_i \circ T] = [T][A_i]. \quad (3.3)$$

Recalling that in writing Eq. (2.6) the A_i were construed as standing for row-matrices, (2.6) takes the equivalent form in the new notation

$$[A_1] + M_2^t[A_2] + \cdots + M_n^t[A_n] = 0, \quad (3.4)$$

where the "t" indicates transposition. From this follows

$$[A_1 \circ T] + M_2^t[A_2 \circ T] + \cdots = 0,$$

and if Eq. (3.3) holds we obtain

$$[T][A_1] + M_2^t[T][A_2] + \cdots = 0,$$

as well as

$$[A_1] + [T]^{-1}M_2^t[T][A_2] + \cdots = 0.$$

Considering the uniqueness of the M, we obtain the following.

Proposition 3.4: Let K be invariant under the linear homogeneous transformation T of $\mathbf{R}^3$. Then $M_k^t[T] = [T]M_k^t$ for $k = 1, 2, \cdots, n$.

4. SCALAR MASSES

Theorem 4.1: Let K be a rotation-invariant interaction. Then the masses are scalar matrices.

This follows from Proposition 3.4, because only a scalar matrix commutes with all rotations.

This result shows that some of the bizarre situations allowed by our rather weak laws (weak compared to Newton's) are ruled out by rotation-invariance. In fact, there are still these two peculiar possibilities, even with rotational symmetry:

Remark 4.2: The masses may be positive, negative, or zero scalars.

Remark 4.3: When there are only two particles, the forces are not necessarily directed along the line connecting their positions.

In the next section we examine Euclidean interactions and find conditions which ensure that Newton's third law hold in its entirety.

5. EUCLIDEAN INTERACTIONS

Consideration of Law 2 makes it clear that interactions are just sums of *binary* interactions; so,

if the binary interactions are all described, then all general interactions will also be described. The same is true for Euclidean interactions.

If $n = 2$, then A_{ij} is merely A_i, except for a tiny notational distinction: whereas

$$A_1(p_1, p_2, v_1, v_2, t) = A_{12}(p_1, p_2, v_1, v_2, t),$$

for A_{21} we have

$$A_2(p_1, p_2, v_1, v_2, t) = A_{21}(p_2, p_1, v_2, v_1, t).$$

Let us denote M_1A_1 by F_1 and M_2A_2 by F_2. Then $F_1 + F_2 = 0$ as far as Cartesian components are concerned.

Let us consider a Euclidean interaction, with $n = 2$.

Theorem 5.1: In a Euclidean binary interaction,

$$\begin{aligned} F_1(p_1, p_2, v_1, v_2, t) &= -F_2(p_1, p_2, v_1, v_2, t) \\ &= (p_1 - p_2)f + v_1 g - v_2 h. \end{aligned} \tag{5.1}$$

This is meant as an equation of Cartesian components. f, g, and h are scalars depending only on t and the six Euclidean invariants

$$\left.\begin{array}{lll} (p_1 - p_2)\cdot(p_1 - p_2), & (p_1 - p_2)\cdot v_1, & (p_1 - p_2)\cdot v_2 \\ & v_1\cdot v_1, & v_1\cdot v_2 \\ & & v_2\cdot v_2 \end{array}\right\} \tag{5.2}$$

Proof: We remark at once that f, g, and h are not uniquely determined, because $p_1 - p_2$, v_1, and v_2 may be linearly dependent. We show that given the values of the array (5.2), we can calculate f, g, and h such that Eq. (5.1) holds.

Select a Euclidean transformation which sends p_1 to the origin, p_2 to a point p_2' on the z axis, and v_1, if it is not collinear with p_1 and p_2, into a vector v_1' in the xz plane. Call this the general v_1 case. The special v_1 case is that in which v_1' lies on the z axis—in this case let v_2' lie in the xz plane. In the general v_1 case, v_2' may lie also in the xz plane—this is the special v_2 case, the other being the general v_2 case.

The case which occurs can be ascertained from the values in (5.2).

In the general v_1, v_2 case, F_1' has a unique representation

$$F_1' = (p_1' - p_2')f + v_1' g - v_2' h. \tag{5.3}$$

In the special v_1, general v_2 case, a reflection in the xz plane preserves p_1', p_2', v_1', v_2' and therefore also preserves F_1' which accordingly lies in the xz plane. Hence (5.3) is again possible, and we choose $g = 0$ to make it unique.

In the special v_1, special v_2 case, any rotation about the z axis preserves p_1', $\cdots$, v_2' and thus also F_1' which accordingly must be a multiple of $p_2' - p_1'$. Hence (5.3) is possible, and unique if $g = h = 0$.

In the general v_1, special v_2 case, we appeal again to the reflection, and let $h = 0$.

Thus (5.3) holds for all values of p_1, p_2, v_1, v_2, where f, g, h do depend only on (5.2). But if the F_1 is Euclidean, then (5.3) implies (5.1). Thus Theorem 5.1 holds.

Call an interaction *Newtonian* if the acceleration functions depend only on p_1, $\cdots$, p_n.

Theorem 5.2: In a Euclidean, Newtonian interaction, the force exerted by one particle on another is along the line through their positions.

Proof: Masses being scalar, it follows that the binary forces (see Theorem 5.1) depend only on p_1, p_2. Thus we must have

$$\begin{aligned} F_1(p_1, p_2, v_1, v_2, t) &= -F_2(p_1, p_2, v_1, v_2, t) \\ &= (p_1 - p_2)f, \end{aligned}$$

where f depends only on $(p_1 - p_2)\cdot(p_1 - p_2)$, and t.

Theorem 5.3: In a Euclidean, Newtonian interaction, all of Newton's laws hold in the sense that they are usually understood, except that masses may have any sign (or vanish). Moreover, such systems are conservative.

Proof: Theorem 5.2 supplies the part of Newton's third law which we had not yet obtained. It is well known that when the force between each pair of particles depends only on the distance, as is the case when

$$F_1 = (p_1 - p_2)f((p_2 - p_1)\cdot(p_2 - p_1)), \tag{5.4}$$

then

$$V((p_2 - p_1)\cdot(p_2 - p_1)) \tag{5.5}$$

is a potential energy expression for the pair (1, 2) provided that

$$2V' = -f. \tag{5.6}$$

Since f is a function of one variable, it would have to be discontinuous if a primitive $-2V$ for it did not exist. It is reasonable to limit the discussion to continuous f, but it is more logical to extend the meaning of "conservative."

Assume, however, that Eq. (5.6) holds for the pair (1, 2). We denote that function V by V_{12}.

Then $V_{12} = V_{21}$. The total potential for the system is

$$V(p_1, \cdots, p_n) = \sum_{i<j} V_{ij}((p_i - p_j)\cdot(p_i - p_j)). \quad (5.7)$$

It is interesting that potential energy is due to *pairs*, and is to be added up over pairs, whereas kinetic energy is made up of contributions from the various individual particles. Moreover,

the sum of the kinetic energy and the potential energy of a Euclidean, Newtonian system is a Euclidean invariant. (5.8)

It is worth noting that, in the Euclidean case, Law 3 may be weakened to the extent that the first sentence may be replaced by

none of $A_{21}, A_{22}, \cdots, A_{2n}$ *vanishes identically,* (5.9)

which is, informally,

no particle is completely indifferent to the first particle. (5.9′)

6. GALILEAN INTERACTIONS

It is well known[4] that Euclidean, Newtonian interactions are invariant under transformations more general than the Euclidean transformations, provided that transformations of *space–time* are allowed.

We will define as a Galilean transformation any 1:1 transformation of space–time onto itself which preserves each Euclidean, Newtonian interaction.[5] By "preserves" we mean "is invariant under," but the meaning has yet to be made precise, since the earlier definition (limited to transformations of space) does not apply without modification.

Definition 6.1: (Definition of invariance). Let K be an interaction, and let G map $\mathbf{R}^4$ into itself. Let P be any member of K. Then $P = (P_1, \cdots, P_n)$, where $P_1, \cdots, P_n$ are curves in $\mathbf{R}^3$. Each P_i determines an *arc* in $\mathbf{R}^4$, the arc consisting of the points $(P_i^1(\tau), P_i^2(\tau), P_i^3(\tau), \tau)$ where τ runs over all real values for which $P_i(\tau)$ is defined. This arc may be called the *graph* of P_i. Let it be denoted by $\text{graph}(P_i)$. Now $\text{graph}(P_i)$ is a subset of $\mathbf{R}^4$ and so $G(\text{graph}(P_i))$ is a subset of $\mathbf{R}^4$. We say that K *is G-invariant* if there exists another member Q of K, $Q = (Q_1, \cdots, Q_n)$, such that

$$\text{graph}(Q_i) = G(\text{graph}(P_i)) \text{ for } i = 1, 2, \cdots, n.$$

As mentioned in Sec. 2, graph (P_i) is usually called the world line of the ith particle in the history P and may be identified with P_i. Thus Definition 6.1 can be given a form which makes it very obvious that this should be called G-invariance:

$$(P_1, \cdots, P_n) \text{ in } K \text{ implies } (G(P_1), \cdots, G(P_n)) \text{ in } K. \quad (6.1)$$

A Galilean transformation G has to be linear because it must preserve the zero interaction. Here the world lines are the straight lines, and if the image of each straight line is a straight line, then G is linear.

The most typical Galilean transformation is obtained by choosing three numbers u, v, w and defining G by

$$\bar{x} = x \circ G = x - ut, \quad \bar{y} = y \circ G = y - vt,$$
$$\bar{z} = z \circ G = z - wt, \quad \bar{t} = t \circ G = t. \quad (6.2)$$

Then there are those which come from Euclidean transformations T in $\mathbf{R}^3$, by means of the formula

$$\bar{T}(a, b, c, d) = (T(a, b, c), d). \quad (6.3)$$

Finally, there are the *time shifts*:

$$(a, b, c, d) \to (a, b, c, d + \tau). \quad (6.4)$$

Each of these transformations (6.2), (6.3), (6.4) is Galilean. In fact, they generate the Galilean group. This can be shown by considering what linear transformations leave invariant the Newtonian "inverse-square law" of interaction.

Thus, the Euclidean, Newtonian interactions define the Galilean group. But does the Galilean group, conversely, define the Euclidean, Newtonian interactions? More explicitly, if an interaction is Galilean (invariant), must it be Euclidean, Newtonian?

The following characterization of Galilean interactions shows that some are not Newtonian.

Theorem 6.3: Let K be a Galilean interaction. Then (compare Theorem 5.1)

$$F_1(p_1, p_2, v_1, v_2, t) = -F_2(p_1, p_2, v_1, v_2)$$
$$= (p_1 - p_2)f + (v_1 - v_2)g, \quad (6.5)$$

where f and g depend only on the Galilean invariants

[4] See M. Born, *Einstein's Theory of Relativity*, prepared with collaboration of G. Leibfried and W. Biem (Dover Publications, Inc., New York, 1962), rev. ed., p. 74.

[5] Born defines the Galilean group of transformations explicitly as a certain group of linear transformations, so that his statement on p. 74 asserts that "his" group is contained in "ours." On the other hand, the opposite inclusion also holds. Perhaps this is regarded as obvious by Born, because if it were not true, he certainly would have mentioned the larger group. Speaking of the Galilean group, we point out that Laws 0–3 are themselves invariant under this group and not invariant under some other (say Lorentz) transformations. However, this cannot be used as a definition of the Galilean group, since Laws 0–3 are invariant under some transformations such as, for example, dilations which are not Euclidean and hence not Galilean.

$$(p_1 - p_2)\cdot(p_1 - p_2),\quad (p_1 - p_2)\cdot(v_1 - v_2),\quad (v_1 - v_2)\cdot(v_1 - v_2). \tag{6.6}$$

Conversely, if f and g are functions of three variables such that

$$\lim_{|x|\to\infty} f(x^2, xy, z)x = 0 \tag{6.7}$$

and

$$\lim_{|x|\to\infty} g(x^2, xy, z) = 0, \tag{6.8}$$

then

$$(p_1 - p_2)f(6.6) + (v_1 - v_2)g(6.6) \tag{6.9}$$

defines a Galilean interaction.

Here the combination of (6.8) and (6.9) is just the relevant form of Law 1.

Proof: Let $P = (P_1, P_2)$ be a member of K. Let G be the transformation (6.2), where u, v, w are the components of $P_2'(0)$. We must calculate $G(P_i)$.

$P_i = \{(P_i(\tau), t) : \tau \text{ real}\}$. So $G(P_i) = \{G(P_i(\tau), \tau)$: τ real$\}$. Now

$$G(P_i(\tau), \tau) = \Big(x\Big(G(P_i(\tau), \tau)\Big),\quad y\Big(G(P_i(\tau), \tau)\Big),$$
$$z\Big(G(P_i(\tau), \tau)\Big),\quad t\Big(G(P_i(\tau), \tau)\Big)\Big).$$

According to (6.2),

$$x(G(P_i(\tau), \tau)) = x(P_i(\tau), \tau) - ut(P_i(\tau), \tau),$$

and this is by definition $= P_i^1(\tau) - u\tau$, where the superscript indicates the component, and similar formulas hold for y and z. But, by (6.2),

$$t(G(P_i(\tau), \tau)) = \tau.$$

Denote $G(P_i)$ by Q_i. Then

$$Q_i = \{(P_i^1(\tau) - u\tau, \cdots, P_i^3(\tau) - w\tau, \tau) : \tau \text{ real}\},$$

so that $Q_i(\tau) = (P_i^1(\tau) - u\tau, \cdots, P_i^3(\tau) - w\tau)$. As a result

$$Q_1(0) = P_1(0),\quad Q_i'(0) = P_i'(0) - (u, v, w) \text{ and } Q''(0) = P_1''(0). \tag{6.10}$$

Now, a Galilean interaction is certainly Euclidean [see Eq. (6.3)]; hence, Theorem 5.1 applies and so

$$m_1Q_1''(0) = [Q_1(0) - Q_2(0)]f + Q_1'(0)g - Q_2'(0)h, \tag{6.11}$$

where f, g, h depend only on K and on the values of (5.2) for Q_1 and Q_2 at $t = 0$. Since $(u, v, w) = P_2(0)$, the array (5.2) reduces to[6]

$$\left.\begin{array}{l}(P_1(0) - P_2(0))\cdot(P_1(0) - P_2(0)),\\ (P_1(0) - P_2(0))\cdot(P_1'(0) - P_2'(0)), 0\\ (P_1'(0) - P_2'(0))\cdot(P_1'(0) - P_2'(0)), 0\\ 0\end{array}\right\}. \tag{6.12}$$

Inserting (6.10) into (6.11),

$$m_1P_1''(0) = [P_1(0) - P_2(0)]f + [P_1'(0) - P_2'(0)]g,$$

where f and g depend only on (6.12).

This establishes (6.5) for $t = 0$. Consideration of (6.4) establishes it for all t.

It is of interest to note which of these Galilean interactions are invariant under the larger group obtained by adjoining the dilations(" generalized Galilean group"). The answer is, those for which f and g are homogeneous functions of the zeroth degree.

Finally, we will discover the most general interaction which is invariant under all *special* Galilean transformations (a transformation is called special if its Jacobian is positive). One might call such interactions "special Galilean" interactions, but should keep in mind that they contain the Galilean interactions as particular cases. The result depends on the following, whose proof is left to the reader.

Lemma 6.4: Suppose that whenever U, V are points of $\mathbf{R}^3$, then $B(U, V)$ is a point of $\mathbf{R}^3$ such that whenever S is a special orthogonal transformation (i.e., a rotation) then $B(SU, SV) = S(B(U, V))$. Then

$$B(U, V) = fU + gV + h(U \times V), \tag{6.13}$$

where f, g, h depend only on

$$U\cdot U,\quad U\cdot V,\quad V\cdot V. \tag{6.14}$$

Theorem 6.5: Let K be an interaction invariant under all special Galilean transformations. Then

$$F_1(p_1, p_2, v_1, v_2, t) = -F_2(p_1, p_2, v_1, v_2, t)$$
$$= (p_1 - p_2)f + (v_1 - v_2)g + [(p_1 - p_2) \times (v_1 - v_2)]h,$$

where[7] f, g, and h depend only on the Galilean invariants (6.6).

Proof: Let $P = (P_1, P_2)$ be a member of K. Let G be the transformation for which [compare (6.2)]

[6] The reader should move the first line of (6.12) to the front of the second line, in order to see the connection with (5.2).

[7] Without this additional assertion, Theorem 6.5 would be trivial.

$\bar{x} = x - ut - a$, $\bar{y} = y - vt - b$, $\bar{z} = z - wt - c$, $\bar{t} = t$, where

$$(a, b, c) = P_2(0), \quad (u, v, w) = P_2'(0).$$

Proceeding as in the proof of Theorem 6.3, we find that [instead of (6.10)]

$$\begin{aligned} Q_1(0) &= P_1(0) - P_2(0), \\ Q_1'(0) &= P_1'(0) - P_2'(0), \quad Q_1''(0) = P_1''(0), \end{aligned} \tag{6.15}$$

and

$$Q_2(0) = 0, \quad Q_2'(0) = 0, \quad Q_2''(0) = P_2''(0). \tag{6.16}$$

Now Q belongs to K, by the special Galilean invariance of K. Hence,

$$m_1 Q_1''(0) = F_1(Q_1(0), Q_2(0), Q_1'(0), Q_2'(0), 0)$$

or

$$m_1 P_1''(0) = F_1(P_1(0) - P_2(0), 0, P_1'(0) - P_2'(0), 0, 0).$$

Now denote $F_1(U, 0, V, 0, 0)$ by $B(U, V)$. This B satisfies the hypotheses of Lemma 6.4. Therefore

$$\begin{aligned} m_1 P_1''(0) = {} & (P_1(0) - P_2(0))f + (P_1'(0) - P_2'(0))g \\ & + [(P_1(0) - P_2(0)) \times (P_1'(0) - P_2'(0))]h, \end{aligned}$$

and on the other hand,

$$m_1 P_1''(0) = F_1(P_1(0), P_2(0), P_1'(0), P_2'(0), 0),$$

so clearly, for $t = 0$ at least,

$$\begin{aligned} F_1(p_1, p_2, v_1, v_2, t) = {} & (p_1 - p_2)f + (v_1 - v_2)g \\ & + [(p_1 - p_2) \times (v_1 - v_2)]h, \end{aligned} \tag{6.17}$$

where f, g, h depend only on (6.6). As before, this extends at once to all values of t.

The proof here did not use Theorem 6.3, so it is worth noting that Theorem 6.3 is an immediate consequence of Lemma 6.4.

7. MOVING ELECTRIC CHARGES

The fact that the motion of two electric charges cannot be explained by (instantaneous) interactions is a consequence of the following.

Theorem 7.1: Suppose K is an interaction between two particles such that, if one is at rest, then the acceleration of the other one is directed toward the first. Suppose that this interaction is invariant under all special Galilean transformations. Then the acceleration of each particle is always along the line of centers.

Proof: By Theorem 6.5, we must have

$$\begin{aligned} F_1 = -F_2 = {} & pf(r^2, p\cdot v, v\cdot v) + vg(r^2, p\cdot v, v\cdot v) \\ & + (p \times v)h(r^2, p\cdot v, v\cdot v), \end{aligned}$$

where $p = p_1 - p_2$, $v = v_1 - v_2$, and $r^2 = p\cdot p$.

Now suppose $p_2 = 0$ and $v_2 = 0$. Then $p = p_1$, $v = v_1$, and

$$\begin{aligned} F_1(p, 0, v, 0, t) = {} & pt(r^2, p\cdot v, v\cdot v) + vg(r^2, p\cdot v, v\cdot v). \\ & + (p \times v)h(r^2, p\cdot v, v\cdot v). \end{aligned}$$

We are told that $F_1(p, 0, v, 0, t)$ is a scalar multiple of p. The values of r^2, $p\cdot v$, $v\cdot v$ can still be varied independently. Hence g and h must be 0. This proves Theorem 7.1.

Doppler Measurement of Space–Time Curvature

BRUNO BERTOTTI*
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England
(Received 3 November 1965)

In special relativity, the Doppler shift between two freely moving identical oscillators is constant if their world lines are coplanar. We show that in general relativity, instead, the rate of change of their Doppler shift is proportional to a component of the space–time curvature, averaged along the light ray. A possible application to the detection of gravitational waves is discussed.

1

A NUMBER of theoretical methods have been proposed to measure the curvature tensor of space-time[1]; they are all based on light signals sent back and forth between freely falling observers. It is interesting to note the existence of another method, in which a bouncing device ("mirror") is not required. We would like to show that it is possible to measure a mean curvature (in the sense to be specified) by measuring the *rate of change* of the Doppler shift between two freely falling observers. Compared with other methods, this one has the advantage of giving a simple and clear-cut result even when the distance between the source and the reciever is not small with respect to the radius of curvature. Although an application of this method is briefly mentioned at the end, this discussion is of theoretical nature and has no direct experimental relevance.

2

Figure 1 illustrates the geometrical setup. Notice that the affine parameter l along each null geodesic is determined to within a linear transformation, whose coefficients may depend on the proper time of the source s. We choose $l = 0$ at the source; the unit interval for l is such that the invariant $\nu = \mathbf{p}(0)\cdot\mathbf{v}(0)$ is the emitted frequency. l and s can be considered as coordinates on the two-dimensional surface spanned by the null geodesics. The observer is characterized by (say) $l_0(s)$. For brevity we do not indicate explicitly the s dependence. The indices s and l indicate partial absolute derivatives.

The unit vector fields $\mathbf{v}(l)$ and $\mathbf{v}'(l)$ are defined by

$$\mathbf{v}_l(l) = \mathbf{v}'_l(l) = 0. \tag{1}$$

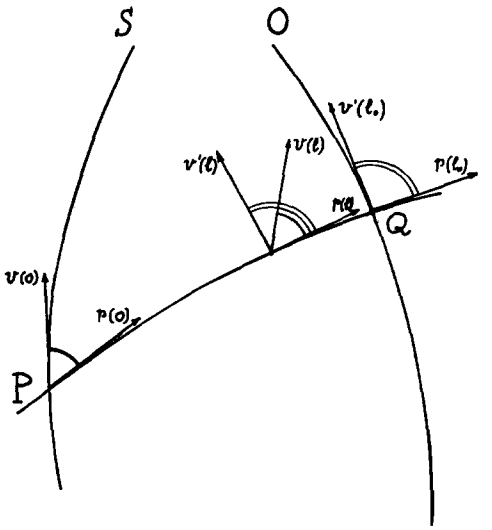


FIG. 1. The geometrical setup. $p^\mu(l, s) \equiv \partial x^\mu/\partial l$: null vector transferred parallel along the null geodesic between P and Q; $\mathbf{v}(l, s)$, $\mathbf{v}'(l, s)$: unit vectors obtained by transferring parallel along the null geodesic the velocities of the source and the observer, respectively.

$\mathbf{v}(0)$ and $\mathbf{v}'(l_0)$ are, respectively, the velocities of the source and the observer.

The frequency shift is given by[2,3]

$$1 + z = \frac{\nu}{\nu_0} = \frac{ds'}{ds} = \frac{\mathbf{p}(0)\cdot\mathbf{v}(0)}{\mathbf{p}(l_0)\cdot\mathbf{v}'(l_0)}. \tag{2}$$

s' indicates the observer's proper time. The interpretation previously given to the scalar $\mathbf{p}\cdot\mathbf{v}$ is consistent with Eq. (2), since z does not depend on the normalization of $\mathbf{p}$.

We now demand the following conditions:

(a) source and observer fall freely:

$$\mathbf{v}_s(0) = \mathbf{v}'_s(l_0) = 0; \tag{3}$$

* Permanent address: Laboratorio Gas Ionizzati, Frascati, Rome.

[1] E. Wigner, Rev. Mod. Phys. **29**, 255 (1957); *ibid.* **120**, 643, (1960); J. L. Synge, *Relativity: The General Theory* (North-Holland Publishing Company, Amsterdam, 1960); B. Bertotti, *Rend. Scuola Intern. Fis. "Enrico Fermi," XX Corso* (Academic Press Inc., New York, 1962), pp. 195–199.

[2] E. Schrödinger, *Expanding Universes* (Cambridge University Press, New York, 1956).

[3] This formula can also be written in a way in which its analogy with its special relativistic counterpart is apparent:

$$\nu_0 = \nu\{1 - \nu^{-1}\mathbf{p}(0)\cdot[\mathbf{v}(0) - \mathbf{v}'(0)]\}; \tag{2'}$$

$\mathbf{v}(0) - \mathbf{v}'(0)$ is the relative velocity of the source with respect to the observer.

(b) the same spectral line is observed; that is to say, the emitted frequency is constant:

$$\nu_s = \mathbf{p}_s(0)\cdot\mathbf{v}(0) = 0; \tag{4}$$

(c) the light arrives at the observer always from the same direction with respect to a local inertial frame of reference[4]:

$$\mathbf{p}_s(l_0) = [\mathbf{p}_s(l_0)\cdot\mathbf{v}'(l_0)]\mathbf{v}'(l_0) + \lambda\{\mathbf{p}(l_0) - [\mathbf{p}(l_0)\cdot\mathbf{v}'(l_0)]\mathbf{v}'(l_0)\}. \tag{5}$$

λ is determined by the condition that $\mathbf{p}$ is a null vector, which leads to

$$\mathbf{p}_s(l_0) = \frac{\mathbf{p}_s(l_0)\cdot\mathbf{v}'(l_0)}{\mathbf{p}(l_0)\cdot\mathbf{v}'(l_0)}\,\mathbf{p}(l_0). \tag{5'}$$

3

We now proceed to compute, using (2), (3), and (4), the derivative of z with respect to s'

$$z_{s'} = \frac{z_s}{1+z} = -\frac{\mathbf{p}_s(l_0)\cdot\mathbf{v}'(l_0)}{\mathbf{p}(l_0)\cdot\mathbf{v}'(l_0)}. \tag{6}$$

Envisage now the scalar

$$\Omega(l) \equiv \mathbf{p}_s(l)\cdot\mathbf{v}(l), \tag{7}$$

which vanishes at $l = 0$ because of (4). It fulfils the differential equation[5]:

$$\Omega_l(l) = R_{\mu\nu\rho\sigma}v^\mu p^\nu q^\rho p^\sigma, \tag{8}$$

where

$$q^\mu(l) = \partial x^\mu/\partial s; \tag{9}$$

hence

$$\Omega(l_0) = \int_0^{l_0} dl\, R_{\mu\nu\rho\sigma}v^\mu p^\nu q^\rho p^\sigma. \tag{10}$$

Using (5') we can obtain the final formula

$$z_{s'} = -\nu^{-1}\int_0^{l_0} dl\, R_{\mu\nu\rho\sigma}v^\mu p^\nu q^\rho p^\sigma. \tag{11}$$

As one expected, the result does not depend on the particular normalization chosen for $\mathbf{p}$.

To gain further insight on this formula, consider the vector

$$\mathbf{u}(l) \equiv \mathbf{q}(l) - \mathbf{v}(l). \tag{12}$$

It fulfills the relationships:

$$\mathbf{u}_l(l) = \mathbf{q}_l(l) = \mathbf{p}_s(l), \tag{13}$$

$$u^\mu_{ll}(l) = q^\mu_{ll}(l) = p^\mu_{sl}(l) = R^\mu_{\mu\nu\rho\sigma}p^\nu q^\rho p^\sigma, \tag{14}$$

$$\mathbf{u}(0) = 0, \quad \mathbf{u}_l(l_0) = \frac{1}{\nu}\int_0^{l_0} dl\, R_{\mu\nu\rho\sigma}v^\mu p^\nu q^\rho p^\sigma \mathbf{p}(l_0).$$

In a flat space $\mathbf{u}(l) = 0$; thus it is convenient to write Eq. (11) in the form

$$z_{s'} = -\nu^{-1}\left[\int_0^{l_0} dl\, K(l) + \int_0^{l_0} dl\, R_{\mu\nu\rho\sigma}v^\mu p^\nu u^\rho p^\sigma\right], \tag{15}$$

where

$$K(l) \equiv R_{\mu\nu\rho\sigma}v^\mu p^\nu v^\rho p^\sigma \tag{16}$$

is proportional to the Gaussian curvature of the geodesic two-dimensional surface determined by the vectors $\mathbf{v}$ and $\mathbf{p}$. The second term in the square bracket is of the second order in the curvature and can be neglected in a linearized calculation.

4

According to Eq. (15), a change in the redshift of a star or a galaxy could be ascribed to a gravitational wave crossing the light ray. It is interesting to see under which conditions this method offers a more sensitive test than the one based on microseisms.[6] A random distribution of linear gravitational waves at the frequency ω gives rise, according to Eq. (15) (notice that νl is a length), to a mean square variation in z of the order of

$$\langle(\Delta z)^2\rangle \approx \frac{1}{\omega^2}\langle z^2_{s'}\rangle \approx \frac{Lc^3}{\omega^3}\langle R^2\rangle, \tag{17}$$

where R is a typical component of the Riemann tensor and L is the distance of the source. A similar equation holds when $\langle(\Delta z)^2\rangle$ and $\langle R^2\rangle$ are replaced with their spectral densities $\langle(\Delta z)^2\rangle_\omega$ and $\langle R^2\rangle_\omega$. When $\omega^{-1} \approx 3000$ sec and $\langle R^2\rangle_\omega \approx 10^{-75}$ cm^{-4} sec, $\langle(\Delta z)^2\rangle_\omega \approx 10^{-15}L$, where L is expressed in light years. Assuming that the spectrum had a bandwidth of order ω, the total fractional change in the light frequency is of the order of

$$(\omega L 10^{-15})^{\frac{1}{2}} \approx \tfrac{1}{3}10^{-9}L^{\frac{1}{2}}.$$

Since the size of the source must not be much greater than $c\omega^{-1}$, in order to prevent finite size effects, this method does not look very promising.

[4] On the concept of local inertial frame of reference, see F. A. E. Pirani, Acta Phys. Polon. 15, 389 (1956); Helv. Phys. Acta, Suppl. IV, 199 (1956); F. A. E. Pirani and A. Schild, Bull. Acad. Polon. Sci. 9, 543 (1961); B. Bertotti, D. Brill, and R. Krotkov, "Experiments on Gravitation," in *Progress Report on General Relativity*, L. Witten, Ed. (John Wiley & Sons, Inc., New York, 1962).

[5] J. L. Synge and A. Schild, *Tensor Calculus* (The University of Toronto Press, Toronto, 1949), Eq. (3.107).

[6] R. L. Forward, D. Zipoy, J. Weber, S. Smith, and H. Benioff, Nature 189, 473 (1961).

A Two-Meson Solution of the Charged Scalar Static Model*

J. B. BRONZAN
Department of Physics and Laboratory for Nuclear Science, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received 3 January 1966)

Presented in this paper is a solution of the charged scalar static model, in which the scattering amplitude is crossing symmetric and has two- and three-particle intermediate states (two-meson approximation). The production and six-point amplitudes are also obtained and are shown to lead to a two- and three-particle unitary scattering matrix.

I. INTRODUCTION

ALTHOUGH static models are much simpler than fully relativistic theories, the approximate solutions obtained for them have not gone very far beyond what is possible for relativistic theories. On the one hand, exact one-meson solutions are available for the neutral,[1] charged,[1] and symmetric[2] scalar theories, and for the neutral pseudoscalar theory.[2] These solutions for the scattering amplitude are crossing symmetric and satisfy elastic unitarity. On the other hand, a solution has been given for $V\theta$ scattering in the Lee model which satisfies two- and three-particle unitarity.[3] However, apart from these solutions and the early strong[4] and intermediate[5] coupling results, the same compromises must be made in evaluating static theories as are made in evaluating relativistic theories. This somewhat limits the usefulness of static theories as models.

In this paper, we give a two-meson solution for the charged scalar static model. As such, our solution is a next step in the refinement of the known approximations for the charged scalar theory. In addition, as far as we know, it is the first example of a two-meson solution in dispersion theory. As we see in Sec. VIII, there is some latitude in specifying what constitutes a two-meson solution, since various dynamical approximations are possible in the calculation of the production and six-point amplitudes. But everyone would agree that a two-meson solution must provide scattering, production, and six-point amplitudes which together form a unitary two- and three-particle scattering matrix. In addition, the scattering amplitude must be crossing symmetric and the production and six-point amplitudes must have the appropriate driving (pole) terms. Our solution has all these properties.

Since our solution is rather complicated, we can simplify our task in two ways. First, we assume that the coupling is not so strong that there is a $\pi^+ - p$ bound state. Presumably a careful continuation of our results in the coupling constant would reproduce the results of a calculation in which bound-state channels are retained at every stage, but we do not pursue the possibility. Second, we do not consider the problem of finding all two-meson solutions of the charged scalar static model. Instead, we present just one solution, the solution which holds when there are no subtractions in dispersion relations or other arbitrary parameters in the scattering matrix.

As an introduction to the central elements of the present calculation, we trace its relationship to previous work. Amado's paper on $V\theta$ scattering in the Lee model is the most important antecedent.[3] Amado realized that, in order to treat production amplitudes successfully, one should work with a dispersion relation in the energy of one of the final-state mesons rather than in the energy of the initial-state meson. It is then necessary to retain only two-particle intermediate states in the production amplitude dispersion relation, even though one is constructing a scattering matrix which is to satisfy two- and three-particle unitarity. Another feature of Amado's calculation is that the final-state scattering amplitude is known independently of the calculated production amplitude. This decoupling of the dynamical equations arises automatically in the Lee model, but it is always a possible and apparently desirable approximation to independently compute one-meson amplitudes for final-state scattering. A third technical development is given in the paper of Bronzan and Brown on the quantitative comparison of the effects of crossing and production

* This work is supported in part by the Atomic Energy Commission under Contract AT(30-1) 2098.
[1] L. Castillejo, R. Dalitz, and F. Dyson, Phys. Rev. 101, 453 (1956).
[2] G. Wanders, Nuovo Cimento 23, 817 (1962).
[3] R. D. Amado, Phys. Rev. 122, 696 (1961).
[4] W. Pauli and S. M. Dancoff, Phys. Rev. 62, 851 (1942).
[5] S. Tomonaga, Progr. Theoret. Phys. (Kyoto) 2, 6 (1947).

in the charged scalar static model.[6] There, it was shown how to solve the scattering amplitude dispersion relation when production is present. We can thereby bypass a part of Amado's calculation which applies only to the Lee model and go on to consider theories with crossing.

II. THE FINAL-STATE SCATTERING AMPLITUDES

In this section, we summarize the elementary properties of the charged scalar static model, and we develop the one-meson approximation scattering amplitudes to be used as final-state scattering amplitudes. The charged scalar static model has a source which exists in two charged states, p (positive) and n (neutral). The source emits and absorbs mesons—π^+ and π^-—in s waves with conservation of charge. The theory is invariant under the simultaneous substitutions $p \leftrightarrow n$, $\pi^+ \leftrightarrow \pi^-$, so that all transition amplitudes are equal in pairs. The Hamiltonian of the theory is

$$H = mZ(\psi_p^+\psi_p + \psi_n^+\psi_n) + \sum_k \omega[a_k^+a_k + b_k^+b_k] + g[\psi_p^+\psi_n A + \psi_n^+A^+\psi_p] + \delta mZ(\psi_p^+\psi_p + \psi_n^+\psi_n), \quad (1)$$

where

$$A = \sum_k \frac{u(\omega)}{(2\omega\Omega)^{\frac{1}{2}}}[a_k + b_k^+], \quad (2)$$

μ is the meson mass, ω is the meson energy, $k = [\omega^2 - \mu^2]^{\frac{1}{2}}$ is the meson momentum, m is the source mass, $\psi_p(\psi_n)$ is the $p(n)$ destruction operator, $a_k(b_k)$ is the destruction operator for a $\pi^+(\pi^-)$ of momentum k, g is the renormalized meson-source coupling constant, Z is the source wave function renormalization constant, δm is the source mass counterterm, Ω is the volume of quantization, and $u(\omega)$ is the cutoff function normalized to $u(0) = 1$. The nonvanishing commutators (or anticommutators) are

$$[a_{k'}, a_k^+] = [b_{k'}, b_k^+] = \delta_{kk'}, \qquad \{\psi_p, \psi_p^+\} = \{\psi_n, \psi_n^+\} = \frac{1}{Z}. \quad (3)$$

The meson current is

$$\begin{aligned} j(t) &= \frac{(2\omega\Omega)^{\frac{1}{2}}}{u(\omega)}\left(-i\frac{d}{dt} + \omega\right)a_k(t) \\ &= \frac{(2\omega\Omega)^{\frac{1}{2}}}{u(\omega)}\left(i\frac{d}{dt} + \omega\right)b_k^+(t) \\ &= \frac{(2\omega\Omega)^{\frac{1}{2}}}{u(\omega)}([H, a_k(t)] + \omega a_k(t)) \\ &= -g\psi_n^+(t)\psi_p(t). \end{aligned} \quad (4)$$

[6] J. B. Bronzan and R. W. Brown, Ann. Phys. (N. Y.) (to be published).

As mentioned in the Introduction, we decouple our dynamical equations by using independently determined one-meson solutions of the charged scalar theory as final-state scattering amplitudes. We denote these one-meson amplitudes by $M_+(\omega)$ for $\pi^+ - p$ and $\pi^- - n$ scattering, and by $M_-(\omega)$ for $\pi^- - p$ and $\pi^+ - n$ scattering. The amplitudes are related to scattering matrix elements by

$$\begin{aligned} S_\pm &= \langle \pi_k^\pm p \text{ out} \mid \pi_{k'}^\pm p \text{ in}\rangle \\ &= \delta_{kk'} + 2\pi i\,\delta(\omega - \omega')[u^2(\omega)/2\omega\Omega]M_\pm(\omega). \end{aligned} \quad (5)$$

Contracting the "in" meson in Eq. (5), we find[7]

$$\begin{aligned} M_+(\omega) &= \frac{(2\omega\Omega)^{\frac{1}{2}}}{u(\omega)}\langle \pi_k^+ p \text{ out} \mid j^+(0) \mid p\rangle, \\ M_-(\omega) &= \frac{(2\omega\Omega)^{\frac{1}{2}}}{u(\omega)}\langle \pi_k^- p \text{ out} \mid j(0) \mid p\rangle. \end{aligned} \quad (6)$$

The dispersion relations for $M_+(\omega)$ and $M_-(\omega)$ are obtained by contracting the "out" meson in Eq. (6). We easily find that

$$M_+(\omega) = \sum_S \left\{\frac{\langle p \mid j(0)\mid S\rangle\langle S \mid j^+(0)\mid p\rangle}{E_S - m - \omega - i\epsilon} + \frac{\langle p \mid j^+(0)\mid S\rangle\langle S \mid j(0)\mid p\rangle}{E_S - m + \omega + i\epsilon}\right\}, \quad (7)$$

with the dispersion relation for $M_-(\omega)$ being given from Eq. (7) by the crossing relation

$$M_-(\omega) = M_+(-\omega - i\epsilon). \quad (8)$$

In deriving Eq. (7), use must be made of the time-translation equation

$$j(t) = e^{iHt}j(0)e^{-iHt}. \quad (9)$$

In addition, Eq. (4) is used to show that there is no equal times commutator contribution to Eq. (7). Since $M_\pm(\omega)$ are to be one-meson amplitudes, we include only zero- and one-meson "out" states in the sum in Eq. (7). We then obtain[8]

$$\begin{aligned} M_+(\omega) &= \frac{g^2}{\omega} + \frac{1}{\pi}\int_\mu^\infty \frac{d\omega_1 k_1 u^2(\omega_1)}{4\pi} \times \left[\frac{|M_+(\omega_1)|^2}{\omega_1 - \omega - i\epsilon} + \frac{|M_-(\omega_1)|^2}{\omega_1 + \omega + i\epsilon}\right] \\ &= \frac{g^2}{\omega} + \frac{1}{\pi}\left\{-\int_{-\infty}^{-\mu} + \int_\mu^\infty\right\} \times \frac{d\omega_1 k_1 u^2(\omega_1)\,|M_+(\omega_1)|^2}{4\pi(\omega_1 - \omega - i\epsilon)}. \end{aligned} \quad (10)$$

[7] The contraction formalism for static theories is reviewed in Ref. 3.

[8] The passage to infinite volume is accomplished by

$$\frac{1}{\Omega}\sum_k \to \frac{1}{(2\pi)^3}\int d^3k = \frac{1}{2\pi^2}\int d\omega\, k\omega.$$

The second form of the equation follows from Eq. (8) and the reflection property $u^2(-\omega) = u^2(\omega)$, which we assume for simplicity. Equation (10) is the variant of the Low equation solved by Castillejo, Dalitz, and Dyson.[1] Since we are constructing only one two-meson solution, we need only one particular solution of Eq. (10), and we choose the solution with no CDD (Castillejo, Dalitz, and Dyson) poles. This solution is

$$M_+(\omega) = M_-(-\omega) = \frac{g^2}{\omega[1 + \alpha(\omega)]},$$
$$\alpha(\omega) = -\frac{2\omega g^2}{\pi}\int_\mu^\infty \frac{d\omega_1 k_1 u^2(\omega_1)}{4\pi\omega_1(\omega_1^2 - \omega^2 - i\epsilon)}. \tag{11}$$

The amplitudes $M_\pm(\omega)$ are related to the one-meson approximation real scattering phase shifts by

$$[ku^2(\omega)/4\pi]M_\pm(\omega) = \sin\delta_\pm(\omega)e^{i\delta_\pm(\omega)}. \tag{12}$$

We require that the cutoff function vanish at infinite energy, and that there is no $\pi^+ - p$ bound state for the range of g we consider. It then follows that we can choose the phase shifts so that $\delta_\pm(\mu) = \delta_\pm(\infty) = 0$. In later sections we have occasion to use the functions

$$\Delta_\pm(z) = \exp\left[\frac{z}{\pi}\int_\mu^\infty \frac{d\omega_1\ \delta_\pm(\omega_1)}{\omega_1(\omega_1 - z)}\right],$$
$$P\Delta_\pm(\omega) = \exp\left[\frac{\omega}{\pi}P\int_\mu^\infty \frac{d\omega_1\ \delta_\pm(\omega_1)}{\omega_1(\omega_1 - \omega)}\right]. \tag{13}$$

III. TWO-MESON APPROXIMATION SCATTERING AMPLITUDE

We are now prepared to begin the derivation of the two-meson solution. We first introduce the production amplitudes. P_- is the amplitude for $\pi^- + p \to \pi^+ + \pi^- + n$ and $\pi^+ + n \to \pi^- + \pi^+ + p$, and P_+ is the amplitude for $\pi^+ + p \to \pi^+ + \pi^+ + n$ and $\pi^- + n \to \pi^- + \pi^- + p$.[9]

$$P_-(\omega_1, \omega_2) = \frac{(2\omega_1\Omega\cdot 2\omega_2\Omega)^{\frac{1}{2}}}{u(\omega_1)u(\omega_2)}\langle\pi^+_{k_1}\pi^-_{k_2}n \text{ out } |j(0)|\ p\rangle,$$
$$P_+(\omega_1, \omega_2) = \frac{(2\omega_1\Omega\cdot 2\omega_2\Omega)^{\frac{1}{2}}}{u(\omega_1)u(\omega_2)}\langle\pi^+_{k_1}\pi^+_{k_2}n \text{ out } |j^+(0)|\ p\rangle. \tag{14}$$

We denote the two-meson scattering amplitudes by $T_\pm(\omega)$. They are the analogs of $M_\pm(\omega)$ introduced in Sec. II, and differ from the M's in that two-meson intermediate states are retained in the dispersion relation, Eq. (7). Using Eq. (14), we find that, in the two-meson approximation, Eq. (7) becomes

$$T_+(\omega) = \frac{g^2}{\omega} + \frac{1}{\pi}\int_\mu^\infty \frac{d\omega_1\ k_1u^2(\omega_1)}{4\pi}$$
$$\times\left[\frac{|T_+(\omega_1)|^2}{\omega_1 - \omega - i\epsilon} + \frac{|T_-(\omega_1)|^2}{\omega_1 + \omega + i\epsilon}\right]$$
$$+ \frac{1}{\pi^2}\int_\mu^\infty\int_\mu^\infty \frac{d\omega_1\ d\omega_2\ k_1k_2u^2(\omega_1)u^2(\omega_2)}{16\pi^2}$$
$$\times\left[\frac{|P_+(\omega_1, \omega_2)|^2}{\omega_1 + \omega_2 - \omega - i\epsilon} + \frac{|P_-(\omega_1, \omega_2)|^2}{\omega_1 + \omega_2 + \omega + i\epsilon}\right]. \tag{15}$$

As before, $T_-(\omega)$ is given from Eq. (15) by the crossing relation $T_-(\omega) = T_+(-\omega - i\epsilon)$. The relation of the two-meson amplitudes to scattering matrix elements is

$$\langle\pi^\pm_k p \text{ out } |\ \pi^\pm_k p \text{ in}\rangle$$
$$= \delta_{k'k} + 2\pi i\ \delta(\omega - \omega')[u^2(\omega)/2\omega\Omega]T_\pm(\omega),$$
$$\langle\pi^+_{k_1}\pi^\pm_{k_2}n \text{ out } |\ \pi^\pm_k p \text{ in}\rangle = 2\pi i\ \delta(\omega_1 + \omega_2 - \omega)$$
$$\times\frac{u(\omega_1)u(\omega_2)u(\omega)P_\pm(\omega_1, \omega_2)}{(2\omega_1\Omega\cdot 2\omega_2\Omega\cdot 2\omega\Omega)^{\frac{1}{2}}}. \tag{16}$$

In order to solve Eq. (15) we must know the production amplitudes, and we now turn to their determination.

IV. PRODUCTION AMPLITUDES

We first consider the amplitude P_-. Contracting the positive meson in Eq. (14) we find

$$P_-(\omega_1, \omega_2) = \frac{(2\omega_2\Omega)^{\frac{1}{2}}}{u(\omega_2)}$$
$$\times\sum_S\langle\pi^-_{k_2}n\ |j(0)|\ S\rangle\langle S\ |j(0)|\ p\rangle$$
$$\times\left\{\frac{1}{E_S - m - \omega_1 - \omega_2 - i\epsilon}\right.$$
$$\left. + \frac{1}{E_S - m + \omega_1 + i\epsilon}\right\}. \tag{17}$$

It is useful to replace P_- by another five-point function Q_-.

$$Q_-(\omega_1, \omega_2) = \frac{(2\omega_1\Omega\cdot 2\omega_2\Omega)^{\frac{1}{2}}}{u(\omega_1)u(\omega_2)}$$
$$\times\{\langle\pi^-_{k_2}n \text{ out } |j(0)|\ \pi^-_{k_1}p \text{ out}\rangle + g\ \delta_{k_1k_2}\}. \tag{18}$$

We contract the meson on the right in Eq. (18) to obtain a dispersion relation for Q_-.

[9] The two-meson states are

$$|\pi^+_{k_1}\pi^+_{k_2}n \text{ out}\rangle = (1/\sqrt{2})a^+_{k_1}(+\infty)a^+_{k_2}(+\infty)\ |n\rangle$$

and

$$|\pi^+_{k_1}\pi^-_{k_2}n \text{ out}\rangle = a^+_{k_1}(+\infty)b^+_{k_2}(+\infty)\ |n\rangle.$$

With this normalization the momentum of each meson takes on all values in sums over states.

$$Q_-(\omega_1, \omega_2) = \frac{(2\omega_2\Omega)^{\frac{1}{2}}}{u(\omega_2)} \times \sum_S \langle \pi^-_{k_2} n \text{ out } |j(0)| S\rangle\langle S |j(0)| p\rangle \times \left\{\frac{1}{E_S - m - \omega_1 + i\epsilon} + \frac{1}{E_S - m + \omega_1 - \omega_2 - i\epsilon}\right\}. \quad (19)$$

Comparing Eqs. (17) and (19) we see that

$$P_-(\omega_1, \omega_2) = Q_-(\omega_1 + \omega_2 + i\epsilon, \omega_2). \quad (20)$$

In Eq. (20) the amplitude Q_- is to be continued in its first-energy variable from the lower half-plane, where it is defined by Eq. (19), to the upper half-plane. Equation (19) shows that there is a gap between the two cuts for $\omega_2 < 2\mu$, so it is always possible to perform the continuation by reducing ω_2 below 2μ (if necessary), performing the continuation, and then restoring ω_2. The continuation has the effect of changing $|\pi^-_{k_1} p \text{ out}\rangle$ to $|\pi^-_{k_1} p \text{ in}\rangle$ in Eq. (18).

We obtain a soluble linear integral equation for Q_- by contracting the meson on the left in Eq. (18).

$$Q_-(\omega_1, \omega_2) = \frac{(2\omega_1\Omega)^{\frac{1}{2}}}{u(\omega_1)} \times \sum_S \left\{\frac{\langle n |j^+(0)| S\rangle\langle S |j(0)| \pi^-_{k_1} p \text{ out}\rangle}{E_S - m - \omega_2 - i\epsilon} + \frac{\langle n |j(0)| S\rangle\langle S |j^+(0)| \pi^-_{k_1} p \text{ out}\rangle}{E_S - m + \omega_2 - \omega_1 + i\epsilon}\right\}. \quad (21)$$

We must define the auxiliary amplitude

$$Q'_-(\omega_1, \omega_2) = \frac{(2\omega_1\Omega \cdot 2\omega_2\Omega)^{\frac{1}{2}}}{u(\omega_1)u(\omega_2)} \times \langle \pi^+_{k_2} n \text{ out } |j^+(0)| \pi^-_{k_1} p \text{ out}\rangle. \quad (22)$$

Then in Eq. (21) we retain intermediate states $|n\pi^-_{k'} \text{ out}\rangle$ in the first term and $|p\rangle$ and $|\pi^+_{k'} n \text{ out}\rangle$ in the second term. As we have argued in Sec. I, it is necessary to retain only two-particle intermediate states, because Eq. (21) is a dispersion relation in the energy of a final-state meson. We then find

$$Q_-(\omega_1, \omega_2) = \frac{g}{\omega_1 - \omega_2 - i\epsilon} \times [T^*_-(\omega_1) - T^*_+(\omega_1)] + \frac{1}{\pi}\int_\mu^\infty \frac{d\omega'\, k'u^2(\omega')}{4\pi} \times \left[\frac{T^*_+(\omega')Q_-(\omega_1, \omega')}{\omega' - \omega_2 - i\epsilon} + \frac{T^*_-(\omega')Q'_-(\omega_1, \omega')}{\omega' + \omega_2 - \omega_1 + i\epsilon}\right]. \quad (23)$$

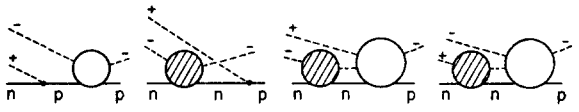


FIG. 1. Dispersion diagrams for Eq. (23). The cross-hatched amplitudes are taken to be one-meson amplitudes in Eq. (24).

The terms in this dispersion relation are displayed diagrammatically in Fig. 1. Figure 1 shows that we need not keep three-particle intermediate states in Eq. (23), because of the presence of the non-interacting meson in the diagrams. For the same reason we can replace the two-meson T amplitudes by one-meson M amplitudes in three out of four places without violating unitarity. Using Eq. (12), the integral equation for Q_- becomes

$$Q_-(\omega_1, \omega_2) = \frac{g}{\omega_1 - \omega_2 - i\epsilon}[T^*_-(\omega_1) - M^*_+(\omega_1)] + \frac{1}{\pi}\int_\mu^\infty d\omega' \left[\frac{e^{-i\delta_+(\omega')} \sin \delta_+(\omega')Q_-(\omega_1, \omega')}{\omega' - \omega_2 - i\epsilon} + \frac{e^{-i\delta_-(\omega')} \sin \delta_-(\omega')Q'_-(\omega_1, \omega')}{\omega' + \omega_2 - \omega_1 + i\epsilon}\right]. \quad (24)$$

We next eliminate Q'_- by contracting the meson on the left in Eq. (22).

$$Q'_-(\omega_1, \omega_2) = \frac{(2\omega_1\Omega)^{\frac{1}{2}}}{u(\omega_1)} \times \sum_S \left\{\frac{\langle n |j(0)| S\rangle\langle S |j^+(0)| \pi^-_{k_1} p \text{ out}\rangle}{E_S - m - \omega_2 - i\epsilon} + \frac{\langle n |j^+(0)| S\rangle\langle S |j(0)| \pi^-_{k_1} p \text{ out}\rangle}{E_S - m + \omega_2 - \omega_1 + i\epsilon}\right\} = Q_-(\omega_1, \omega_1 - \omega_2 - i\epsilon). \quad (25)$$

The last equality follows from Eq. (21). Equation (21) shows that the continuation in the second-energy variable can be carried out for $\omega_1 < 2\mu$. We now write the integral equation (24) in a form which stresses the analytic properties of Q_- as a function of its second-energy variable.

$$Q_-(\omega_1, z) = \frac{g}{\omega_1 - z}[T^*_-(\omega_1) - M^*_+(\omega_1)] + \frac{1}{\pi}\int_\mu^\infty d\omega' \left[\frac{e^{-i\delta_+(\omega')} \sin \delta_+(\omega')Q_-(\omega_1, \omega' + i\epsilon)}{\omega' - z} + \frac{e^{-i\delta_-(\omega')} \sin \delta_-(\omega')Q_-(\omega_1, \omega_1 - \omega' - i\epsilon)}{\omega' - \omega_1 + z}\right]. \quad (26)$$

Although the undetermined function T_- appears in this equation, it occurs parametrically, and we can therefore find Q_- and P_- in terms of T_-.

To construct a solution of Eq. (26), we generalize the treatment of Gartenhaus and Blankenbecler[10] to the case where the phases on the left and right cuts are different. We assume that ω_1 is real and smaller than μ. The solution for other values of ω_1 can be obtained by analytic continuation. We introduce the function $F(z)$ through the definition[11]

$$F(z)\Delta_+(z)\Delta_-(\omega_1 - z) \equiv \frac{1}{2i\pi}\int_\mu^\infty d\omega' \left[\frac{e^{-i\delta_+(\omega')}\sin\delta_+(\omega')Q_-(\omega_1, \omega' + i\epsilon)}{\omega' - z} + \frac{e^{-i\delta_-(\omega')}\sin\delta_-(\omega')Q_-(\omega_1, \omega_1 - \omega' - i\epsilon)}{\omega' - \omega_1 + z}\right]. \quad (27)$$

$F(z)$ has cuts from $-\infty$ to $\omega_1 - \mu$, and from μ to $+\infty$ in the complex z plane and is otherwise analytic. From Eq. (26) we find that its boundary values are

$$\begin{aligned} &2iF(\omega + i\epsilon)e^{i\delta_+(\omega)}\,\Delta_-(\omega_1 - \omega)P\Delta_+(\omega) \\ &\quad = Q_-(\omega_1, \omega + i\epsilon) - A(\omega + i\epsilon) \quad (\omega > \mu), \\ &2iF(\omega - i\epsilon)e^{i\delta_-(\omega_1 - \omega)}\,\Delta_+(\omega)P\Delta_-(\omega_1 - \omega) \\ &\quad = Q_-(\omega_1, \omega - i\epsilon) - A(\omega - i\epsilon) \quad (\omega < \omega_1 - \mu), \end{aligned} \quad (28)$$

where

$$A(z) = [g/(\omega_1 - z)][T^*_-(\omega_1) - M^*_+(\omega_1)]. \quad (29)$$

The discontinuity of F across the cuts is

$$\begin{aligned} &[F(\omega + i\epsilon)e^{i\delta_+(\omega)} - F(\omega - i\epsilon)e^{-i\delta_+(\omega)}]\Delta_-(\omega_1 - \omega)P\Delta_+(\omega) \\ &\quad = e^{-i\delta_+(\omega)}\sin\delta_+(\omega)Q_-(\omega_1, \omega + i\epsilon) \quad (\omega > \mu), \\ &[F(\omega + i\epsilon)e^{-i\delta_-(\omega_1 - \omega)} - F(\omega - i\epsilon)e^{i\delta_-(\omega_1 - \omega)}]\Delta_+(\omega)P\Delta_-(\omega_1 - \omega) \\ &\quad = -e^{-i\delta_-(\omega_1 - \omega)}\sin\delta_-(\omega_1 - \omega)Q_-(\omega_1, \omega - i\epsilon) \\ &\qquad (\omega < \omega_1 - \mu). \end{aligned} \quad (30)$$

Eliminating Q_- from Eqs. (29) and (30),

$$\begin{aligned} &F(\omega + i\epsilon) - F(\omega - i\epsilon) = \frac{\sin\delta_+(\omega)A(\omega + i\epsilon)}{\Delta_-(\omega_1 - \omega)P\Delta_+(\omega)} \quad (\omega > \mu), \\ &F(\omega + i\epsilon) - F(\omega - i\epsilon) = -\frac{\sin\delta_-(\omega_1 - \omega)A(\omega - i\epsilon)}{\Delta_+(\omega)P\Delta_-(\omega_1 - \omega)} \quad (\omega < \omega_1 - \mu). \end{aligned} \quad (31)$$

Therefore,

$$F(z) = \frac{K(z)}{2i} - \frac{1}{2\pi i}\int_{-\infty}^{\omega_1 - \mu}\frac{d\omega'\sin\delta_-(\omega_1 - \omega')A(\omega' - i\epsilon)}{\Delta_+(\omega')P\Delta_-(\omega_1 - \omega')[\omega' - z]} + \frac{1}{2\pi i}\int_\mu^\infty \frac{d\omega'\sin\delta_+(\omega')A(\omega' + i\epsilon)}{\Delta_-(\omega_1 - \omega')P\Delta_+(\omega')[\omega' - z]}, \quad (32)$$

where $K(z)$ is an arbitrary polynomial. Then Q_- becomes

$$\begin{aligned} Q_-(\omega_1, \omega_2) &= K(\omega_2)\Delta_+(\omega_2 + i\epsilon)\Delta_-(\omega_1 - \omega_2) + g[T^*_-(\omega_1) - M^*_+(\omega_1)] \\ &\times\left[\frac{1}{\omega_1 - \omega_2 - i\epsilon} + \frac{\Delta_+(\omega_2 + i\epsilon)\Delta_-(\omega_1 - \omega_2)}{\pi}\left\{-\int_{-\infty}^{\omega_1 - \mu}\frac{d\omega'\sin\delta_-(\omega_1 - \omega')}{\Delta_+(\omega')P\Delta_-(\omega_1 - \omega')[(\omega' - \omega_2)(\omega_1 - \omega')]}\right.\right. \\ &\left.\left. + \int_\mu^\infty \frac{d\omega'\sin\delta_+(\omega')}{\Delta_-(\omega_1 - \omega')P\Delta_+(\omega')[(\omega' - \omega_2 - i\epsilon)(\omega_1 - \omega')]}\right\}\right]. \end{aligned} \quad (33)$$

The integrals in Eq. (33) may be evaluated by noting that

$$\begin{aligned} \frac{\sin\delta_-(\omega_1 - \omega)}{P\Delta_-(\omega_1 - \omega)} &= \frac{1}{2i}\left[\frac{1}{\Delta_-(\omega_1 - \omega - i\epsilon)} - \frac{1}{\Delta_-(\omega_1 - \omega + i\epsilon)}\right], \\ \frac{\sin\delta_+(\omega)}{P\Delta_+(\omega)} &= \frac{1}{2i}\left[\frac{1}{\Delta_+(\omega - i\epsilon)} - \frac{1}{\Delta_+(\omega + i\epsilon)}\right]. \end{aligned} \quad (34)$$

Thus,

$$\begin{aligned} Q_-(\omega_1, \omega_2) &= K(\omega_2)\Delta_+(\omega_2 + i\epsilon)\Delta_-(\omega_1 - \omega_2) \\ &+ g[T^*_-(\omega_1) - M^*_+(\omega_1)] \\ &\times\left[\frac{1}{\omega_1 - \omega_2 - i\epsilon} + \Delta_+(\omega_2 + i\epsilon)\Delta_-(\omega_1 - \omega_2)\frac{1}{2i\pi}\right. \\ &\left.\times\int_C \frac{d\omega'}{\Delta_+(\omega')\,\Delta_-(\omega_1 - \omega')[(\omega' - \omega_1)(\omega' - \omega_2 - i\epsilon)]}\right], \end{aligned} \quad (35)$$

where C is the contour around the cuts of the Δ's, as shown in Fig. 2. Since $\Delta_\pm$ approach a constant

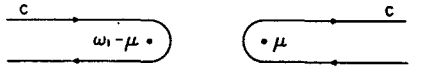


FIG. 2. Contour of integration for Eqs. (35), (38), and (39).

[10] R. Blankenbecler and S. Gartenhaus, Phys. Rev. 116, 1297 (1959).

[11] We do not display the dependence of F on ω_1.

at infinity, the great circle at infinity gives no contribution, and by Cauchy's theorem

$$Q_-(\omega_1, \omega_2) = K(\omega_2)\Delta_+(\omega_2 + i\epsilon)\Delta_-(\omega_1 - \omega_2) + \frac{g}{\omega_1 - \omega_2 - i\epsilon}[T^*_-(\omega_1) - M^*_+(\omega_1)] \times \frac{\Delta_+(\omega_2 + i\epsilon)\Delta_-(\omega_1 - \omega_2)}{\Delta_+(\omega_1)}. \quad (36)$$

The arbitrary polynomial $K(z)$ can be nonzero only if $K(z)\Delta_+(z)\Delta_-(\omega_1 - z)$ is a solution of the homogeneous equation obtained from Eq. (26). There are two points to be considered here. First, if K is of too high a degree, the integrals in Eq. (26) may fail to exist. The question hinges on how fast the one-meson phase shifts decrease at infinity, that is, on how fast the cutoff function decreases at infinity. Let us simply assume that the cutoff function has been chosen so that the integrals exist for low-degree polynomials, and proceed to the second point. The second point is that $K(z)\Delta_+(z)\Delta_-(\omega_1 - z)$ may be a solution of a subtracted integral equation rather than the unsubtracted equation we are considering. We therefore try evaluating the integral

$$I = \frac{1}{\pi}\int_\mu^\infty d\omega' \frac{e^{-i\delta_+(\omega')}\sin\delta_+(\omega')K(\omega')\Delta_+(\omega' + i\epsilon)\Delta_-(\omega_1 - \omega')}{\omega' - z} + \frac{1}{\pi}\int_\mu^\infty d\omega' \frac{e^{-i\delta_-(\omega')}\sin\delta_-(\omega')K(\omega_1 - \omega')\Delta_+(\omega_1 - \omega')\Delta_-(\omega' + i\epsilon)}{\omega' + z - \omega_1}. \quad (37)$$

Equation (37) is easily changed to a contour integral,

$$I = \frac{1}{2i\pi}\int_C \frac{d\omega'}{\omega' - z} K(\omega')\Delta_+(\omega')\Delta_-(\omega_1 - \omega'). \quad (38)$$

In order to close the contour at infinity and calculate I by Cauchy's theorem, one must add an appropriate entire function to the numerator of Eq. (38), so that the numerator of Eq. (38) vanishes at infinity and the great circle gives no contribution. (Since the required entire function does not possess the cuts of $\Delta_\pm$, it gives no contribution to I.) This requirement eliminates the possibility of a solution of the unsubtracted homogeneous equation. For instance, if $K(z) = K$, a constant, one finds

$$\begin{aligned} I &= \frac{1}{2\pi i}\int_C \frac{d\omega'}{\omega' - z}[K\Delta_+(\omega')\Delta_-(\omega_1 - \omega') - K\Delta_+(\infty)\Delta_-(\infty)] \\ &= K\Delta_+(z)\Delta_-(\omega_1 - z) - K\Delta_+(\infty)\Delta_-(\infty) \\ &\neq K\Delta_+(z)\Delta_-(\omega_1 - z). \end{aligned} \quad (39)$$

Thus Q_- and P_- are uniquely determined:

$$\begin{aligned} Q_-(\omega_1, \omega_2) &= \frac{g}{\omega_1 - \omega_2 - i\epsilon}[T^*_-(\omega_1) - M^*_+(\omega_1)] \times \frac{\Delta_+(\omega_2 + i\epsilon)\Delta_-(\omega_1 - \omega_2 - i\epsilon)}{\Delta_+(\omega_1 - i\epsilon)}, \\ P_-(\omega_1, \omega_2) &= \frac{g}{\omega_1}[T_-(\omega_1 + \omega_2) - M_+(\omega_1 + \omega_2)] \times \frac{\Delta_-(\omega_1 + i\epsilon)\Delta_+(\omega_2 + i\epsilon)}{\Delta_+(\omega_1 + \omega_2 + i\epsilon)}. \end{aligned} \quad (40)$$

A direct substitution of Q_- into Eq. (26) verifies the fact that we have a solution. In writing Eq. (40) we have given all variables their physical imaginary parts.

The second production amplitude, P_+, can be obtained by methods analogous to those used for P_-. The derivation is somewhat simplified by the indentity of the final-state mesons, and we therefore simply state some of the key equations. We first replace P_+ by a second five-point amplitude Q_+:

$$P_+(\omega_1, \omega_2) = 1/\sqrt{2}\, Q_+(\omega_1 + \omega_2 + i\epsilon, \omega_2), \quad (41)$$

where

$$Q_+(\omega_1, \omega_2) = \frac{(2\omega_1\Omega\cdot 2\omega_2\Omega)^{\frac{1}{2}}}{u(\omega_1)u(\omega_2)} \times \{\langle \pi^+_{k_2} n \text{ out} | j(0) | \pi^+_{k_1} p \text{ out}\rangle + g\,\delta_{k_1k_2}\}. \quad (42)$$

The integral equation for Q_+ is

$$\begin{aligned} Q_+(\omega_1, \omega_2) &= g[T^*_+(\omega_1) - M^*_-(\omega_1)]\left[\frac{1}{\omega_1 - \omega_2 - i\epsilon} + \frac{1}{\omega_2 + i\epsilon}\right] \\ &+ \frac{1}{\pi}\int_\mu^\infty d\omega'\, e^{-i\delta_-(\omega')}\sin\delta_-(\omega')Q_+(\omega_1, \omega' + i\epsilon) \times \left[\frac{1}{\omega' - \omega_2 - i\epsilon} + \frac{1}{\omega' - \omega_1 + \omega_2 + i\epsilon}\right], \end{aligned} \quad (43)$$

and its solution is

$$Q_+(\omega_1,\omega_2) = \frac{g\omega_1}{\omega_2(\omega_1-\omega_2-i\epsilon)}\,[T_+^*(\omega_1) - M^*(\omega_1)]\,\frac{\Delta_-(\omega_2+i\epsilon)\Delta_-(\omega_1-\omega_2-i\epsilon)}{\Delta_-(\omega_1-i\epsilon)},$$
$$P_+(\omega_1,\omega_2) = \frac{g(\omega_1+\omega_2)}{\sqrt{2}\,\omega_1\omega_2}\,[T_+(\omega_1+\omega_2) - M_-(\omega_1+\omega_2)]\,\frac{\Delta_-(\omega_1+i\epsilon)\Delta_-(\omega_2+i\epsilon)}{\Delta_-(\omega_1+\omega_2+i\epsilon)}. \tag{44}$$

V. SOLUTION OF THE SCATTERING AMPLITUDE DISPERSION RELATION

From Eqs. (15), (40), and (44), the dispersion relation for $T_+(\omega)$ is

$$T_+(\omega) = \frac{g^2}{\omega} + \frac{1}{\pi}\int_\mu^\infty \frac{d\omega_1 k_1 u^2(\omega_1)}{4\pi}\times\left[\frac{|T_+(\omega_1)|^2}{\omega_1-\omega-i\epsilon} + \frac{|T_-(\omega_1)|^2}{\omega_1+\omega+i\epsilon}\right] + \frac{g^2}{\pi^2}\int_\mu^\infty\int_\mu^\infty \frac{d\omega_1\,d\omega_2\,k_1k_2u^2(\omega_1)u^2(\omega_2)}{16\pi^2}\times\left[\frac{(\omega_1+\omega_2)^2}{2\omega_1^2\omega_2^2(\omega_1+\omega_2-\omega-i\epsilon)}\left|\{T_+(\omega_1+\omega_2) - M_-(\omega_1+\omega_2)\}\,\frac{\Delta_-(\omega_1+i\epsilon)\Delta_-(\omega_2+i\epsilon)}{\Delta_-(\omega_1+\omega_2+i\epsilon)}\right|^2 + \frac{1}{\omega_1^2(\omega_1+\omega_2+\omega+i\epsilon)}\left|\{T_-(\omega_1+\omega_2) - M_+(\omega_1+\omega_2)\}\,\frac{\Delta_-(\omega_1+i\epsilon)\Delta_+(\omega_2+i\epsilon)}{\Delta_+(\omega_1+\omega_2+i\epsilon)}\right|^2\right]. \tag{45}$$

To solve this equation we consider the function $f(\omega)$:

$$f(\omega) \equiv \frac{2}{[g^2/\omega T_+(\omega)] - [g^2/\omega M_-(\omega)]}, \tag{46}$$

which has the following properties.

(1) $f(0) = 1$.

(2) $f(\omega)$ is meromorphic in the cut ω plane with cuts from $-\infty$ to $-\mu$ and μ to ∞, and with additional branch points at $\pm 2\mu$. When use is made of the crossing relations

$$T_-(\omega) = T_+(-\omega-i\epsilon),\quad M_-(\omega) = M_+(-\omega-i\epsilon), \tag{47}$$

we find from Eq. (45) that $f(\omega)$ has no discontinuity arising from the elastic cuts, and its discontinuities from the inelastic cuts of $T_+(\omega)$ are determined.

$$f(\omega+i\epsilon) - f(\omega-i\epsilon) = 2i\rho_+(\omega)\quad(\omega>2\mu),$$
$$f(\omega+i\epsilon) - f(\omega-i\epsilon) = 2i\rho_-(-\omega)\quad(\omega<-2\mu), \tag{48}$$

where

$$\rho_+(\omega) = \frac{\omega^3}{16\pi^3}\left|\frac{M_-(\omega)}{\Delta_-(\omega)}\right|^2\int_\mu^{\omega-\mu}\frac{d\omega_1\,k_1k_{-1}u^2(\omega_1)u^2(\omega_{-1})}{\omega_1^2\omega_{-1}^2}\times|\Delta_-(\omega_1+i\epsilon)\Delta_-(\omega_{-1}+i\epsilon)|^2,$$
$$\rho_-(\omega) = \frac{\omega}{8\pi^3}\left|\frac{M_+(\omega)}{\Delta_+(\omega)}\right|^2\int_\mu^{\omega-\mu}\frac{d\omega_1\,k_1k_{-1}u^2(\omega_1)u^2(\omega_{-1})}{\omega_1^2}\times|\Delta_-(\omega_1+i\epsilon)\Delta_+(\omega_{-1}+i\epsilon)|^2, \tag{49}$$

and $\omega_{-1} = \omega-\omega_1$, $k_{-1} = [\omega_{-1}^2-\mu^2]^{\frac{1}{2}}$.

(3) We observe that $\rho_+(\omega)$ vanishes at infinity.

$$\rho_+(\omega) = \frac{\omega^3}{8\pi^3}\left|\frac{M_-(\omega)}{\Delta_-(\omega)}\right|^2\int_\mu^{\frac{1}{2}\omega}\frac{d\omega_1\,k_1k_{-1}u^2(\omega_1)u^2(\omega_{-1})}{\omega_1^2\omega_{-1}^2}\times|\Delta_-(\omega_1+i\epsilon)\Delta_-(\omega_{-1}+i\epsilon)|^2 \le \frac{\lambda\omega^3}{8\pi^3}\left|\frac{M_-(\omega)}{\Delta_-(\omega)}\right|^2\frac{4}{\omega^2}\left[\left(\frac{\omega}{2}\right)^2-\mu^2\right]^{\frac{1}{2}}u^2(\bar\omega)\times\int_\mu^{\frac{1}{2}\omega}\frac{d\omega_1\,k_1u^2(\omega_1)}{\omega_1^2}, \tag{50}$$

where λ is the maximum value of

$$|\Delta_-(\omega_1+i\epsilon)\Delta_-(\omega_{-1}+i\epsilon)|^2$$

in the range $\mu\le\omega$, $\omega_1<\infty$, and $\frac{1}{2}\omega\le\bar\omega\le\omega$. Since $M_-(\omega)$ vanishes like ω^{-1} at infinity, $\rho_+(\omega)$ vanishes at infinity provided $u^2(\omega)$ vanishes at infinity. A slightly modified argument shows that $\rho_-(\omega)$ vanishes at infinity.

(4) We assume that $f(\omega)$ has no poles, that is, the equation $T_+(\omega) = M_-(\omega)$ has no solutions. Poles of this type correspond to arbitrary parameters in the scattering matrix, and our assumption singles out a particular two-meson solution.

(5) We assume that $f(\omega)$ approaches a constant at infinity. Since $\omega M_-(\omega)$ approaches a constant, our assumption is equivalent to the demand that $\omega T_+(\omega)$ approach a constant at infinity. This asymptotic behavior is consistent with the unsubtracted dispersion relation, Eq. (45).

These considerations determine $f(\omega)$:

$$f(\omega) = 1 + \frac{\omega}{\pi}\int_{2\mu}^\infty\frac{d\omega_1}{\omega_1}\left[\frac{\rho_+(\omega_1)}{\omega_1-\omega-i\epsilon} + \frac{\rho_-(\omega_1)}{\omega_1+\omega+i\epsilon}\right] \equiv 1+\omega C(\omega). \tag{51}$$

We now have the two-meson scattering amplitude from Eq. (46).[12]

$$T_+(\omega) = g^2\omega^{-1}\{\alpha(\omega) + [1-\omega C(\omega)][1+\omega C(\omega)]^{-1}\}^{-1}. \tag{52}$$

[12] We write Eq. (52) in this form to stress its similarity to amplitudes given in Refs. 3 and 6 which also include production.

VI. SIX-POINT AMPLITUDES

In deriving Eq. (52), we had to assume that we could replace the final-state scattering T amplitudes by M amplitudes. While it is plausible that this may be done without violating unitarity, we have yet to show explicitly that we have constructed an unitary scattering matrix. In order to do this we must construct the six-point amplitudes.

We define the connected amplitudes

$$R_-(\omega_1,\omega_2,\omega_3)=\frac{(2\omega_1\Omega\cdot 2\omega_2\Omega\cdot 2\omega_3\Omega)^{\frac{1}{2}}}{u(\omega_1)u(\omega_2)u(\omega_3)}\times\Big[\langle\pi^-_{k_1}n \text{ out } |j(0)|\,\pi^-_{k_2}\pi^+_{k_3}n \text{ out}\rangle-\delta_{k_1k_2}\frac{u(\omega_3)}{(2\omega_3\Omega)^{\frac{1}{2}}}M^{\pm}(\omega_3)\Big],\tag{53}$$

$$R_+(\omega_1,\omega_2,\omega_3)=\frac{(2\omega_1\Omega\cdot 2\omega_2\Omega\cdot 2\omega_3\Omega)^{\frac{1}{2}}}{u(\omega_1)u(\omega_2)u(\omega_3)}\times\Big[\langle\pi^+_{k_1}n \text{ out } |j(0)|\,\pi^+_{k_2}\pi^+_{k_3}n \text{ out}\rangle-\frac{1}{\sqrt{2}}\delta_{k_1k_2}\frac{u(\omega_2)}{(2\omega_2\Omega)^{\frac{1}{2}}}M^{\pm}(\omega_2)-\frac{1}{\sqrt{2}}\delta_{k_1k_3}\frac{u(\omega_3)}{(2\omega_3\Omega)^{\frac{1}{2}}}M^{\pm}(\omega_3)\Big].$$

Since we are dealing with three-particle states, consistency requires that we interpret the disconnected amplitudes we subtract off in Eq. (53) as being M amplitudes. We obtain dispersion relations by contracting the mesons on the left. As before, we need retain only two-particle intermediate states, and we find[13]

$$R_-(\omega_1,\omega_2,\omega_3)=\frac{gP^*_{\pm}(\omega_3,\omega_2)}{\omega_2+\omega_3-\omega_1-i\epsilon}+\frac{1}{\pi}\int_\mu^\infty d\omega'\times\Big[\frac{e^{-i\delta_+(\omega')}\sin\delta_+(\omega')R_-(\omega'+i\epsilon,\omega_2,\omega_3)}{\omega'-\omega_1-i\epsilon}+\frac{e^{-i\delta_-(\omega')}\sin\delta_-(\omega')R_-(\omega_2+\omega_3-\omega'-i\epsilon,\omega_2,\omega_3)}{\omega'-\omega_2-\omega_3+\omega_1+i\epsilon}\Big],\tag{54}$$

$$R_+(\omega_1,\omega_2,\omega_3)=gP^*_+(\omega_2,\omega_3)\Big[\frac{1}{\omega_2+\omega_3-\omega_1-i\epsilon}+\frac{1}{\omega_1+i\epsilon}\Big]+\frac{1}{\pi}\int_\mu^\infty d\omega'\,e^{-i\delta_-(\omega')}\sin\delta_-(\omega')R_+(\omega'+i\epsilon,\omega_2,\omega_3)\Big[\frac{1}{\omega'-\omega_1-i\epsilon}+\frac{1}{\omega'-\omega_2-\omega_3+\omega_1+i\epsilon}\Big].$$

The solutions of these equations may be read off immediately from the solutions of Eqs. (26) and (43).

$$R_-(\omega_1,\omega_2,\omega_3)=\frac{gP^*_{\pm}(\omega_3,\omega_2)\Delta_+(\omega_1+i\epsilon)\Delta_-(\omega_2+\omega_3-\omega_1-i\epsilon)}{(\omega_2+\omega_3-\omega_1-i\epsilon)\Delta_+(\omega_2+\omega_3-i\epsilon)}=\frac{g^2[T^*_{\pm}(\omega_2+\omega_3)-M^*_{\pm}(\omega_2+\omega_3)]\Delta_-(\omega_3-i\epsilon)\Delta_+(\omega_2-i\epsilon)\Delta_-(\omega_2+\omega_3-\omega_1-i\epsilon)\Delta_+(\omega_1+i\epsilon)}{\omega_3(\omega_2+\omega_3-\omega_1-i\epsilon)[\Delta_+(\omega_2+\omega_3-i\epsilon)]^2},$$

$$R_+(\omega_1,\omega_2,\omega_3)=\frac{g(\omega_2+\omega_3)P^*_+(\omega_2,\omega_3)\Delta_-(\omega_1+i\epsilon)\Delta_-(\omega_2+\omega_3-\omega_1-i\epsilon)}{\omega_1(\omega_2+\omega_3-\omega_1-i\epsilon)\Delta_-(\omega_2+\omega_3-i\epsilon)}=\frac{g^2[T^*_+(\omega_2+\omega_3)-M^*_{\pm}(\omega_2+\omega_3)](\omega_2+\omega_3)^2\Delta_-(\omega_3-i\epsilon)\Delta_-(\omega_2-i\epsilon)\Delta_-(\omega_2+\omega_3-\omega_1-i\epsilon)\Delta_-(\omega_1+i\epsilon)}{\sqrt{2}\,\omega_1\omega_2\omega_3(\omega_2+\omega_3-\omega_1-i\epsilon)[\Delta_-(\omega_2+\omega_3-i\epsilon)]^2}.\tag{55}$$

The R amplitudes do not have the correct boundary conditions to give scattering matrix elements as they stand. To obtain these elements, we note that by contracting the positive "in" meson we find

$$\begin{aligned}&\langle\pi^-_{k_1}\pi^+_{k_2}n \text{ in } |\pi^-_{k_3}\pi^+_{k_4}n \text{ out}\rangle\\&=\delta_{k_1k_3}\,\delta_{k_2k_4}-2\pi i\,\delta_{k_2k_4}\,\delta(\omega_1-\omega_3)\times\frac{u^2(\omega_1)}{2\omega_1\Omega}M^*_+(\omega_1)-2\pi i\,\delta(\omega_1+\omega_2-\omega_3-\omega_4)\\&\times\frac{u(\omega_2)}{(2\omega_2\Omega)^{\frac{1}{2}}}\langle\pi^-_{k_1}n \text{ in } |j(0)|\,\pi^-_{k_3}\pi^+_{k_4}n \text{ out}\rangle.\end{aligned}\tag{56}$$

We write the last matrix element in terms of a sum over state:

$$\begin{aligned}&\langle\pi^-_{k_1}n \text{ in } |j(0)|\,\pi^-_{k_3}\pi^+_{k_4}n \text{ out}\rangle\\&=\sum_S\langle\pi'_{k_1}n \text{ in } |\,S \text{ out}\rangle\times\langle S \text{ out } |j(0)|\,\pi^-_{k_3}\pi^+_{k_4}n \text{ out}\rangle\\&=\sum_{k'}\langle\pi^-_{k_1}n \text{ in } |\,\pi^-_{k'}n \text{ out}\rangle\times\langle\pi^-_{k'}n \text{ out } |j(0)|\,\pi^-_{k_3}\pi^+_{k_4}n \text{ out}\rangle.\end{aligned}\tag{57}$$

[13] In writing Eq. (54) an amplitude R'_-, analogous to Q'_-, must be introduced and eliminated in favor or R_-.

It is consistent with our approximation to keep only the two-particle intermediate state, and to use Eq. (5) to replace the two-particle matrix element by an M amplitude. Use of Eq. (53) then gives the six-point scattering matrix element.

$$\begin{aligned}
&\langle \pi^-_{k_3}\pi^+_{k_4} n \text{ out} \mid \pi^-_{k_1}\pi^+_{k_2} n \text{ in}\rangle \\
&= \delta_{k_1k_3}\,\delta_{k_2k_4} + 2\pi i\,\delta_{k_2k_4}\,\delta(\omega_1-\omega_3) \\
&\times \frac{u^2(\omega_1)}{2\omega_1\Omega} M_+(\omega_1) + 2\pi i\,\delta_{k_1k_3}\,\delta(\omega_2-\omega_4) \\
&\times \frac{u^2(\omega_2)}{2\omega_2\Omega} M_-(\omega_2) + (2\pi i)^2\,\delta(\omega_1-\omega_3)\,\delta(\omega_2-\omega_4) \\
&\times \frac{u^2(\omega_1)u^2(\omega_2)}{2\omega_1\Omega\cdot 2\omega_2\Omega} M_+(\omega_1)M_-(\omega_2) \\
&+ 2\pi i\,\delta(\omega_1+\omega_2-\omega_3-\omega_4) \\
&\times \frac{u(\omega_1)u(\omega_2)u(\omega_3)u(\omega_4)}{(2\omega_1\Omega\cdot 2\omega_2\Omega\cdot 2\omega_3\Omega\cdot 2\omega_4\Omega)^{\frac{1}{2}}} \\
&\times e^{2i\delta_+(\omega_1)} R^*_-(\omega_1,\omega_3,\omega_4). \qquad (58)
\end{aligned}$$

Similar considerations yield the equation

$$\begin{aligned}
&\langle \pi^+_{k_3}\pi^+_{k_4} n \text{ out} \mid \pi^+_{k_1}\pi^+_{k_2} n \text{ in}\rangle \\
&= \tfrac{1}{2}(\delta_{k_1k_4}\,\delta_{k_2k_3} + \delta_{k_1k_3}\,\delta_{k_2k_4}) \\
&+ \frac{2\pi i}{2}\,[\delta_{k_1k_4}\,\delta(\omega_2-\omega_3) \\
&+ \delta_{k_1k_3}\,\delta(\omega_2-\omega_4)]\,\frac{u^2(\omega_2)}{2\omega_2\Omega} M_-(\omega_2) \\
&+ \frac{2\pi i}{2}\,[\delta_{k_2k_3}\,\delta(\omega_1-\omega_4) \\
&+ \delta_{k_2k_4}\,\delta(\omega_1-\omega_3)]\,\frac{u^2(\omega_1)}{2\omega_1\Omega} M_-(\omega_1) \\
&+ \frac{(2\pi i)^2}{2}\,[\delta(\omega_1-\omega_3)\,\delta(\omega_2-\omega_4) \\
&+ \delta(\omega_1-\omega_4)\,\delta(\omega_2-\omega_3)]\,\frac{u^2(\omega_1)u^2(\omega_2)}{2\omega_1\Omega\cdot 2\omega_2\Omega} \\
&\times M_-(\omega_1)M_-(\omega_2) + \frac{2\pi i}{\sqrt{2}}\,\delta(\omega_1+\omega_2-\omega_3-\omega_4 \\
&\times \frac{u(\omega_1)u(\omega_2)u(\omega_3)u(\omega_4)}{(2\omega_1\Omega\cdot 2\omega_1\Omega\cdot 2\omega_3\Omega\cdot 2\omega_4\Omega)^{\frac{1}{2}}}\, e^{2i\delta_-(\omega_1)} R^*_+(\omega_1,\omega_3,\omega_4). \qquad (59)
\end{aligned}$$

VII. UNITARITY

Our object in this section is to verify that our two- and three-particle scattering matrix is unitary. The statement of unitarity is

$$SS^+ = \mathbf{1}, \quad S_{ij} = \langle i \text{ out} \mid j \text{ in}\rangle, \quad \mathbf{1}_{ij} = \langle i \text{ out} \mid j \text{ out}\rangle. \qquad (60)$$

The matrix element of Eq. (60) between two-particle states requires that $T_\pm(\omega)$ satisfy Eq. (15) and the crossed dispersion relation. Our solution has this property by construction. Similarly, the matrix elements of Eq. (60) with a two-particle state on one side and a three-particle state on the other side specify Im P_- and Im P_+, and the matrix elements between three-particle states specify Im R_- and Im R_+. Although we have verified all four unitarity relations, in this paper we reproduce only the calculation for Im P_-. For this case, Eq. (60) requires that

$$\begin{aligned}
0 = &\sum_{k'} \langle \pi^+_{k_1}\pi^-_{k_2} n \text{ out} \mid \pi^-_{k'} p \text{ in}\rangle \\
&\times \langle \pi^-_{k'} p \text{ out} \mid \pi^-_{k'} p \text{ in}\rangle^* \\
&+ \sum_{k'k''} \langle \pi^+_{k_1}\pi^-_{k_2} n \text{ out} \mid \pi^+_{k'}\pi^-_{k''} n \text{ in}\rangle \\
&\times \langle \pi^-_{k'} p \text{ out} \mid \pi^+_{k'}\pi^-_{k''} n \text{ in}\rangle^*. \qquad (61)
\end{aligned}$$

Using time reversal invariance on the last matrix element, and Eqs. (16) and (58), Eq. (61) becomes

$$\begin{aligned}
0 = &\sum_{k'} \left\{ 2\pi i\,\delta(\omega_1+\omega_2-\omega')\,\frac{u(\omega_1)u(\omega_2)u(\omega')P_-(\omega_1,\omega_2)}{(2\omega_1\Omega\cdot 2\omega_2\Omega\cdot 2\omega'\Omega)^{\frac{1}{2}}} \right\} \\
&\times \left\{ \delta_{kk'} - 2\pi i\,\delta(\omega-\omega')\,\frac{u^2(\omega)}{2\omega\Omega}\,T^*_-(\omega) \right\} \\
&+ \sum_{k'k''} \Big\{ \delta_{k_1k'}\,\delta_{k_2k''} + 2\pi i\,\delta_{k_1k'}\,\delta(\omega_2-\omega'') \\
&\times \frac{u^2(\omega_2)}{2\omega_2\Omega} M_+(\omega_2) + 2\pi i\,\delta_{k_2k''}\,\delta(\omega_1-\omega') \\
&\times \frac{u^2(\omega_1)}{2\omega_1\Omega} M_-(\omega_1) + (2\pi i)^2\,\delta(\omega_1-\omega')\,\delta(\omega_2-\omega'') \\
&\times M_+(\omega_2)M_-(\omega_1) + 2\pi i\,\delta(\omega_1+\omega_2-\omega'-\omega'') \\
&\times \frac{u(\omega_1)u(\omega_2)u(\omega')u(\omega'')e^{2i\delta_+(\omega'')}}{(2\omega_1\Omega\cdot 2\omega_2\Omega\cdot 2\omega'\Omega\cdot 2\omega''\Omega)^{\frac{1}{2}}} R^*_-(\omega'',\omega_2,\omega_1) \Big\} \\
&\times \Big\{ -2\pi i\,\delta(\omega'+\omega''-\omega) \\
&\times \frac{u(\omega')u(\omega'')u(\omega)}{(2\omega'\Omega\cdot 2\omega''\Omega\cdot 2\omega\Omega)^{\frac{1}{2}}} P^*_-(\omega',\omega'') \Big\}. \qquad (62)
\end{aligned}$$

When the summations are carried out, we find that unitarity requires that

$$\begin{aligned}
\text{Im } P_-(\omega_1,\omega_2) = &\frac{ku^2(\omega)}{4\pi} P_-(\omega_1,\omega_2)T^*_-(\omega) \\
&+ \frac{k_1u^2(\omega_1)}{4\pi} M_-(\omega_1)P^*_-(\omega_1,\omega_2) \\
&+ \frac{k_2u^2(\omega_2)}{4\pi} M_+(\omega_2)P^*_-(\omega_1,\omega_2) \\
&+ 2i\,\frac{k_1u^2(\omega_1)}{4\pi} M_-(\omega_1)\,\frac{k_2u^2(\omega_2)}{4\pi} M_+(\omega_2)P^*_-(\omega_1,\omega_2) \\
&+ \frac{1}{16\pi^3}\int_\mu^{\omega-\mu} d\omega'\,k'k'_-u^2(\omega')u^2(\omega'_-) \\
&\times e^{2i\delta_+(\omega')} R^*_-(\omega',\omega_2,\omega_1)P^*_-(\omega'_-,\omega'), \qquad (63)
\end{aligned}$$

where $\omega'_- = \omega - \omega'$, and in this and following equations $\omega = \omega_1 + \omega_2$. From Eqs. (40) and (55), the integral in Eq. (63) may be written as

$$\frac{1}{16\pi^3}\int_\mu^{\omega-\mu} d\omega'\, k'k'_-u^2(\omega')u^2(\omega'_-) \times e^{2i\delta_+(\omega')}R^*_-(\omega', \omega_2, \omega_1)P^*_-(\omega'_-, \omega') = \frac{g^3}{16\pi^3\omega_1}\left|\frac{T_-(\omega) - M_+(\omega)}{\Delta_+(\omega)}\right|^2 \times \frac{\Delta_-(\omega_1 + i\epsilon)\,\Delta_+(\omega_2 + i\epsilon)}{\Delta_+(\omega + i\epsilon)} \times \int_\mu^{\omega-\mu} \frac{d\omega'\, k'k'_-u^2(\omega')u^2(\omega'_-)}{(\omega_-)^2} \times |\Delta_-(\omega'_- + i\epsilon)\,\Delta_+(\omega' + i\epsilon)|^2. \quad (64)$$

On the other hand, we note from Eq. (45) that

$$\text{Im}\, T_-(\omega) = \frac{ku^2(\omega)}{4\pi}|T_-(\omega)|^2 + \frac{g^2}{16\pi^3}\left|\frac{T_-(\omega) - M_+(\omega)}{\Delta_+(\omega)}\right|^2 \times \int_\mu^{\omega-\mu} \frac{d\omega'\, k'k'_-u^2(\omega')u^2(\omega'_-)}{(\omega_-)^2} \times |\Delta_-(\omega'_- + i\epsilon)\,\Delta_+(\omega' + i\epsilon)|^2. \quad (65)$$

Equations (63), (64), (65), (12), and (40) now yield

$$\text{Im}\, P_-(\omega_1, \omega_2) = \frac{g}{\omega_1}\Big\{\text{Im}\, T_-(\omega) \times \frac{\Delta_+(\omega_2 + i\epsilon)\Delta_-(\omega_1 + i\epsilon)}{\Delta_+(\omega + i\epsilon)} - \frac{ku^2(\omega)}{4\pi}T^*_-(\omega) \times M_+(\omega)\frac{\Delta_+(\omega_2 + i\epsilon)\Delta_-(\omega_1 + i\epsilon)}{\Delta_+(\omega + i\epsilon)} + [T^*_-(\omega) - M^*_+(\omega)]\frac{P\Delta_+(\omega_2)P\Delta_-(\omega_1)}{\Delta_+(\omega - i\epsilon)} \times [e^{-i\delta_+(\omega_2)}\sin\delta_-(\omega_1) + e^{-i\delta_-(\omega_1)} \times \sin\delta_+(\omega_2) + 2i\sin\delta_-(\omega_1)\sin\delta_+(\omega_2)]\Big\}. \quad (66)$$

After some rearrangement, Eq. (66) becomes

$$\text{Im}\, P_-(\omega_1, \omega_2) = \frac{g}{\omega_1}\Big\{\text{Re}\, T_-(\omega)\sin[\delta_-(\omega_1) + \delta_+(\omega_2) - \delta_+(\omega)] + \text{Im}\, T_-(\omega)\cos[\delta_-(\omega_1) + \delta_+(\omega_2) - \delta_+(\omega)] - M_+(\omega)e^{-i\delta_+(\omega)}\sin[\delta_-(\omega_1) + \delta_+(\omega_2)]\Big\}\frac{P\Delta_+(\omega_2)P\Delta_-(\omega_1)}{P\Delta_+(\omega)}. \quad (67)$$

Since Eq. (67) is just the relation implied by Eq. (40), the unitarity relation for P_- is verified. Similar demonstrations may be carried out for P_+, R_-, and R_+, and the scattering matrix we have constructed is therefore unitary.

VIII. CROSSING SYMMETRY

In the Introduction we mentioned that there is more than one set of approximate dynamical equations whose solution leads to a two-meson S matrix. One point at which this latitude manifests itself is in the crossing properties of the production, and the six-point amplitudes one obtains. While it is fundamental that the scattering amplitudes must be crossing symmetric in a two-meson solution, the crossing properties of the multiparticle amplitudes have no such status, and they depend on the dynamical equations chosen. For example, if the amplitude P_- were crossing symmetric, then by crossing the positive outgoing meson with the negative incoming meson one would find

$$P_-(-\omega_1, \omega_2) = P_-(\omega_1 - \omega_2, \omega_2), \quad (68)$$

a particularly simple crossing relation. We note from Eq. (40) that Eq. (68) is not satisfied by our solution. On the other hand, an examination of Eq. (17) shows that, if we had used that equation to find P_-, Eq. (68) would be satisfied. Of course, no one has shown how to solve the resulting system of equations, and there is no guarantee that crossing relations other than Eq. (68) would be satisfied. Equation (17) has the property that Eq. (68) in particular is satisfied, just as Eq. (26), which we used in our calculation, has the property that the final-state mesons are on the same footing. In the case of P_+, our choice ensures that Bose statistics are maintained in the two-meson state.

The violation of Eq. (68) by our solution can be understood easily. It is not a consequence of the substitution of M's for T's in Eq. (24), but rather it is due to the fact that in Eq. (24) the incoming meson has both two- and three-particle cuts while the final mesons have only two-particle cuts (and no crossed cuts). Moreover, this asymmetric treatment of incoming and outgoing mesons is not corrected by including any finite number of intermediate states in Eq. (21), since if the outgoing mesons have n-body cuts, the incoming meson has an $(n + 1)$-body cut. Eq. (68) becomes valid only for the exact solution of Eq. (21).

In conclusion, we note that the methods we have used in this paper are probably not widely ap-

plicable to the development of two-mesons solutions for other static models. Leaving aside production, even the development of one-meson solutions proceeds quite differently for the symmetric scalar[2] and neutral pseudoscalar[2] theories than for the neutral[1] and charged[1] scalar theories. When production is included, it seems to be crucial that each channel be coupled to a single three-body state, and this requirement is met only by the neutral and charged scalar theories. It therefore seems probable that only these theories can have two-meson solutions constructed by our method. In the case of the neutral theory, the Hamiltonian version is known to give no scattering unless unstable states of the source are included. In the dispersion theoretic treatment, this means that CDD poles, or the new poles of $f(\omega)$ [Eq. (46)], or subtractions in the production dispersion relations are required. Such a solution would look somewhat different from the one we have presented and would arise naturally in the context of a discussion of the most general two-meson solution of the charged scalar theory.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 8 AUGUST 1966

A Class of Representations of the Generalized Bondi–Metzner Group*

VITTORIO CANTONI†
Department of Mathematics, King's College, London, England
(Received 19 October 1965)

A procedure is given for the construction of a faithful linear representation of the generalized Bondi–Metzner group from each faithful linear representation of the inhomogeneous orthochronous Lorentz group. Unitary representations can be obtained in this way.

I. INTRODUCTION

THE structure and meaning of the generalized Bondi–Metzner group (GBM group)[1-3] have been extensively investigated and discussed by Sachs.[4] In particular, its possible relevance to microphysics has been suggested,[4-11] and a Hermitian representation of the GBM Lie algebra has been analysed in this connection.[4]

In this paper it is shown that from any faithful linear representation of the inhomogeneous orthochronous Lorentz group a faithful linear representation of the GBM group can be constructed. First it is pointed out that, in a sufficiently small neighborhood of any given "ray direction," a GBM transformation is asymptotically equivalent, in a sense which is made precise, to a uniquely determined Lorentz transformation. It is then possible to construct a representation of the GBM group by forming the infinite semidirect product of any given representation of the Lorentz group with itself. If one starts from a unitary representation of the Lorentz group, an inner product can be defined in the new representation space such that the corresponding representation of the GBM group is unitary.

II. GBM TRANSFORMATIONS

Consider a normal hyperbolic Riemannian manifold assumed to admit a global coordinate system $(u, r, \theta, \phi) \equiv (x^0, x^1, x^2, x^3)$ in which the metric takes the form[3,4]

$$ds^2 = e^{2\beta}Vr^{-1}\,du^2 - 2e^{2\beta}\,du\,dr + r^2h_{AB}(dx^A - U^A\,du)(dx^B - U^B\,du) \qquad (1)$$

$(A, B = 2, 3)$, with

$$V = -r + 2M(u, \theta, \phi) + O(r^{-1}) = -C(u, \theta, \phi)C^* + O(r^{-4}),$$

$$h_{AB}\,dx^A\,dx^B = d\theta^2 + \sin^2\theta\,d\phi^2 + O(r^{-1}),$$

$$U^A = O(r^{-2})$$

* This work was sponsored in part by the Aerospace Research Laboratories, OAR, under Contract AF 61(052)-877, through the European Office of Aerospace Research, U. S. Air Force.

† N.A.T.O. research scholar, on leave from Gruppo di ricerca No. 36, C.N.R., Rome.

[1] H. Bondi, M. G. J. van der Burg, and A. W. K. Metzner, Proc. Roy. Soc. (London) **A269**, 21 (1962).

[2] E. T. Newman and T. W. J. Unti, J. Math. Phys. **3**, 891 (1962).

[3] R. K. Sachs, Proc. Roy. Soc. (London) **A270**, 103 (1962).

[4] R. K. Sachs, Phys. Rev. **128**, 2851 (1962).

[5] R. K. Sachs, in *Relativity, Groups and Topology* (the 1963 Les Houches lectures) (Gordon and Breach Science Publishers, New York, 1964.)

[6] R. Penrose, in *Relativity, Groups and Topology* (the 1963 Les Hauches lectures) (Gordon and Breach Science Publishers, New York, 1964).

[7] R. Penrose, Phys. Rev. Letters **10**, 66 (1963).

[8] R. Penrose, Proc. Roy. Soc. (London) **A284**, 159 (1965).

[9] E. T. Newman, Nature **206**, 811 (1965).

[10] A. Komar, Proceedings of the International Conference on General Relativity and Gravitation (London, 1965).

[11] A. Komar, Phys. Rev. Letters **15**, 76 (1965).

plicable to the development of two-mesons solutions for other static models. Leaving aside production, even the development of one-meson solutions proceeds quite differently for the symmetric scalar[2] and neutral pseudoscalar[2] theories than for the neutral[1] and charged[1] scalar theories. When production is included, it seems to be crucial that each channel be coupled to a single three-body state, and this requirement is met only by the neutral and charged scalar theories. It therefore seems probable that only these theories can have two-meson solutions constructed by our method. In the case of the neutral theory, the Hamiltonian version is known to give no scattering unless unstable states of the source are included. In the dispersion theoretic treatment, this means that CDD poles, or the new poles of $f(\omega)$ [Eq. (46)], or subtractions in the production dispersion relations are required. Such a solution would look somewhat different from the one we have presented and would arise naturally in the context of a discussion of the most general two-meson solution of the charged scalar theory.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 8 AUGUST 1966

A Class of Representations of the Generalized Bondi–Metzner Group*

VITTORIO CANTONI†
Department of Mathematics, King's College, London, England
(Received 19 October 1965)

A procedure is given for the construction of a faithful linear representation of the generalized Bondi–Metzner group from each faithful linear representation of the inhomogeneous orthochronous Lorentz group. Unitary representations can be obtained in this way.

I. INTRODUCTION

THE structure and meaning of the generalized Bondi–Metzner group (GBM group)[1-3] have been extensively investigated and discussed by Sachs.[4] In particular, its possible relevance to microphysics has been suggested,[4-11] and a Hermitian representation of the GBM Lie algebra has been analysed in this connection.[4]

In this paper it is shown that from any faithful linear representation of the inhomogeneous orthochronous Lorentz group a faithful linear representation of the GBM group can be constructed. First it is pointed out that, in a sufficiently small neighborhood of any given "ray direction," a GBM transformation is asymptotically equivalent, in a sense which is made precise, to a uniquely determined Lorentz transformation. It is then possible to construct a representation of the GBM group by forming the infinite semidirect product of any given representation of the Lorentz group with itself. If one starts from a unitary representation of the Lorentz group, an inner product can be defined in the new representation space such that the corresponding representation of the GBM group is unitary.

II. GBM TRANSFORMATIONS

Consider a normal hyperbolic Riemannian manifold assumed to admit a global coordinate system $(u, r, \theta, \phi) \equiv (x^0, x^1, x^2, x^3)$ in which the metric takes the form[3,4]

$$ds^2 = e^{2\beta}Vr^{-1}\,du^2 - 2e^{2\beta}\,du\,dr + r^2h_{AB}(dx^A - U^A\,du)(dx^B - U^B\,du) \qquad (1)$$

$(A, B = 2, 3)$, with

$$V = -r + 2M(u, \theta, \phi) + O(r^{-1}) = -C(u, \theta, \phi)C^* + O(r^{-4}),$$

$$h_{AB}\,dx^A\,dx^B = d\theta^2 + \sin^2\theta\,d\phi^2 + O(r^{-1}),$$

$$U^A = O(r^{-2})$$

* This work was sponsored in part by the Aerospace Research Laboratories, OAR, under Contract AF 61(052)-877, through the European Office of Aerospace Research, U. S. Air Force.

† N.A.T.O. research scholar, on leave from Gruppo di ricerca No. 36, C.N.R., Rome.

[1] H. Bondi, M. G. J. van der Burg, and A. W. K. Metzner, Proc. Roy. Soc. (London) **A269**, 21 (1962).

[2] E. T. Newman and T. W. J. Unti, J. Math. Phys. **3**, 891 (1962).

[3] R. K. Sachs, Proc. Roy. Soc. (London) **A270**, 103 (1962).

[4] R. K. Sachs, Phys. Rev. **128**, 2851 (1962).

[5] R. K. Sachs, in *Relativity, Groups and Topology* (the 1963 Les Houches lectures) (Gordon and Breach Science Publishers, New York, 1964.)

[6] R. Penrose, in *Relativity, Groups and Topology* (the 1963 Les Hauches lectures) (Gordon and Breach Science Publishers, New York, 1964).

[7] R. Penrose, Phys. Rev. Letters **10**, 66 (1963).

[8] R. Penrose, Proc. Roy. Soc. (London) **A284**, 159 (1965).

[9] E. T. Newman, Nature **206**, 811 (1965).

[10] A. Komar, Proceedings of the International Conference on General Relativity and Gravitation (London, 1965).

[11] A. Komar, Phys. Rev. Letters **15**, 76 (1965).

$(-\infty < u < +\infty;\ r_0 \leq r < +\infty;\ 0 \leq \theta \leq \pi;\ 0 \leq \phi \leq 2\pi)$. To each set of values for θ, ϕ, and u, there corresponds a null geodesic (or "ray") lying in the null hypersurface $u = \text{const}$.

GBM transformations preserve, by definition, the above form of the metric and the character of the coordinate system; they are given by

$$\begin{aligned} u' &= K^{-1}(\theta, \phi)[u + \alpha(\theta, \phi)] + O(r^{-1}), \\ r' &= K(\theta, \phi)r + J(u, \theta, \phi) + O(r^{-1}), \\ \theta' &= G^2(\theta, \phi) + O(r^{-1}), \\ \phi' &= G^3(\theta, \phi) + O(r^{-1}), \end{aligned} \tag{2}$$

where $\alpha(\theta, \phi)$ is an arbitrary twice-differentiable continuous function on the surface of the unit sphere, while G^2 and G^3 represent a conformal mapping of the surface of the unit sphere onto itself; α, G^2, and G^3 determine the transformation completely; in particular, they determine the functions K and J. One needs here the explicit expression for J only in the special cases where $G^2 = \theta$, $G^3 = \phi$, which is

$$J = \frac{1}{2}\left[\left(\frac{\partial^2\alpha}{\partial\theta^2}\right) + \left(\frac{1}{\sin^2\theta}\right)\left(\frac{\partial^2\alpha}{\partial\phi^2}\right) + \left(\frac{\partial\alpha}{\partial\theta}\right)\cot\theta\right]. \tag{3}$$

In the general case, the expression for K is

$$K(\theta, \phi) = (\sin\theta)^{\frac{1}{2}}(\sin G^2)^{-\frac{1}{2}}[\partial(G^2, G^3)/\partial(\theta, \phi)]^{-\frac{1}{2}}. \tag{3'}$$

The subgroup of the GBM group obtained by putting $\alpha = 0$ is isomorphic with the homogeneous orthochronous Lorentz group. The subgroup obtained by restricting α to the form

$$\alpha(\theta, \phi) = a^1 \sin\theta\cos\phi + a^2 \sin\theta\sin\phi + a^3\cos\theta - a^0$$

(a^i constants) is isomorphic with the inhomogeneous orthochronous Lorentz group. The subgroup of transformations with $G^2 = \theta$, $G^3 = \phi$ is the "supertranslation" subgroup.

The generic GBM transformation (2), characterized by α, G^2, and G^3, can be regarded as the product of the supertranslation characterized by the function α, followed by the homogeneous Lorentz transformation Λ characterized by G^2 and G^3. It is denoted by (Λ, α). Thus symbols of the type $(1, \alpha)$ and $(\Lambda, 0)$ represent supertranslations and homogeneous Lorentz transformations, respectively, and $(\Lambda, \alpha) = (\Lambda, 0)(1, \alpha)$.

III. ASYMPTOTIC TANGENCY

Let $(x^i) \equiv (x^0, x^1, x^2, x^3) \equiv (u, r, \theta, \phi)$ by any coordinate system of the type introduced above.

By "ray direction" $(\bar\theta, \bar\phi)$ is meant the set of all rays labeled by the same couple of values $\bar\theta \equiv \bar x^2$ and $\bar\phi \equiv \bar x^3$ of the angular coordinates x^2 and x^3, and any possible value of the coordinate x^0.

Let t' and t'' be two GBM transformations, so that

$$(x^i) \xrightarrow{t'} (x'^i), \qquad (x^i) \xrightarrow{t''} (x''^i) \qquad (i = 0, 1, 2, 3).$$

By definition t' and t'' are called "asymptotically tangent in the ray direction $(\bar\theta, \bar\phi)$" [briefly AT $(\bar\theta, \bar\phi)$] if, for any fixed value of x^0,

$$\begin{aligned} &\lim_{x^1\to\infty}[x''^i(x^0, x^1, \bar x^2, \bar x^3) - x'^i(x^0, x^1, \bar x^2, \bar x^3)] = 0, \\ &\lim_{x^1\to\infty}[x''^0(x^0, x^1, \bar x^2 + \Delta x^2, \bar x^3 + \Delta x^3) \\ &\quad - x'^0(x^0, x^1, \bar x^2 + \Delta x^2, \bar x^3 + \Delta x^3)] = O(\Delta x^B \cdot \Delta x^C), \\ &\lim_{x^1\to\infty}[x''^A(x^0, x^1, \bar x^2 + \Delta x^2, \bar x^3 + \Delta x^3) \\ &\quad - x'^A(x^0, x^1, \bar x^2 + \Delta x^2, \bar x^3 + \Delta x^3)] = O(\Delta x^B \cdot \Delta x^C) \\ &\qquad (i = 0, 1, 2, 3; \quad A \cdot B \cdot C = 2, 3). \end{aligned} \tag{4}$$

(See Ref. 12.)

The following remarks are useful:

(1) Two supertranslations $(1, \alpha_1)$ and $(1, \alpha_2)$ are AT $(\bar\theta, \bar\phi)$ if and only if

$$\begin{aligned} &\alpha_1(\bar\theta, \bar\phi) = \alpha_2(\bar\theta, \bar\phi), \\ &(\partial\alpha_1/\partial\theta)_{\bar\theta,\bar\phi} = (\partial\alpha_2/\partial\theta)_{\bar\theta,\bar\phi}, \\ &\left(\frac{\partial^2\alpha_1}{\partial\theta^2} + \frac{1}{\sin^2\theta}\frac{\partial^2\alpha_1}{\partial\phi^2} + \frac{\partial\alpha_1}{\partial\theta}\cot\theta\right)_{\bar\theta,\bar\phi} \\ &\quad = \left(\frac{\partial^2\alpha_2}{\partial\theta^2} + \frac{1}{\sin^2\theta}\frac{\partial^2\alpha_2}{\partial\phi^2} + \frac{\partial\alpha_2}{\partial\theta}\cot\theta\right)_{\bar\theta,\bar\phi}. \end{aligned} \tag{5}$$

This follows from (2) and (3) as an immediate consequence of the above definition.

(2) If two homogeneous Lorentz transformations $(\Lambda_1, 0)$ and $(\Lambda_2, 0)$ are AT $(\bar\theta, \bar\phi)$, they necessarily coincide. This may be proved by expressing G^2 and G^3, which represent a conformal transformation of the surface of the unit sphere, explicitly in terms of six parameters: denoting by ξ_1 and ξ_2 the sets of parameters which correspond to Λ_1 and Λ_2, respectively, and noting that conditions (4) imply in this case,

$$\begin{aligned} &G^A(\bar\theta, \bar\phi i\xi_1) = G^A(\bar\theta, \bar\phi i\xi_2), \\ &[\partial G^A(\theta, \phi i\xi_1)/\partial x^B]_{\bar\theta,\bar\phi} = [\partial G^A(\theta, \phi i\xi_2)/\partial x^B]_{\bar\theta,\bar\phi}, \\ &\qquad (A, B = 2, 3), \end{aligned}$$

one obtains six relations among the twelve quantities ξ_1 and ξ_2. These relations are satisfied if and only if ξ_1 and ξ_2 correspond to the same transformation.

(3) More generally, two GBM transformations

[12] $0(\Delta x^B \cdot \Delta x^C)$ is to be interpreted as the order of magnitude of the largest of the products $\Delta\theta\cdot\Delta\theta$, $\Delta\theta\cdot\Delta\phi$, $\Delta\phi\cdot\Delta\phi$.

(Λ_1, α_1) and (Λ_2, α_2) are AT$(\bar\theta, \bar\phi)$ if and only if $\Lambda_1 = \Lambda_2$ and Eqs. (5) hold. It is convenient at this stage to introduce the symbols $\Lambda\bar\theta$ and $\Lambda\bar\phi$ to denote $G^2(\bar\theta, \bar\phi)$ and $G^3(\bar\theta, \bar\phi)$, respectively, whenever G^2 and G^3 correspond to the homogeneous Lorentz transformation Λ.

The following statement is easily seen to follow from conditions (4):

(4) If (Λ, α_1) and (Λ, α_2) are AT $(\bar\theta, \bar\phi)$, and if (Σ, β_1) and (Σ, β_2) are AT $(\Lambda\bar\theta, \Lambda\bar\phi)$, then the products $(\Sigma, \beta_1)(\Lambda, \alpha_1)$ and $(\Sigma, \beta_2)(\Lambda, \alpha_2)$ are AT $(\bar\theta, \bar\phi)$.

IV. TRANSLATIONS ASYMPTOTICALLY TANGENT TO A SUPERTRANSLATION

Let $(1, \alpha)$ be any given supertranslation. Corresponding to each ray direction $(\bar\theta, \bar\phi)$, there exists a uniquely determined translation $(1, a_{\bar\theta\bar\phi})$ which is AT $(\bar\theta, \bar\phi)$ to $(1, \alpha)$; $[a_{\bar\theta\bar\phi}(\theta, \phi) \equiv a^1_{\bar\theta\bar\phi} \sin\theta\cos\phi + a^2_{\bar\theta\bar\phi} \sin\theta\sin\phi + a^3_{\bar\theta\bar\phi}\cos\theta - a^0_{\bar\theta\bar\phi}$; $a^i_{\bar\theta\bar\phi}$ constants]. In fact, writing (5) with $\alpha_1 = \alpha$, $\alpha_2 = a_{\bar\theta\bar\phi}$ and solving with respect to the coefficients $a^i_{\bar\theta\bar\phi}$, one gets

$$
\begin{aligned}
a^0 &= \alpha + \frac{1}{2}\left(\frac{\partial^2\alpha}{\partial\theta^2} + \frac{1}{\sin^2\theta}\frac{\partial^2\alpha}{\partial\phi^2} + \frac{\partial\alpha}{\partial\theta}\cot\bar\theta\right),\\
a^1 &= (a^0 - \alpha)\sin\bar\theta\cos\bar\phi\\
&\qquad + \frac{\partial\alpha}{\partial\phi}\frac{\sin\bar\phi}{\sin\bar\theta} - \frac{\partial\alpha}{\partial\theta}\cos\bar\theta\cos\bar\phi,\\
a^2 &= (a^0 - \alpha)\sin\bar\theta\sin\bar\phi\\
&\qquad - \frac{\partial\alpha}{\partial\phi}\frac{\cos\bar\phi}{\sin\bar\theta} - \frac{\partial\alpha}{\partial\theta}\cos\bar\theta\sin\bar\phi,\\
a^3 &= (a^0 - \alpha)\cos\bar\theta + (\partial\alpha/\partial\theta)\sin\bar\theta,
\end{aligned}
\tag{6}
$$

where α and its derivatives are evaluated at the point $(\bar\theta, \bar\phi)$.

Thus, to every supertranslation may be associated a set of translations, each corresponding to a ray direction $(\bar\theta, \bar\phi)$. Conversely, the first equation of (5) shows that a supertranslation is unambiguously determined by the set of its asymptotically tangent translations.

It can be seen, from Eqs. (6), that when $(1, \alpha)$ is itself a translation it coincides with all its AT translations.

If two inhomogeneous Lorentz transformations (Λ_1, a_1) and (Λ_2, a_2) are AT in one ray direction $(\bar\theta, \bar\phi)$, they necessarily coincide, for remark (3) in Sec. III implies in this case $\Lambda_1 = \Lambda_2$ and $a_1 = a_2$.

V. CONSTRUCTION OF THE REPRESENTATION

Consider now any faithful linear representation[13] of the inhomogeneous orthochronous Lorentz group L. $T(l)$ denotes the operator representing the generic element $l \in L$; $\mathfrak{I}$ and $\mathfrak{K}$ denote the space of the operators T and the representation space, respectively.

Let $g \equiv (\Lambda, \alpha) = (\Lambda, 0)(1, \alpha)$ be a GBM transformation. To $(\Lambda, 0)$ there corresponds an element of the homogeneous Lorentz group and, therefore an element $T(\Lambda) \in \mathfrak{I}$. To $(1, \alpha)$ and to every ray direction (θ, ϕ), there corresponds a definite AT translation $(a_{\theta\phi}, 1)$ represented by the operator $T(a_{\theta\phi}) \in \mathfrak{I}$ (bars on θ and ϕ have been dropped).

With the GBM transformation g, one can associate the mapping $\tilde T_g$ of the surface of the unit sphere into the space $\mathfrak{I}$ defined by

$$
(\theta, \phi) \xrightarrow{\tilde T_g} T(\Lambda)T(a_{\theta\phi}) \equiv \tilde T_g(\theta, \phi)
\qquad (0 \le \theta \le \pi;\ 0 \le \phi \le 2\pi). \tag{7}
$$

Let $h \equiv (\Sigma, \beta)$ be another GBM transformation and $\tilde T_h$ the corresponding mapping according to the above definition. One can define the product $\tilde U = \tilde T_h \tilde T_g$ of the mappings $\tilde T_g$ and $\tilde T_h$ by

$$
\begin{aligned}
\tilde U(\theta, \phi) &\overset{\text{def}}{=} T(\Sigma)T(b_{\Lambda\theta, \Lambda\phi})T(\Lambda)T(a_{\theta,\phi})\\
&= \tilde T_h(\Lambda\theta, \Lambda\phi)\tilde T_g(\theta, \phi),
\end{aligned}
\tag{8}
$$

where $b_{\theta,\phi}$ is the translation AT (θ, ϕ) to $(1, \beta)$.

It is now shown that *the mappings $\tilde T$ constitute a representation*, that is to say,

$$\tilde T_{hg} = \tilde T_h \tilde T_g.$$

In fact, for any value of (θ, ϕ), the Lorentz transformation

$$(\Sigma, 0)(1, b_{\Lambda\theta, \Lambda\phi})(\Lambda, 0)(1, a_{\theta,\phi}) \tag{9}$$

is AT (θ, ϕ) to hg [Sec. III, remark (4)]. On the other hand, hg can be written in the form $hg = (\Sigma\Lambda, 0)(1, \gamma)$, and denoting by $(1, c_{\theta,\phi})$ the translation AT (θ, ϕ) to $(1, \gamma)$, the Lorentz transformation

$$(\Sigma\Lambda, 0)(1, c_{\theta,\phi}) \tag{10}$$

is also AT (θ, ϕ) to hg. From remark (3) we see that (9) and (10), being AT (θ, ϕ) to the same GBM transformation hg, are also AT (θ, ϕ) to each other. Being both Lorentz transformations, they coincide:

$$(\Sigma, 0)(1, b_{\Lambda\theta, \Lambda\phi})(\Lambda, 0)(1, a_{\theta,\phi}) = (\Sigma\Lambda, 0)(1, C_{\theta,\phi}) \tag{11}$$

(see the end of Sec. IV). When θ and ϕ vary, the left-hand member of this equation defines $\tilde T_h \tilde T_g$ according to (8), while the right-hand member defines $\tilde T_{hg}$ according to (7). Thus the assertion is proved.

It now remains only to show that the mappings $\tilde T$ can be regarded as linear operators on a linear space $\tilde{\mathfrak{K}}$.

[13] For simplicity we assume the representation to be single-valued.

VI. CONSTRUCTION OF THE REPRESENTATION SPACE

Let $\tilde{\mathcal{H}}$ be the space of all continuous mappings of the surface of the unit sphere into the space $\mathcal{H}$

$$\phi : (\theta, \phi) \to \tilde{\Phi}(\theta, \phi); \qquad \tilde{\Phi}(\theta, \phi) \in \mathcal{H}, \qquad \tilde{\Phi} \in \tilde{\mathcal{H}},$$
$$0 \leq \theta \leq \pi, \qquad 0 \leq \phi \leq 2\pi.$$

$\tilde{\mathcal{H}}$ is a linear space with the following definition of sum and of composition with scalars:

$$a\tilde{\Phi} + b\tilde{\Phi} : (\theta, \phi) \to a\tilde{\Phi}(\theta, \phi) + b\tilde{\Phi}(\theta, \phi), \quad \tilde{\Phi}, \tilde{\Phi} \in \tilde{\mathcal{H}};$$
$$\tilde{\Phi}(\theta, \phi), \quad \tilde{\Phi}(\theta, \phi) \in \mathcal{H}; \quad a, b \text{ complex numbers.}$$

If $\tilde{T}_g \in \tilde{\mathfrak{I}}$ is the mapping associated with the GBM transformation $g \equiv (\Lambda, \alpha)$, the result $\tilde{\Phi}' = \tilde{T}_g\tilde{\Phi}$ of the operation of $\tilde{T}_g$ on $\tilde{\Phi} \in \tilde{\mathcal{H}}$ may be defined as follows:

$$\tilde{\Phi}'(\theta, \phi) = K_\Lambda(\Lambda^{-1}\theta, \Lambda^{-1}\phi)\tilde{T}_g \times (\Lambda^{-1}\theta, \Lambda^{-1}\phi)\tilde{\Phi}(\Lambda^{-1}\theta, \Lambda^{-1}\phi), \qquad (12)$$

where $K_\Lambda(\Lambda^{-1}\theta, \Lambda^{-1}\phi) \equiv (\sin \Lambda^{-1}\theta)^{\frac{1}{2}} (\sin \theta)^{-\frac{1}{2}} [\partial(\Lambda^{-1}\theta, \Lambda^{-1}\phi)/\partial(\theta, \phi)]^{\frac{1}{2}}$ [see (3′)].

It is easy to verify that the operation just defined is linear and that the result of the action of the product of two operators is the same as the result of the successive application of the two operators.

Assuming now that the operators $T \in \mathfrak{I}$ are unitary with respect to a scalar product defined in $\mathcal{H}$, one can define a scalar product in $\tilde{\mathcal{H}}$ with respect to which the operators $\tilde{T} \in \tilde{\mathfrak{I}}$ are also unitary, in the following way:

$$[\tilde{\Psi}, \tilde{\Phi}] = \int_0^\pi \int_0^{2\pi} \Big(\tilde{\Psi}(\theta, \phi), \tilde{\Phi}(\theta, \phi)\Big) \sin \theta \, d\theta \, d\phi, \qquad (13)$$

where $(\ ,\ )$ denotes the scalar product in $\mathcal{H}$, while $[\ ,\ ]$ is the scalar product in $\tilde{\mathcal{H}}$. The operators $\tilde{T}$ are unitary because, taking (12) and (13) into account, and the fact that the operators T are unitary in $\mathcal{H}$, one has:

$$[\tilde{T}\tilde{\Psi}, \tilde{T}\tilde{\Phi}] = \int_0^\pi \int_0^{2\pi} K_\Lambda^2(\Lambda^{-1}\theta, \Lambda^{-1}\phi)\Big(\tilde{T}(\Lambda^{-1}\theta, \Lambda^{-1}\phi) \times \tilde{\Psi}(\Lambda^{-1}\theta, \Lambda^{-1}\phi), \tilde{T}(\Lambda^{-1}\theta, \Lambda^{-1}\phi) \times \tilde{\Phi}(\Lambda^{-1}\theta, \Lambda^{-1}\phi)\Big) \sin \theta \, d\theta \, d\phi$$

$$= \int_0^\pi \int_0^{2\pi} K_\Lambda^2(\Lambda^{-1}\theta, \Lambda^{-1}\phi)\Big(\tilde{\Psi}(\Lambda^{-1}\theta, \Lambda^{-1}\phi), \tilde{\Phi}(\Lambda^{-1}\theta, \Lambda^{-1}\phi)\Big) \sin \theta \, d\theta \, d\phi$$

$$= \int_0^\pi \int_0^{2\pi} \Big(\tilde{\Psi}(\Lambda^{-1}\theta, \Lambda^{-1}\phi), \tilde{\Phi}(\Lambda^{-1}\theta, \Lambda^{-1}\phi)\Big) \sin \Lambda^{-1}\theta \, d\Lambda^{-1}\theta \, d\Lambda^{-1}\phi = [\tilde{\Psi}, \tilde{\Phi}].$$

The faithfulness of the representations obtained is a consequence of the fact that a GBM transformation is unambiguously determined by the set of its asymptotically tangent Lorentz transformations.

VII. REMARK ON THE REST-MASS OPERATOR

The inhomogeneous orthochronous Lorentz group L being a subgroup of the GBM group, every representation of the GBM group induces a representation of L. No attempt is made here to determine which representations of L are induced by the GBM representations obtained above. Note only that whenever the representation of L on which the construction is based is irreducible, so that the rest-mass operator is a multiple of the identity with eigenvalue m^2, one obtains an induced representation of L in which the rest-mass operator is also a multiple of the identity with the same eigenvalue m^2. In fact, let $A_k(t) \equiv 1 + itP_k + O(t^2)$, $(k = 0, 1, 2, 3; -\infty < t < +\infty)$, represent the one-parameter subgroup of translations parallel to the kth axis, in the original representation of L, so that $P_k \in \mathfrak{I}$ corresponds to the kth component of the linear momentum. From definition (7) it follows that the same one-parameter subgroup is represented, in the GBM representation, by the operator $\tilde{A}_k \in \tilde{\mathfrak{I}}$ such that

$$\tilde{A}_k(\theta, \phi; t) = A_k(t) = 1 + itP_k + O(t^2),$$

and the operator $\tilde{P}_k$ corresponding to the kth component of the linear momentum in the induced representation of L is such that

$$\tilde{P}_k(\theta, \phi) = P_k.$$

Thus, for the rest-mass operator in this representation, one has

$$\tilde{M}^2(\theta, \phi) = M^2.$$

From (12) one sees that if, by assumption, every element of $\mathcal{H}$ is an eigenvector of M^2 with eigenvalue m^2, then every element of $\tilde{\mathcal{H}}$ is an eigenvector of $\tilde{M}^2$ with the same eigenvalue m^2.

ACKNOWLEDGMENTS

The author is deeply indebted to Dr. E. T. Newman, who suggested this problem, for his constant encouragement and enlightning discussions, to Dr. R. Penrose for several valuable discussions, and to Dr. F. A. E. Pirani for his careful reading of the manuscript and helpful suggestions.

Role of Causality in Quantum Field Theory and the Dynamical Postulate*

T. W. CHEN, F. ROHRLICH,† AND M. WILNER
Physics Department, Syracuse University, Syracuse, New York
(Received 26 October 1965)

It is shown that Bogoliubov causality together with the usual assumptions of quantum field theory suffice to determine the off-mass-shell behavior of the S-operator as that resulting from the φ-product. The same assumptions lead to an equation for the current which replaces the field equation for the interpolating field. The specification of the interaction term in this equation corresponds to the specification of an interaction Hamiltonian in the usual formalism. This interaction term is an operator distribution whose support is a single point and which is otherwise severly restricted.

I. INTRODUCTION AND SUMMARY

THE new, asymptotic formulation of quantum field theory has suffered since its inception from two basic difficulties. The first was the problem of "going off the mass shell"; it seemed to be completely arbitrary how this is to be done and, short of solving infinite simultaneous systems of equations in order to find a possible consistency criterion as suggested by Lehmann, Symanzik, and Zimmermann, an additional assumption seemed to be required. The second question was how the interaction is to be specified in a theory which is not (at least not explicitly) based on a Lagrangian or Hamiltonian formalism. The present paper suggests a solution to both of these problems.

Ever since the bold assumption was made by Pugh[1] that the continuation off the mass shell is to be accomplished in terms of the φ-product rather than any other product, and since his verification that this leads to the correct quantum electrodynamics, we have had the problem of justifying this procedure. It appears in both Pugh's original formulation as well as in his later operator formulation[2] that a new assumption is needed [Pugh's dynamical axiom, Eq. (5.6) below]. We show below that this is not the case, but that *causality*, in the form of Bogoliubov's condition on the current rather than local commutativity of the interpolating field (together with the other assumptions also made by Pugh such as strong unitarity and completeness of the asymptotic fields) suffices to single out the φ product from all others. No additional axiom is needed. Since the theory is formulated in terms of strong operator equations, the problem of "going off the mass shell" is then completely resolved.

While the general equations of the theory are thus determined, the question of how and where to insert the specific interaction is not thereby answered. To this end, it is shown that the axioms of the theory lead to an equation for the "current," defined by (2.3), which contains an undetermined operator distribution whose functional dependence is, however, limited. This operator distribution appears to be the natural input for the specific interaction. Thus, this equation plays the role that the field equation for the interpolating field used to play and it contains an interaction term. Equation (4.8) below is nonlinear and contains the causality condition within it. It therefore also implies certain analytic properties (which are, however, not pursued in the present paper).

The operator distribution whose choice specifies a particular interaction can be related to the interaction Hamiltonian of the usual formalism. This specification is therefore the *dynamical postulate* of the theory. The basic technical problem in the theory is now the solution of the equation for the current, for a given interaction. The S-operator is then easily obtained from the current.

The study outlined above is carried out in detail for the neutral scalar field. The generalization to other fields is fairly straightforward. The basic mathematical tool will be differentiation with respect to a free field.[3] Since the theory is symmetric in in-fields and out-fields we choose the in-fields and use the notation

$$\delta F/\delta x \equiv \delta F/\delta A_{\text{in}}(x), \tag{1.1}$$

where F is an arbitrary operator functional and $A_{\text{in}}(x)$ is the free field operator,

$$KA_{\text{in}}(x) \equiv (\Box - m^2)A_{\text{in}}(x) = 0.$$

* This work was supported by the National Science Foundation.

† Part of this work was carried out while one of the authors (F. R.) was a member of the Physics Division of the Aspen Institute for Humanistic Studies, Aspen, Colorado. The hospitality of the Institute is gratefully acknowledged.

[1] R. E. Pugh, Ann. Phys. (N. Y.) **23**, 335 (1963).

[2] R. E. Pugh, J. Math. Phys. **6**, 740 (1965).

[3] F. Rohrlich, J. Math. Phys. **5**, 324 (1964); F. Rohrlich and M. Wilner, *ibid.* **7**, 482 (1966).

In Sec. II we discuss various causality statements in terms of the derivatives (1.1). The desire of manifest covariance makes the use of invariant step functions a welcome tool (Sec. III). The equation for the current is then derived (Sec. IV) and the restrictions on the interaction term are obtained (Sec. V); in the course of this study Pugh's dynamical axiom is derived from causality. Finally, it is shown how the S operator follows easily from the current (Sec. VI) and how the interaction term in the fundamental current equation can be related to the Hamiltonian of the usual theory.

II. CAUSALITY

With the assumption of strong unitarity,[4]

$$S^* \overset{s}{=} S^{-1}, \tag{2.1}$$

we easily derive the relation

$$S^* \frac{\delta^2 S}{\delta x_1\, \delta x_2} = \frac{\delta}{\delta x_1}\left(S^* \frac{\delta S}{\delta x_2}\right) - \frac{\delta S^*}{\delta x_1} S S^* \frac{\delta S}{\delta x_2} = -i \frac{\delta j(x_2)}{\delta x_1} - j(x_1) j(x_2). \tag{2.2}$$

The current $j(x)$ is here defined by

$$j(x) \equiv iS^*(\delta S/\delta x) = -i(\delta S^*/\delta x)S. \tag{2.3}$$

Now, it can easily be proven[3] that the second derivative is symmetric,

$$\delta^2 S/\delta x_1\, \delta x_2 = \delta^2 S/\delta x_2\, \delta x_1. \tag{2.4}$$

Therefore, (2.2) can also be written as [note $j_i \equiv j(x_i)$]

$$S^* \frac{\delta^2 S}{\delta x_1\, \delta x_2} = -i \frac{\delta j_1}{\delta x_2} - j_2 j_1 = -\frac{i}{2}\left(\frac{\delta j_1}{\delta x_2} + \frac{\delta j_2}{\delta x_1}\right) - \tfrac{1}{2}\{j_1, j_2\}. \tag{2.2'}$$

We also find

$$i(\delta j_1/\delta x_2) - i(\delta j_2/\delta x_1) = {}_i[j_1, j_2]. \tag{2.5}$$

Since the current operator $j(x)$ is an observable, the causality statement usually made is

$$[j_1, j_2] = 0 \qquad (x_1 \cong x_2), \tag{2.6}$$

where $x_1 \sim x_2$ means that the two events x_1 and x_2 are spacelike relative to each other. This causality statement follows from Lorentz invariance alone, if one assumes the commutator to be a multiple of the unit operator (c number) for $x_1 \cong x_2$. It is referred to as *local current causality*.

Equation (2.5) implies that local current causality can also be written in the form

$$\delta j_1/\delta x_2 = \delta j_2/\delta x_1 \qquad (x_1 \cong x_2). \tag{2.7}$$

The equivalence of the two statements (2.6) and (2.7) is assured by strong unitarity, (2.1), and the "integrability condition" (2.4).

A much stronger causality requirement is the Bogoliubov causality condition[5] as generalized to operator derivatives,[6]

$$\delta j_1/\delta x_2 = 0 \qquad (x_1 \lesssim x_2). \tag{2.8}$$

The notation $x_1 < x_2$ signifies that $x_1 - x_2$ is timelike or null and $x_1^0 < x_2^0$. The point $x_1 = x_2$ is not included in (2.8).

The statement (2.8) is referred to as *strong Bogoliubov causality*. It implies (2.7), i.e., local current causality.

There exists a weaker statement than (2.8) which also implies local current causality, viz.,

$$\delta j_1/\delta x_2 = 0 \qquad (x_1 < x_2), \tag{2.9a}$$

$$\delta j_1/\delta x_2 = \delta j_2/\delta x_1 \qquad (x_1 \cong x_2). \tag{2.9b}$$

These two conditions together are referred to as *weak* Bogoliubov causality. Obviously, (2.9b) and (2.7) are identical, so that local current causality is obviously implied.

From Sec. V on we always use *strong* Bogoliubov causality and do not always say so explicitly.

III. INVARIANT STEP FUNCTION

The function $\tilde{\theta}(x)$ is defined as follows:

$$\tilde{\theta}(x) \equiv 1 \quad \text{for} \quad x^0 > 0, \quad x^2 \text{ timelike or null} \qquad \equiv 0 \quad \text{otherwise.} \tag{3.1}$$

This function is obviously Lorentz-invariant. It is also convenient to define

$$\tilde{\epsilon}(x) \equiv \tilde{\theta}(x) - \tilde{\theta}(-x), \tag{3.2}$$

which is also Lorentz-invariant and which vanishes everywhere outside the light cone as well as at $x = 0$.

We note that $\tilde{\theta}(x) + \tilde{\theta}(-x)$ vanishes in exactly the same domain as $\tilde{\epsilon}(x)$ and is consequently *not* equal to unity.

[4] Strong equalities ($\overset{s}{=}$) between operators imply that they continue to hold upon operator differentiation to arbitrary order. In the following, all equations are strong equations unless noted otherwise, so that we can omit the index s on the equality sign.

[5] N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Interscience Publishers, Inc., London, 1959).

[6] F. Rohrlich and J. C. Stoddart, J. Math. Phys. 6, 495 (1965). It is shown here that (2.8) implies local commutativity, $[A(x_1), A(x_2)] = 0$, $(x_1 \sim x_2)$, also for charged fields and higher spins.

These functions are very convenient when one states the causality conditions. The local current causality (2.6) can be written as [we put $\tilde{\theta}_{12} \equiv \tilde{\theta}(x_1 - x_2)$]

$$(1 - \tilde{\theta}_{12} - \tilde{\theta}_{21})[j_1, j_2] = 0. \tag{3.3}$$

With the assumption of Lorentz invariance, weak Bogoliubov causality can be stated as

$$\tilde{\theta}_{21}(\delta j_1/\delta x_2) = 0. \tag{3.4}$$

The strong Bogoliubov causality statement is then just the statement that

$$(1 - \tilde{\theta}_{21})(\delta j_2/\delta x_1) = 0, \quad \text{for} \quad x_1 \neq x_2. \tag{3.5}$$

Consider next the time-ordered product of the current at two points x_1 and x_2. In terms of the usual step functions $\theta(x)$ and $\epsilon(x)$ we have

$$\begin{aligned} T(j_1 j_2) &\equiv \theta_{12} j_1 j_2 + \theta_{21} j_2 j_1 \\ &= \tfrac{1}{2}\{j_1, j_2\} + \tfrac{1}{2}\epsilon_{12}[j_1, j_2]. \end{aligned} \tag{3.6}$$

Assuming that local current causality (2.6) holds, ϵ_{12} can be replaced by $\tilde{\epsilon}_{12}$ without changing the equation; this leads to the following alternative forms of the T product which are all manifestly covariant:

$$\begin{aligned} T(j_1 j_2) &= \tfrac{1}{2}\{j_1, j_2\} + \tfrac{1}{2}\tilde{\epsilon}_{12}[j_1, j_2] \\ &= \tfrac{1}{2}\{j_1 j_2(1 + \tilde{\theta}_{12} - \tilde{\theta}_{21}) \\ &\qquad + j_2 j_1(1 - \tilde{\theta}_{12} + \tilde{\theta}_{21})\} \\ &= \tilde{\theta}_{12}[j_1, j_2] + j_2 j_1 \\ &= -\tilde{\theta}_{21}[j_1, j_2] + j_1 j_2. \end{aligned} \tag{3.7}$$

IV. DYNAMICAL POSTULATE

The equations derived in the previous sections permit us to obtain the following strong equation for the second derivative of the S operator and the current T product,

$$\begin{aligned} S^* \frac{\delta^2 S}{\delta x_1\, \delta x_2} &+ T(j_1 j_2) \\ &= -i \frac{\delta j_2}{\delta x_1} - \tilde{\theta}_{21}[j_1, j_2] \\ &= -i \frac{\delta j_2}{\delta x_1} - \tilde{\theta}_{21}\left(i \frac{\delta j_1}{\delta x_2} - i \frac{\delta j_2}{\delta x_1}\right). \end{aligned} \tag{4.1}$$

Equations (2.2), (3.7), and (2.5) used here are based only on strong unitarity and Lorentz invariance.

The requirement of weak Bogoliubov causality, (3.4), reduces this relation to

$$\begin{aligned} S^*(\delta^2 S/\delta x_1\, \delta x_2) &+ T(j_1 j_2) \\ &= -i(1 - \tilde{\theta}_{21})(\delta j_2/\delta x_1). \end{aligned} \tag{4.2}$$

If we also require (3.5) (strong Bogoliubov causality), we find that the right side of (4.2) vanishes everywhere except possibly at the single point $x_1 = x_2$ which we denote by O_{12}. In that case,

$$\operatorname{supp}\left[S^* \frac{\delta^2 S}{\delta x_1\, \delta x_2} + T(j_1 j_2)\right] = \text{point } O_{12}, \tag{4.3}$$

i.e., the support of this operator-valued distribution is the single point O_{12}. We repeat that this is a consequence of Lorentz invariance, strong unitarity, and strong Bogoliubov causality only.

Let us denote this operator [right side of (4.2)], whose support is the point O_{12} only, $\beta(x_1, x_2) \equiv \beta_{12}$

$$\beta_{12} \equiv -(1 - \tilde{\theta}_{21}) i \frac{\delta j_2}{\delta x_1} = -(1 - \tilde{\theta}_{12}) i \frac{\delta j_1}{\delta x_2}. \tag{4.4}$$

We note that (4.3) demands that β be a symmetric function of its arguments. Equation (4.2) then becomes the following equation for the S operator in terms of β_{12}:

$$S^* \frac{\delta^2 S}{\delta x_1\, \delta x_2} = \beta_{12} + T\left(S^* \frac{\delta S}{\delta x_1} S^* \frac{\delta S}{\delta x_2}\right). \tag{4.5}$$

The operator (distribution) β_{12} is apparently expressible in terms of $\delta(x_1 - x_2)$ and its derivatives multiplied by an operator function of one of the two variables. Since (4.4) can also be written as

$$(1 - \tilde{\theta}_{12} - \tilde{\theta}_{21}) i(\delta j_2/\delta x_1) + \beta_{12} = 0, \tag{4.6}$$

this could be interpreted by saying that the current is unaffected by a change of the in-field at a spacelike point, but that it is affected at the *same* point corresponding to a direct interaction. We note that (4.6) is completely symmetric in x_1 and x_2, β being a symmetric function of its arguments. It is also symmetric with respect to in- and out-fields in the sense that derivatives with respect to the out-fields lead to an equation of exactly the same form as (4.6) with a function $\beta_{\text{out}}(x_1, x_2)$ which is of exactly the same nature.

If $\beta(x_1, x_2)$ is known, (4.5) is an equation for the S operator. We refer to this equation with known $\beta(x_1 x_2)$ as the *dynamical postulate*. Since the structure of the equation follows from strong unitarity and causality, the dynamical postulate consists only in the specification of $\beta(x_1 x_2)$.

We conclude this section by casting (4.5) into

another, equivalent form which is useful later. Using (2.2), (2.4), and (3.7), Eq. (4.5) becomes

$$-i\frac{\delta j_2}{\delta x_1} - i\frac{\delta j_1}{\delta x_2} - \{j_1, j_2\} = 2\beta_{12} - (\{j_1, j_2\} + \bar{\epsilon}_{12}[j_1, j_2])$$

or

$$i\frac{\delta j_2}{\delta x_1} + i\frac{\delta j_1}{\delta x_2} = -2\beta_{12} + \bar{\epsilon}_{12}[j_1, j_2]. \quad (4.7)$$

Because of (2.5) this can also be written as

$$i\,\delta j_1/\delta x_2 = -\beta_{12} + \bar{\theta}_{12}[j_1, j_2], \quad (4.8)$$

in which 1 and 2 can of course be interchanged. As a check one sees easily that this equation is equivalent to (4.4) provided strong unitarity (2.5) and weak causality (3.4) hold.

Equation (4.8) shows clearly that the dynamical postulate (the knowledge of β_{12}) gives one an equation for S (4.5) and an equation for j, and that these two equations are equivalent. Obviously, one can find j if S is known, but the converse is less obvious. We show in Sec. VII how S can be found easily when j is known.

The dynamical postulate can also be expressed in a different form. A general operator $j(x)$ can always be written in the form

$$j(x) = g(x)1 + \sum_{n=1}^{\infty}\frac{(-i)^n}{n!} \times \int g(x; \xi_1 \cdots \xi_n) :A_{\text{in}}(\xi_1) \cdots A_{\text{in}}(\xi_n): (d\xi) \quad (4.9)$$

because of the completeness of the A_{in} field. Current causality is then expressed [cf. (2.7)] by

$$g(x_1; x_2\xi_1 \cdots \xi_n) = g(x_2; x_1\xi_1 \cdots \xi_n) \quad (x_1 \sim x_2), \text{ all } n, \quad (4.10)$$

weak causality is expressed by (4.10) and

$$g(x_1; x_2\xi_1 \cdots \xi_n) = 0 \quad (x_1 < x_2), \text{ all } n, \quad (4.11)$$

and strong causality is expressed by

$$g(x_1; x_2\xi_1 \cdots \xi_n) = 0 \quad (x_1 \lesssim x_2), \text{ all } n. \quad (4.12)$$

Equation (4.4) then expresses $\beta(x_1, x_2)$ by

$$\beta(x_1, x_2) = (\bar{\theta}_{12} - 1)\Big[g(x_1; x_2)1 + \sum_{n=1}^{\infty}\frac{(-i)^n}{n!}\int g(x_1; x_2\xi_1 \cdots \xi_n) \times :A_{\text{in}}(\xi_1) \cdots A_{\text{in}}(\xi_n): (d\xi)\Big]. \quad (4.13)$$

Thus, the dynamical postulate specifies the dependence of $g(x; \xi_1 \cdots \xi_n)$ on x and ξ_1 whenever $(x - \xi_1)^2 \geq 0$, a distribution whose only support is the point $x = \xi_1$. But, since g is symmetric in the ξ, this must hold for all ξ_k. Therefore, unless it vanishes, g has the form

$$g(x; \xi_1 \cdots \xi_n) = [\sum_{\text{perm}} \beta_{i_1}(x - \xi_1) \cdots \beta_{i_n}(x - \xi_n)] \times h(\xi_1 \cdots \xi_n) \quad \text{for} \quad (x - \xi_k)^2 \geq 0, \quad k = 1, \cdots, n, \quad (4.14)$$

where each $\beta_k(x)$ is a distribution whose only support is the point $x = 0$, and h is a symmetric function of its arguments. Equation (4.9) together with (4.14) is the general solution of (4.4) or (4.8).

V. RESTRICTION ON β_{12}

Surprisingly, the assumptions of the theory give a restriction on β_{12} which could not have been anticipated: the functional dependence of β_{12} is essentially restricted to a δ function with no more than three derivatives.

In order to show this, we have to introduce the interpolating field which so far has not occurred. It plays a subordinate role since it is defined in terms of the current, i.e., the S operator, either by

$$A(x) \equiv A_{\text{in}}(x) - \int \Delta_R(x - \xi)j(\xi)\, d^4\xi \quad (5.1)$$

or by

$$A(x) \equiv A_{\text{out}}(x) - \int \Delta_A(x - \xi)j(\xi)\, d^4\xi.$$

We work with the in-fields throughout.

As was shown previously,[6] the commutator of $A(x)$ can be expressed in terms of $j(x)$, using strong unitarity, as follows:

$$[A_1, A_2] = -i\Delta_{12} + i\int \Delta(x_1 - y_1)\,\Delta(x_2 - y_2)\, d^4y_1\, d^4y_2 \times \Big[\theta(x_2 - y_2)\theta(y_1 - x_1)\frac{\delta j(y_2)}{\delta y_1} - \theta(x_1 - y_1)\theta(y_2 - x_2)\frac{\delta j(y_1)}{\delta y_2}\Big]. \quad (5.2)$$

From this one easily obtains

$$\theta_{12}[A_1, A_2] = -i\Delta^R_{12} + i\int \Delta(x_1 - y_1)\Delta(x_2 - y_2)\, d^4y_1\, d^4y_2 \times \left[\theta(y_1x_1x_2y_2)\frac{\delta j(y_2)}{\delta y_1} + \theta(y_2x_2x_1y_1)\frac{\delta j(y_1)}{\delta y_2} - \theta(x_1 - y_1)\theta(y_2 - x_2)\frac{\delta j(y_1)}{\delta y_2}\right], \quad (5.3)$$

where

$$\theta(xyrs) \equiv \theta(x - y)\theta(y - r)\theta(r - s). \quad (5.4)$$

Because of (strong) Bogoliubov causality, the first two terms in parenthesis vanish and one is left with[7]

$$\theta_{12}[A_1, A_2] = -i\Delta^R_{12} + i\int \Delta_R(x_1 - y_1) \times \Delta_A(x_2 - y_2)\frac{\delta j(y_1)}{\delta y_2}\, d^4y_1\, d^4y_2. \quad (5.5)$$

The Klein–Gorden operator K_2 acting on x_2 on the right reduces that part of the equation to $+i\delta A(x_1)/\delta x_2$ because of (5.1). Therefore

$$i(\delta A_1/\delta x_2) = K_2(\theta_{12}[A_1, A_2]). \quad (5.6)$$

This is exactly the dynamical axiom assumed by Pugh,[2] which follows here from Bogoliubov causality in addition to the other assumptions which we share with his treatment. Assuming the latter, it is in fact easy to show that (5.6) is equivalent to Bogoliubov causality. To this end we note that (5.6) was just obtained from causality and that, conversely, (5.6) implies

$$i\frac{\delta j_1}{\delta x_2} = iK_1\frac{\delta(A_1 - A_1^{\text{in}})}{\delta x_2} = K_1K_2(\theta_{12}[A_1, A_2]) - iK_1\,\delta(x_1 - x_2). \quad (5.7)$$

But the right side vanishes for $x_1 < x_2$ and $x_1 \sim x_2$ and therefore implies Bogoliubov causality.

We now recall the result[2] that strong unitarity together with (5.6) gives

$$i^nS^*\frac{\delta^nS}{\delta x_1 \cdots \delta x_n} = K_1K_2 \cdots K_n\varphi(A_1 \cdots A_n). \quad (5.8)$$

The φ product is defined in terms of a sum of time-ordered products with the same coefficients that the expansion of a normal-ordered product of free fields has.

For $n = 1$ (5.8) is just $KA = j$ which is a consequence of the defining equation (5.1) for A. For $n = 2$ it is a consequence of (5.6),

$$\begin{aligned} K_1K_2\varphi(A_1A_2) &= K_1K_2[T(A_1A_2) + i\Delta^c_{12}] \\ &= K_1K_2(\theta_{12}[A_1, A_2] + A_2A_1 + i\Delta^c_{12}) \\ &= iK_1(\delta A_1/\delta x_2) + j_2j_1 - iK_1\,\delta(x_1 - x_2) \\ &= i(\delta j_1/\delta x_2) + j_2j_1 = -S^*(\delta^2S/\delta x_1\,\delta x_2). \end{aligned} \quad (5.9)$$

The last equation follows from (2.2′). For $n > 2$ the equations (5.8) follow by induction using (5.6) and (2.1), as was shown by Pugh.[2]

We conclude that only strong unitarity and Bogoliubov causality need to be added to the usual assumptions of asymptotic quantum field theory to assure (5.8).

Having established the equivalence of our formalism with that of Pugh, we can take over another result previously established: Eq. (5.9) implies (4.5) with the restriction that β_{12} be a solution of the equation

$$B\beta = \beta, \quad (5.10)$$

where the B operator was defined previously.[1] This restricts β as indicated at the beginning of this Section.

If we apply this restriction to Eqs. (4.13) and (4.14) we find that each factor β_k must be a solution of (5.10).

VI. SOLUTION FOR S IN TERMS OF j

We now show how the S operator is obtained explicitly from the solution of (4.8), i.e., from $j(x)$ in the form (4.9) with known g.

Define the symmetric operator function

$$J_n \equiv J(x_1x_2 \cdots x_n) \equiv S^*i^n\frac{\delta^nS}{\delta x_1 \cdots \delta x_n}; \quad (6.1)$$

since

$$S = 1 + \sum_{n=1}^{\infty}\frac{(-i)^n}{n!}\int \langle J_n\rangle_0 \times :A_1^{\text{in}} \cdots A_n^{\text{in}}: dx_1 \cdots dx_n, \quad (6.2)$$

it follows that the S operator is known when the vacuum expectation values of (6.1) are known on the mass shell.

Since the operator $j(x) = J(x)$ is now assumed

[7] For $y_1^0 = y_2^0$, one has necessarily $y_1^0 = x_1^0$ and $y_2^0 = x_2$ because of the θ functions. But in that case, the Δ-functions vanish.

to be known, it is sufficient to establish a recursion relation for J_n. In fact,

$$J_n = iS^* \frac{\delta}{\delta x_n}\left(SS^* i^{n-1} \frac{\delta^{n-1} S}{\delta x_{n-1} \cdots \delta x_1}\right)$$

$$= j(x_n) J_{n-1} + i \frac{\delta}{\delta x_n} J_{n-1}. \tag{6.3}$$

Thus, the knowledge of $j(x)$ determines J_n for all $n \geq 2$ and therefore the S-operator.

We note an alternative expression to (6.3)

$$J(x_1 \cdots x_n) = \prod_{k=2}^{n}\left(j_k + i\frac{\delta}{\delta x_k}\right) j_1, \tag{6.4}$$

which follows from (6.3) and the symmetry of J_n.

VII. DYNAMICAL POSTULATE AND HAMILTONIAN

In order to establish the connection between the Hamiltonian formulation of field theory and the dynamical postulate, let H be the space–time integral over the Hamiltonian density and assume the following relation between S and H,

$$S = (e^{iH})_+. \tag{7.1}$$

From (4.5) one then finds[7]

$$\beta(x_1 x_2) = iS^*[(\delta^2 H/\delta x_1\, \delta x_2) S]_+. \tag{7.2}$$

The subscript $+$ indicates positive time ordering with respect to the asymptotic free fields $A_{1n}(x)$.

As an example, assume

$$H = \frac{1}{n!}\int :A_{1n}^n(x_1): dx_1. \tag{7.3}$$

It yields

$$\beta(x_1 x_2) = i\,\delta(x_1 - x_2)$$

$$\times S^*(:A_{1n}^{n-2}(x_1): S)_+/(n-2)! \qquad (n \geq 2). \tag{7.4}$$

The special cases

$$\beta(x_1 x_2) = i\,\delta(x_1 - x_2)1 \qquad (n = 2) \tag{7.5}$$

and

$$\beta(x_1 x_2) = i\,\delta(x_1 - x_2) A(x_1) \qquad (n = 3) \tag{7.6}$$

are especially noteworthy. The cases $n = 1$ and $n = 2$ are trivial and give no scattering.

It is clear from these considerations that so far there seems to be nothing in the theory that indicates an essential difference between the cases $n \leq 4$ and $n > 4$. These would be expected to correspond to renormalizable and nonrenormalizable theories, respectively. The difference appears, however, when one goes to momentum space. The solution of the fundamental equation (4.8) with β as given in (7.4) and $n > 4$ leads to S-matrix elements $\langle J_n\rangle_0$ which are no longer bounded by polynomials but have in fact an essential singularity in the high-energy limit. But this is established only in perturbation expansion. Nonperturbation solutions are not known.

It is therefore desirable to solve (4.8) by nonperturbative methods. However, if one does carry out a perturbation expansion, one must first separate the second term on the right into two parts corresponding to the partition of our space $1 = (1 - B) + B$, where B is the idempotent operator mentioned in (5.10). Because of this, Eq. (4.8) becomes, after some calculation,

$$(1 - B) i \frac{\delta j_1}{\delta x_2} = (1 - B)\theta_{12}[j_1, j_2]$$

$$= -K_1 K_2 \int \Delta(x_1 - \xi_1)$$

$$\times \Delta(x_2 - \xi_2)\theta(\xi_1 x_1 x_2 \xi_2)[j(\xi_1), j(\xi_2)]\, d\xi_1\, d\xi_2$$

or[8]

$$i\frac{\delta j_1}{\delta x_2} = Bi\frac{\delta j_1}{\delta x_2} + K_1 K_2 \theta_{12}$$

$$\times \int \Delta_A(x_1 - \xi_1)\, \Delta_R(x_2 - \xi_2) i \frac{\delta j(\xi_1)}{\delta \xi_2}\, d\xi_1\, d\xi_2. \tag{7.7}$$

One then starts the perturbation expansion by assuming this integral on the right side to be zero in first approximation and one solves by iteration. Note that the integral term of (7.7) must vanish when operated on by B.

VIII. CONCLUSIONS

The results which we have obtained here for the neutral scalar field, and which clearly permit generalization to charged fields, other spins, and, in particular, to quantum electrodynamics, permit one to formulate quantum field theory very briefly as follows.

The axioms of the asymptotic free field, Lorentz invariance, and the completeness of the asymptotic field, must be amended by two more axioms: strong unitarity and Bogoliubov causality. These axioms then lead to an S operator which together with all its derivatives is determined by the φ product of the interpolating field. No terms which vanish on the mass shell can be added to this product and the

[8] An equivalent equation was recently published by R. E. Pugh, [J. Math. Phys. 7, 376 (1966)].

off mass-shell values of the S-matrix elements are therefore determined by the interpolating field. The latter is defined by the current for which we obtained an equation. This equation requires the specification of the interaction (dynamical postulate). Its solution is therefore the central mathematical problem of the theory.

This formulation of quantum field theory is free of divergences and does not require renormalization, as was shown explicitly for the case of quantum electrodynamics by Pugh.[1] It is therefore the basis of a satisfactory theory for all those cases where one has polynomial boundedness in momentum space. This corresponds exactly to those theories of the conventional type which are called renormalizable.

Whether the present formulation also works for some or all of the so-called nonrenormalizable theories is not known at present. For this purpose it will be necessary to know whether the equation for the current has nonperturbative solutions.

Note added in proof: The restriction on β_{12} discussed in Sec. 5 [i.e., that it be a solution of Eq. (5.10) and that it therefore contains no more than three time-derivatives of a delta function] has the following origin. The first two terms in the square bracket on the right side of (5.3) vanish due to Bogoliubov causality only when $\delta j(y_1)/\delta y_2$ for $y_1^0 = y_2^0$ is of the form $\partial_0^n[\delta(y_1 - y_2)j(y_1y_2)]$ with $n < 4$; otherwise $\delta j_1/\delta y_2$ can be made to contain two Klein–Gordon operators which (differentiating by parts) replace the $\Delta_A\Delta_R$ by Dirac delta functions. These terms then do *not* vanish. Because of (4.4), this same restriction must then also hold for β_{12}.

In a future publication, it will be shown that this theory leads to undefined products of distributions unless the distribution $\delta j_1/\delta y_2$ in convolutions such as (5.3) are restricted as indicated. This is the mathematical reason for the need to extend the present formalism if it is to be applicable to nonrenormalizable theories.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 8 AUGUST 1966

Wave Mechanics in Classical Phase Space, Brownian Motion, and Quantum Theory

GIACOMO DELLA RICCIA*
Department of Mathematics, Indiana University, Bloomington, Indiana
AND
NORBERT WIENER†
Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received 8 March 1965; revised manuscript received 22 December 1965)

A wave dynamics of fields $\varphi(p, q; t) \in L_2(\Gamma)$ over the phase space $\Gamma(p, q)$ of a classical system S is derived from the Liouville theorem. We define the energy contained in a given field $\varphi(p, q; t)$. We show that for a special class of fields, selected on physical grounds, the energy spectrum is given by a time-independent Schrödinger equation. This allows us to associate with S an ordinary quantum system Q such that the values of the quantized energy coincide for the fields in the phase space of S and for Q. Then we make use of Wiener's stochastic integral based on the theory of Brownian motion to derive probabilities which are the same as those one would obtain through Born's statistical postulate of quantum theory. From this it follows that we can regard normalized fields $\varphi(p, q; t)$ as "probability amplitudes" leading to a probability density function $\rho(p, q; t) = \varphi\varphi^*$ in the sense of Gibbs' statistical mechanics. Our work therefore appears as a bridge between a statistical theory (in the sense of Gibbs) of a mechanical system S and the usual quantum theory of the related quantum system Q.

INTRODUCTION

IN the usual basis for the Schrödinger equation, there are two radical departures from classical physics. One of them is to replace the observables occurring in classical dynamics by operators, and in terms of these operators to deduce a new probability theory of the von Neumann type, which gives a postulational basis for the probabilities of quantum theory by starting with a totally new set of formal assumptions. The other departure is that the dynamics by which we replace classical dynamics is itself a dynamics of operators. In both cases we are replacing classical physics by a new physics in which the first principles are different.

From the very beginning of quantum theory, there has been a widespread suspicion that this modification of classical physics is too radical and that there is a more direct transition to be made from a genuine dynamics of the Hamiltonian type to quantum theory. This suspicion has been expressed by L. de Broglie, D. Bohm, and J. P. Vigier and by a considerable group of theoretical physicists working together.

In the work of de Broglie, the classical dynamics which he has tried to use as the basis of quantum theory is a classical dynamics to which quantum mechanics is asymptotic when we treat the quantum constant $\hbar$ as very small. However, in one way or another this approach has run into many difficulties.

The present paper is based on the suggestion that, although quantum theory can be carried back to a classical dynamics, this classical dynamics is not that to which quantum dynamics is asymptotic in the sense of de Broglie. Philosophically, this means that our point of view is identical to that of de Broglie, Bohm, etc., in the reduction of quantum physics to a deterministic system in which probabilities are dependent on our ignorance of certain hidden variables; but here the hidden variables occur in a different manner from that in which they have occurred in previous works. Nevertheless, as with the previous authors, the basis for the ignoring of hidden variables is that of statistical mechanics.

In Gibbs statistical mechanics the basic quantity is a probability density function $\rho(p, q; t)$ defined over the phase space of the mechanical system S under observation. In our work we introduce in phase space a new quantity $\varphi(p, q; t)$, which by definition is a normalized square-integrable, real, or complex-valued function. We call it a "probability amplitude" field. $\varphi(p, q; t)$ is required to satisfy the usual equation of continuity derived from the Liouville theorem. We then build a field dynamics in phase space, and consequently, we can define the energy contained in a given field. An important result is that there exists a class of fields, characterized by physical properties, for which the values of the energy are quantized. We show that this

* G. Della Riccia acknowledges partial support for this work by the National Science Foundation under Contract NSF-GP-149, while he served as Research Associate in the Department of Mathematics of MIT, during 1963–64, on leave of absence from Istituto di Fisica Teorica dell' Università-Naples, Italy.
† Deceased.

quantization is completely determined by a time-independent Schrödinger equation in configuration space $\mathcal{C}(q)$. The method we present of arriving at this Schrödinger equation starting from the Liouville theorem is new. Thanks to these result, it is possible to associate with a given quantum system Q an appropriate mechanical system $\mathcal{S}$ such that the energy spectrum of the quantum system and of the system of fields in phase space of $\mathcal{S}$ coincide. Another interesting consequence of the existence of this Schrödinger equation is that ordinary statistical states of the quantum system can be obtained from statistical states of $\mathcal{S}$ by averaging the "probability amplitudes" in phase space over all possible values of the classical momenta p. As a result, the p variables no longer appear in the quantum theory associated with the statistical theory of the classical system $\mathcal{S}$. In this sense, the p's play the role of hidden variables for the quantum system Q for which we define new momenta by the standard method. Clearly, at no time do we violate the uncertainty principle since the p's which can be measured simultaneously with the q's and with the same accuracy, belong to a classical system.

In the last part of our work we are concerned with the problem of constructing probabilities out of "probability amplitudes." We use known results based on Wiener's mathematical theory of Brownian motion to derive probabilities which agree with those obtained from Born's statistical postulate. The desired result is that it is possible to interpret the quantity $\rho(p, q; t) = \varphi\varphi^*$ as a probability density in the sense of Gibbs.

Hence the use of "probability amplitudes" is justified since it can be tied up with the methods of Gibbs. More generally, we wish to present our theory as a bridge between the Gibbs theory for classical systems and the usual quantum theory for quantum systems. It may be possible that further investigation will prove that the ultimate physical reality is the mechanical system $\mathcal{S}$.

Because of the sad demise of Norbert Wiener in March 1964, the treatment given in this paper is due to the first-named author. For the same reason it seemed desirable that the results should be presented, however incomplete they may be.

I. DEFINITION, IN A CLASSICAL PHASE SPACE, OF "PROBABILITY AMPLITUDES" AND OF A FIELD DYNAMICS

Let us consider a mechanical system $\mathcal{S}$ with N degrees of freedom. Its phase space is denoted by $\Gamma(p, q)$, or simply Γ, where q stands for the N coordinates $q_1, \cdots, q_N$ in configuration space $\mathcal{C}(q)$ and p represents the associated N momenta $p_1, \cdots, p_N$. Let us assume that $\mathcal{S}$ is conservative, so that its Hamiltonian function $H(p, q)$ does not depend explicitly on the time t. Under fairly general conditions, the Hamilton equations of motion,

$$\begin{aligned} dq_n/dt &= \partial H/\partial p_n, \\ dp_n/dt &= -\partial H/\partial q_n, \end{aligned} \qquad (n = 1, 2, \cdots, N) \tag{1}$$

admit a unique solution for prescribed initial conditions (p, q). Hence, in accordance with (1), to each point $M(p, q) \in \Gamma$ considered at some initial time, say $t_0 = s$, we can associate after an interval of time t (positive or negative) one and only one point $M_t = T^t M$. T^t represents a one-parameter group of one-to-one point transformations of Γ onto itself:

$$T^{s+t} = T^s T^t = T^t T^s \quad \text{for all} \quad t, s,$$
$$T^0 = 1.$$

Under the group of transformations T^t, a point M describes a trajectory in phase space (one and only one for each M), and the velocity $\mathbf{v}(p, q)$ of $M(p, q)$ on its trajectory has $2N$ components $(dp/dt, dq/dt)$ which are defined by (1). The position of M at a given time t characterizes completely the state of the system $\mathcal{S}$.

The celebrated Liouville theorem states that T^t is a measure preserving transformation on Γ; the measure whose value remains invariant under the flow T^t is the Lebesgue measure $dp\ dq = dp_1 \cdots dp_N\ dq_1 \cdots dq_N$. Let us now consider a function $f_0(p, q)$ which is locally L_1 (Lebesgue) and define

$$f(p, q; t) = f_0(T^{-t}p, T^{-t}q).$$

If f_0 is sufficiently differentiable, one can readily show, as a corollary of Liouville's theorem, that f must satisfy an equation of continuity

$$\partial f/\partial t + \operatorname{div}(f\mathbf{v}) = 0,$$

which because of (1) reduces to

$$\partial f/\partial t + [H, f] = 0, \tag{2}$$

where

$$[H, \cdot] = \sum_{n=1}^{n=N} \left(\frac{\partial H}{\partial p_n} \frac{\partial}{\partial q_n} - \frac{\partial H}{\partial q_n} \frac{\partial}{\partial p_n} \right)$$

is the classical Poisson bracket operator. It is easy to see that an equivalent expression for the Poisson bracket is $[H, f] = \mathbf{v} \cdot \nabla f$. The latter expression shows that $[H, f]$ represents the time rate of change of the value of the function f as M describes its trajectory.

In usual Gibbs statistical mechanics, one considers only those solutions of (2), $\rho(p, q; t)$ say, which are real, nonnegative, and belong to the space of functions L_1. The reason for that special choice is that in the usual theory one needs to interpret $\rho(p, q; t)$ as a probability density function. In the present work we are rather interested in solutions of (2), $\varphi(p, q; t)$ say, which may be real or complex-valued and which belong to Hilbert space L_2 (Γ, Lebesgue). To emphasize this fact, we rewrite (2) as

$$\partial\varphi/\partial t + [H, \varphi] = 0, \qquad \varphi \in L_2. \tag{3}$$

We call (3) the *Liouville equation* and a solution $\varphi(p, q; t)$ of (3) the "probability amplitude" wave or field. The reason for borrowing the name for φ from the vocabulary of quantum theory is that, as shown later (in Sec. III), when φ is normalized, the quantity $\rho(p, q; t)\, dp\, dq = \varphi\varphi^*\, dp\, dq = |\varphi|^2\, dp\, dq$ can be interpreted as the probability of finding, at time t, the mechanical system $\mathcal{S}$ in a state M in the interval $[(p, q), (p + dp, q + dq)]$. This is, of course, very similar to the relationship between Schrödinger waves $\Psi(q, t)$ and probability densities $\rho(q, t) = \Psi\Psi^*$ of quantum theory. It should be noticed that if φ is a solution of (3), then not only φ^* but also $\rho(p, q; t) = \varphi\varphi^*$ satisfies the same equation. This means that the probability densities defined through the above relationship with the "probability amplitudes" coincide with the probability densities considered by Gibbs. In this sense, we can say that the present work is consistent with Gibbs statistical mechanics. Before deriving any result from this new concept of "probability amplitudes" in phase space we wish to investigate further the wave character of the φ's. For this purpose let us introduce space, we wish to investigate further the wave character of the φ's. For this purpose, let us introduce the Liouville operator $\mathfrak{L} = i[H, \cdot]$, where $i = (-1)^{\frac{1}{2}}$. It can be shown[1] that if $H(p, q)$ is a sufficiently smooth function (and we assume that it is always our case), then $\mathfrak{L}$ is a self-adjoint operator; thus, its spectrum is real. Equation (3) becomes

$$-i(\partial\varphi/\partial t) = \mathfrak{L}\varphi, \qquad \varphi \in L_2. \tag{4}$$

A solution is completely determined by its value $\varphi_0(p, q)$ at $t = 0$. Indeed the solution of (4) is formally $\varphi(p, q; t) = \exp(i\mathfrak{L}t)\varphi_0(p, q)$, $\varphi_0 \in L_2$. Here $U^t = \exp(i\mathfrak{L}t)$ is a one-parameter group of unitary transformations since $\mathfrak{L}$ is self-adjoint. The propagation laws in phase space can be easily discussed if one considers eigenfunctions of (4), that is, solutions $\Phi(p, q; t)$ such that

$$\Phi(p, q; t) = U^t\Phi_0 = \exp(i\omega t)\Phi_0(p, q),$$

or equivalently $\mathfrak{L}\Phi_0 = \omega\Phi_0$, where it is assumed that the spectrum $\{\omega\}$ of $\mathfrak{L}$ is discrete. It is easy to verify that the same discussion would hold in the general case where the spectrum is continuous but with an appropriate change of notations.

Let us write $\Phi_0 = A_0(p, q) \exp[-i\theta_0(p, q)]$ with A_0 and θ_0 being two real functions independent of time. Their dependence on the particular value of ω in the spectrum is not shown explicitly. Let us substitute this expression of Φ_0 into $\mathfrak{L}\Phi_0 = \omega\Phi_0$, which is satisfied if the real and the imaginary parts separately vanish, that is,

$$[H, A_0] = 0, \tag{5a}$$

$$[H, \theta_0] = \omega. \tag{5b}$$

From (5b) we see immediately that if $\omega \neq 0$ then $\theta_0(p, q)$ is not identically 0, thus the corresponding eigenfunction Φ_0 is necessarily complex. Furthermore, if $\mathfrak{L}\Phi_0 = \omega\Phi_0$, then $\mathfrak{L}\Phi_0^* = -\omega\Phi_0^*$, which follows from the fact that $\mathfrak{L}^* = -\mathfrak{L}$ (see definition of $\mathfrak{L}$). Hence the spectrum of $\mathfrak{L}$ is always symmetric with respect to $\omega = 0$ and if Φ_0 is the eigenfunction associated with ω, Φ_0^* is also an eigenfunction corresponding to the eigenvalue $-\omega$.

Let us first discuss the case $\omega \neq 0$. The surfaces of constant value of $\theta_0(p, q)$ have fixed locations in Γ. A surface characterized by a constant value of the phase $\theta(p, q; t) = \omega t - \theta_0(p, q)$ of the eigenfunction $\Phi(p, q; t)$ must coincide, at given t, with some particular surface of constant θ_0. However, the value of θ_0 corresponding to a fixed value of θ changes with time. For instance, if at $t = 0$ the surface $\theta = a$ coincides with the surface $\theta_0 = -a$, at a time dt later the surface $\theta = a$ will coincide with the surface characterized by $\theta_0 = -a + \omega\, dt$; thus the time rate of change of the value of the constant θ_0 value for which the coincidence with $\theta = a$ occurs is equal to ω. These surfaces of constant θ values are the *wave fronts*. Equation (5b) tells us something more. Since it can also be written $\mathbf{v}(p, q) \cdot \nabla\theta_0(p, q) = \omega$, we see that as we follow the flow T^t, we find that the time rate of change of the value of θ_0 is again ω. It is therefore obvious that the traveling of the wave fronts keeps in step with the flow T^t, and it is not difficult to see that points of two distinct wave fronts are put in one-to-one correspondence under that flow: Equivalent points are obtained by intersecting wave fronts with the

[1] G. W. Mackey, *Mathematical Foundations of Quantum Mechanics* (W. A. Benjamin, Inc., New York, 1963).

rays defined as trajectories of T^t. Notice that the rays are not in general orthogonal trajectories of the family of wave fronts; the angle at which they intersect the wave fronts is given by (5b). On the other hand, Eq. (5a) indicates that the amplitude $A_0(p, q)$ of the wave $\Phi(p, q; t)$ is a constant of the motion: $A_0(p, q) = A_0(T_p^t, T_q^t)$ for all t. The rays are lines of constant value of the amplitude. In the case where $\omega = 0$, Eq. (5b) becomes $[H, \theta_0] = 0$, meaning that θ_0 is also a constant of the motion. But in this case we can take $\theta_0 \equiv 0$ without loss of generality. In fact, there is no need to talk about waves, since, in this case, we have a steady (time-independent) solution of the Liouville equation in spite of the fact that the flow itself T^t never stops. A general solution of the Liouville equation is a linear superposition of waves corresponding to a set (in general continuous) of values of ω in the spectrum of $\mathfrak{L}$.

From the previous discussion we might think that the Liouville equation is the only wave equation we need. This is not so because, as we have seen, it only describes waves whose wave fronts all propagate in step with the velocity field $\mathbf{v}(p, q)$ defined in (1). A complete mathematical description of waves in Γ space should also allow waves to propagate according to the reversed motion corresponding to a velocity field $-\mathbf{v}(p, q)$. In this sense, the Liouville equation is too restrictive, and we must associate with it a "modified" Liouville equation describing this new type of waves.

Since the Liouville equation (3) is

$$(\partial\varphi/\partial t) + \mathbf{v}(p, q)\cdot\nabla\varphi = 0,$$

the "modified" Liouville equation has to be

$$\partial\varphi/\partial t - \mathbf{v}(p, q)\cdot\nabla\varphi = 0.$$

Thus the complete equation of motion in Γ space must include all solutions of either equation below,

$$(\partial\varphi/\partial t) + [H, \varphi] = 0, \tag{6a}$$

$$(\partial\varphi/\partial t) - [H, \varphi] = 0, \tag{6b}$$

with $\varphi \in L_2$ in all cases.

An essential physical difference between solutions $\varphi(p, q; t) = \exp(i\mathfrak{L}t)\varphi_0(p, q)$ of (6a) and solutions $\varphi(p, q; t) = \exp(-i\mathfrak{L}t)\varphi_0(p, q)$ of (6b) is made clear by the following remark.

Remark: In classical mechanics the time-reversal operation $t \to -t$ is necessarily followed by the change $p \to -p$. Usually these two transformations leave the equations of motion (1) invariant. Thus the result of $t \to -t$ is simply to replace a solution of (1) corresponding to initial conditions $M(p_0, q_0)$ at $t = 0$ by a solution corresponding to a new set of initial conditions $M(-p_0, q_0)$. In general, the trajectories in phase space corresponding, respectively, to $M(p_0, q_0)$ and $M(-p_0, q_0)$ are different. In our case the situation is completely different since the replacement of $\mathbf{v}(p, q)$ by $-\mathbf{v}(p, q)$ amounts to the change $t \to -t$ in (1) while (p, q) remain the same. But the transformation $t \to -t$ and $(p, q) \to (p, q)$ is itself equivalent to the transformation $t \to t$, $(p, q) \to (p, q)$, and $H(p, q) \to -H(p, q)$. Now a change of sign of this sort in the Hamiltonian function can only be achieved by a change of sign of the physical constants (masses, electric charges, coupling constants, etc.) which appear in H. This means that we have to replace the mechanical system characterized by H by a "modified" system $\mathcal{S}'$. If at $t = 0$ we assign the same initial conditions $M(p_0, q_0)$ to $\mathcal{S}$ and $\mathcal{S}'$, then the *same* trajectory in phase space will describe the time evolution of the two system but the representative point in each case will move in opposite directions. However, the system $\mathcal{S}'$ has no physical existence in the classical sense since it corresponds to negative masses, etc. Unless, of course, we use the type of arguments which, in relativistic quantum mechanics, lead to the concepts of electrons and positrons, particles and antiparticles, or in the theory of semiconductors to the concepts of electrons and holes. We do not wish for the moment to introduce such a physical interpretation of $\mathcal{S}'$, although it may turn out to have interesting implications in future developments. In the present work we use $\mathcal{S}'$ as a mathematical device which is convenient for the purpose of deriving the complete wave equation in phase space.

We now return to the pair of equations (6). By partial differentiation with respect to t, we obtain the pair of equations

$$(\partial^2\varphi/\partial t^2) + [H, \partial\varphi/\partial t] = 0, \tag{6a'}$$

$$(\partial^2\varphi/\partial t^2) - [H, \partial\varphi/\partial t] = 0. \tag{6b'}$$

Then we replace in (6a') $\partial\varphi/\partial t$ by $-[H, \varphi]$ obtained from (6a) and in (6b') $\partial\varphi/\partial t$ by $[H, \varphi]$ obtained from (6b). In both cases we obtain the same equation:

$$-\partial^2\varphi/\partial t^2 = -[H, [H, \varphi]] = \mathfrak{L}^2\varphi, \quad \varphi \in L_2. \tag{7}$$

To be sure, all solutions of either (6a) or (6b) are solutions of (7), but the converse is not true. For instance,

$$\varphi(p, q; t) = \exp(i\mathfrak{L}t)f_0(p, q) + \exp(-i\mathfrak{L}t)g_0(p, q), \qquad f_0, g_0 \in L_2$$

is solution of (7) but it does not satisfy (6a) nor (6b). Equation (7) has the required time-reversal invariance property. We postulate that (7) is the *field equation of motion* in phase space.

In the next section we show how the properties of a general field $\varphi(p, q; t)$ are derived from (7). But as a rule we always particularize the results to the case where φ is also a solution of the Liouville equation (or solution of the "modified" Liouville equation) because we are only interested in waves whose propagation can be associated with the point-like dynamics of a mechanical system $\mathcal{S}$ (or with $\mathcal{S}'$).

II. THE FIELD DYNAMICS IN PHASE SPACE

This section is divided into two parts. In the first part we derive the field equation of motion (7) from a variational principle. This allows us to define the energy of a field $\varphi(p, q; t)$. In the last part we show that there exists a special class of fields for which the values of the energy are quantized. The interesting feature of these fields is that the values of the quantized energy can be made to coincide with the energy spectrum of an associated ordinary quantum system. The procedure shows a new way of arriving at the time-independent Schrödinger equation. This suggests that some aspects of ordinary quantum systems can be discussed in terms of a classical statistical theory of the corresponding mechanical systems.

A. Variational Method in Phase Space and the Field Energy

Let us use the Lagrangian and Hamiltonian formulations for continuous media. We only consider scalar fields $\varphi(p, q; t)$. Let the Lagrangian density of the field φ be

$$\lambda(p, q; t) = \tfrac{1}{2}A\{\dot{\varphi}\dot{\varphi}^* - [H, \varphi][H, \varphi^*]\}, \qquad (8)$$

where $\dot{\varphi} = \partial\varphi/\partial t$, the asterisk indicates complex conjugation, and A is a constant scale factor to be specified later.

The total Lagrangian is

$$L = \int_\Gamma \lambda(p, q; t)\, dp\, dq,$$

and the total action between two times t_1 and t_2 is

$$S = \int_{t_1}^{t_2} L\, dt = \frac{A}{2}\int_{t_1}^{t_2}\int_\Gamma \{\dot{\varphi}\dot{\varphi}^* - [H, \varphi][H, \varphi^*]\}\, dp\, dq\, dt. \qquad (9)$$

Let us employ the principle of least action, which tells us that (9) must be stationary when φ and φ^* are separately varied by arbitrary small quantities $\delta\varphi$ and $\delta\varphi^*$ both vanishing at times t_1 and t_2. By standard computation we find that if φ vanishes sufficiently fast when p or q become infinite [practically the condition $\varphi \in L_2(\Gamma)$ is sufficient] then the total action S is stationary if $\partial^2\varphi/\partial t^2 = [H, [H, \varphi]]$, which is the equation of motion (7) that was derived by a different method in Sec. I.

With the field variables φ and φ^*, we can associate the canonical conjugate variables π and π^*, respectively, by the usual definition

$$\pi = \partial\lambda(p, q; t)/\partial\dot{\varphi} = \tfrac{1}{2}A\dot{\varphi}^*,$$

$$\pi^* = \partial\lambda(p, q; t)/\partial\dot{\varphi}^* = \tfrac{1}{2}A\dot{\varphi},$$

where we used the expression (8) for λ.

As usual, we define the Hamiltonian or energy density of the field to be

$$h(p, q; t) = \pi\varphi + \pi^*\varphi^* - \lambda,$$

which in our case becomes

$$h(p, q; t) = \tfrac{1}{2}A\{\dot{\varphi}\dot{\varphi}^* + [H, \varphi][H, \varphi^*]\}.$$

Consequently, the total energy E of a field φ is

$$E = \tfrac{1}{2}A\int_\Gamma \{\dot{\varphi}\dot{\varphi}^* + [H, \varphi][H, \varphi^*]\}\, dp\, dq.$$

As mentioned before we are only interested in those fields φ, solutions of (7), which at the same time satisfy the Liouville equation (6a) or the "modified" Liouville equation (6b).

With this restriction it is easy to see that the energy becomes

$$E = A\int_\Gamma [H, \varphi][H, \varphi^*]\, dp\, dq.$$

By integration by parts and under the condition that $\varphi \in L_2$ so that the integrated parts vanish at infinity, we find that

$$E = -A\int_\Gamma \varphi^*[H, [H, \varphi]]\, dp\, dq = -A\int \varphi^*\frac{\partial^2\varphi}{\partial t^2}\, dp\, dq, \qquad (10a)$$

where the last equality is obtained by employing the field equation of motion (7). If we introduce the Liouville operator $\mathfrak{L}$, then

$$E = A\int \varphi^*\mathfrak{L}^2\varphi\, dp\, dq = A\langle\varphi, \mathfrak{L}^2\varphi\rangle, \quad \varphi \in L_2. \qquad (10b)$$

The notation $\langle\cdot, \cdot\rangle$ stands for the usual scalar product in $L_2(\Gamma)$.

To sum up the results: fields φ which satisfy the

Liouville equation (or the "modified" Liouville equation) have an energy E equal to the expectation value $\langle \varphi, A\mathfrak{L}^2\varphi \rangle$ of the operator $A\mathfrak{L}^2$ where the scale constant A is not yet determined.

B. The Quantization of the Field Energy

We assume that the statistical properties of the mechanical system $\mathcal{S}$ under consideration are best represented by the so-called canonical ensemble. It is well known that one chooses the canonical ensemble as a representative ensemble when it is assumed that $\mathcal{S}$ is not completely isolated from its surroundings $\mathcal{O}$. Because of the interaction with $\mathcal{O}$, the energy of $\mathcal{S}$ is allowed to fluctuate. However, the assumption is that this interaction is weak enough so that it is possible with a good approximation to assign to $\mathcal{S}$ a Hamiltonian function $H(p, q)$ which is independent of the degrees of freedom of $\mathcal{O}$. The nature of $\mathcal{O}$ need not be specified further; it is only required to play the role of a thermostat (heat bath) which, in the state of statistical equilibrium, imposes its temperature T to the system $\mathcal{S}$. According to Gibbs, this state of thermal equilibrium is represented by the so-called canonical distribution

$$\rho_0(p, q) = C \exp(-2\beta H), \qquad 2\beta = 1/kT, \tag{11}$$

where k is the Boltzmann constant, T the absolute temperature, and C a normalization constant such that $\int_\Gamma \rho_0(p, q)\, dp\, dq = 1$.

Let us specify $\mathcal{S}$, further assuming that, in an appropriate system of Cartesian coordinates in the $2N$-dimensional Γ space, its Hamiltonian function is of the form

$$H(p, q) = \sum_{n=1}^{N} \frac{p_n^2}{2m} + F(q_1, \cdots, q_N). \tag{12}$$

The mass m is assumed to be the same for all N degrees of freedom. As in most cases of interest, the potential function $F(q)$ is assumed to depend only on the configuration coordinates $q(q_1, \cdots, q_N)$ and to be independent of the momenta $p(p_1, \cdots, p_N)$. Furthermore, we only consider systems whose $F(q)$ is such that $\int_\Gamma \exp(-H/kT)\, dp\, dq$ exists so that Gibbs theory makes sense.

According to our previous definition of "probability amplitude," it is natural to associate with the canonical distribution ρ_0 the normalized field in L_2

$$\Phi_0(p, q) = C^{\frac{1}{2}} \exp(-\beta H) = C^{\frac{1}{2}} \exp\left[-\beta \sum_{n=1}^{N} \frac{p_n}{2m}\right] \exp[-\beta F(q)], \tag{13}$$

with $\beta = 1/2kT$ as above. Of course, the field Φ_0, as well as ρ_0, is a steady solution of the Liouville equation; it is thus a suitable representation of an equilibrium situation.

Now we wish to generalize the previous problem in the following sense. We consider fields that at some arbitrary initial time, say $t = 0$, are of the particular form

$$\Phi(p, q) = C^{\frac{1}{2}} \exp\left[-\beta \sum_{n=1}^{N} \frac{p_n^2}{2m}\right] W(q), \tag{14}$$

where $W(q)$ is an arbitrary function (real or complex) belonging to $L_2(\mathfrak{C})$ [the constant C taking care of the normalization of $\Phi(p, q)$ whenever this condition is required]. It is clear that elements of the type occurring in (14) generate only a sub-Hilbert space of $L_2(\Gamma)$. This subspace is the direct product of the complete Hilbert space $L_2(\mathfrak{C})$ and the subspace of momentum space spanned by the N-dimensional Hermite function of order 0 of the variables

$$\left\{P_n = \left(\frac{2\beta}{m}\right)^{\frac{1}{2}} p_n, \quad n = 1, \cdots, N\right\},$$

$$\text{that is,} \quad \exp\left[-\beta \sum_{n=1}^{N} \frac{p_n^2}{2m}\right].$$

The thermal equilibrium field Φ_0 defined in (13) is itself of the special form (14). This suggests that we may interpret fields of the form (14) as representing statistical states which has been slightly perturbed, at $t = 0$, from the state of equilibrium. It is physically conceivable that small perturbations will leave the Maxwell distribution of momenta untouched but will affect the distribution in configuration space $\mathfrak{C}(q)$. Hence the result of the disturbance is the replacement of $\exp[-\beta F(q)]$ by any other $W(q) \in L_2(\mathfrak{C})$.

According to the Liouville equation, the "probability amplitude," at a later time t, will be

$$\varphi(p, q; t) = U^t \Phi(p, q) = \exp(i\mathfrak{L}t)\Phi(p, q).$$

In general, $\varphi(p, q; t)$ will no longer be of the form (14) (a function of p times a function of q); however, the value of the field energy is independent of time since it was derived from a Lagrangian formulation. Therefore, for convenience, let us compute this energy at $t = 0$, where we can take advantage of the particular form assumed in (14). According to (10a),

$$E = -A \int_\Gamma \Phi^*[H, [H, \Phi]]\, dp\, dq.$$

Using expression (12) for H we have

$$[H, \Phi] = \sum_{n=1}^{N}\left(\frac{\partial\Phi}{\partial q_n}\frac{\partial H}{\partial p_n} - \frac{\partial\Phi}{\partial p_n}\frac{\partial H}{\partial q_n}\right)$$
$$= C^{\frac{1}{2}} \exp\left[-\beta \sum_{n=1}^{N}\frac{p_n^2}{2m}\right]$$
$$\times\left\{\sum_{n=1}^{N}\frac{p_n}{m}\left(\frac{\partial}{\partial q_n} + \beta\frac{\partial F}{\partial q_n}\right)W(q)\right\}$$

and

$$[H, [H, \Phi]] = C^{\frac{1}{2}} \exp\left[-\beta \sum_{n=1}^{N}\frac{p_n^2}{2m}\right]$$
$$\times\left\{\sum_{n,k=1}^{N}\sum\left(\frac{\partial}{\partial q_n} + \beta\frac{\partial F}{\partial q_n}\right)\left(\frac{\partial}{\partial q_k} + \beta\frac{\partial F}{\partial q_k}\right)\right.$$
$$\left. - \sum_{n=1}^{N}\frac{1}{m}\frac{\partial F}{\partial q_n}\left(\frac{\partial}{\partial q_n} + \beta\frac{\partial F}{\partial q_n}\right)\right\}W(q).$$

The terms of this last expression can be grouped in the following way:

$$[H, [H, \Phi]] = C^{\frac{1}{2}} \exp\left[-\beta \sum_{n=1}^{N}\frac{p_n^2}{2m}\right]$$
$$\times\left\{\frac{1}{2\beta m}\Delta + \frac{1}{2m}\left[\Delta F - \beta\sum_{n=1}^{N}\left(\frac{\partial F}{\partial q_n}\right)^2\right]\right\}W(q)$$
$$+ \frac{C^{\frac{1}{2}}}{2\beta m}\exp\left[-\beta\sum_{n=1}^{N}\frac{p_n^2}{2m}\right]\left\{\sum_n\left(\frac{2\beta}{m}p_n^2 - 1\right)\right.$$
$$\left.\times\left(\frac{\partial}{\partial q_n} + \beta\frac{\partial F}{\partial q_n}\right)^2\right\}W(q) + \frac{C^{\frac{1}{2}}}{2\beta m}\exp\left[-\beta\sum_{n=1}^{N}\frac{p_n^2}{2m}\right]$$
$$\times\left\{\sum_{\substack{n,k=1\\ n\neq k}}^{N}\sum\frac{2\beta}{m}p_n p_k\left(\frac{\partial}{\partial q_n} + \beta\frac{\partial F}{\partial q_n}\right)\right.$$
$$\left.\times\left(\frac{\partial}{\partial q_k} + \beta\frac{\partial F}{\partial q_k}\right)\right\}W(q), \qquad (15)$$

where Δ is the the Laplacian operator with respect to the q variables.

Equation (15) is the sum of three terms; the first involves the N-dimensional Hermite function of order 0 of the variables

$$\{P_n = (2\beta/m)^{\frac{1}{2}}p_n; n = 1, \cdots, N\},$$

the second and the third involve similar Hermite functions but of higher order. On account of the orthogonality of Hermite function of different order, it is readily seen that these two last terms, which are orthogonal to Φ^* because it involves the 0th-order Hermite function, give no contribution to the energy integral. Thus we have simply that

$$E = -\frac{AC}{\beta}\int_{\Gamma}\exp\left[-2\beta\sum_n\frac{p_n^2}{2m}\right]W^*(q)$$
$$\times\left\{\frac{1}{2m}\Delta + \frac{\beta}{2m}\left[\Delta F - \beta\sum_n\left(\frac{\partial F}{\partial q_n}\right)^2\right]\right\}W(q)\,dp\,dq$$

or

$$E = -\frac{A}{\beta}\int_{\mathfrak{C}(q)}W^*(q)\left\{\frac{1}{2m}\Delta + \frac{\beta}{2m}\right.$$
$$\left.\times\left[\Delta F - \beta\sum_n\left(\frac{\partial F}{\partial q_n}\right)^2\right]\right\}W(q)\,dq, \qquad (16)$$

where we assume that C is such that

$$C\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left[-2\beta\sum_n\frac{p_n^2}{2m}\right]dp = 1$$

and $\int_{\mathfrak{C}} W^*(q)W(q)\,dq = 1$.

Let us define the potential function $F(q)$ by the relation

$$\frac{\beta}{2m}\left[\beta\sum_n\left(\frac{\partial F}{\partial q_n}\right)^2 - \Delta F\right] = \frac{1}{\hbar^2}[V(q) + a], \qquad (17)$$

where $V(q)$ is some given function and a some given constant. We want $V(q)$ and a to have dimensions of "energy." From a consideration of the various quantities appearing in the left-hand side of (17), it appears immediately that $\hbar$ must have dimensions of an action (momentum times position). We can choose $\hbar$ to be the reduced Planck's constant $h/2\pi$ without loss of generality since we still have at our disposal the scale constant A which will be fixed shortly. Substituting (17) into (16), we find

$$E = -\frac{A}{\beta\hbar^2}\int_{\mathfrak{C}}W^*(q)\left[\frac{\hbar^2}{2m}\Delta - V(q) - a\right]W(q)\,dq$$

or

$$E = \frac{A}{\beta\hbar^2}\int_{\mathfrak{C}}W^*(q)[\mathfrak{K} + a]W(q)\,dq,$$

where $\mathfrak{K} = -(\hbar^2/2m)\Delta + V(q)$ is an operator of the Schrödinger type.

Let us choose $A = \beta\hbar^2 = \hbar^2/2kT$. With this choice we obtain

$$E = \int_{\mathfrak{C}}W^*(q)[\mathfrak{K} + a]W(q)\,dq,$$
$$W(q) \in L_2(\mathfrak{C}), \quad W(q) \text{ normalized.} \qquad (18)$$

Thus the quantized values of the field energy are related to the spectrum of $\mathfrak{K}$. We are particularly interested in the case where $\mathfrak{K}$ is self-adjoint and where at least part of its spectrum is discrete.

All the following results can be extended to the continuous part of the spectrum with the usual appropriate change in the notations.

Let $\{e_k\}$ be the discrete spectrum and $\{W_k(q)\}$ the corresponding orthonormal set of functions defined by the eigenvalue problem

$$\mathfrak{K}W_k(q) = e_kW_k(q), \qquad k = 0, 1, 2, \cdots.$$

This set of functions induces an orthonormal basis $\{\Phi_k(p, q)\}$ in the subspace of $L_2(\Gamma)$ whose elements are characterized in (14), simply by taking $\Phi_k(p, q) = C^{\frac{1}{2}} \exp(-\beta \sum_n p_n^2/2m)\, W_k(q)$ for all k. Let $\Phi(p,q) = C^{\frac{1}{2}} \exp[-\beta \sum_n (p_n^2/2m)]W(q)$ be an element in the space spanned by $\{\Phi_k(p, q)\}$. The corresponding series expansion is

$$\Phi(p, q) = C^{\frac{1}{2}} \exp\left(-\beta \sum_n \frac{p_n^2}{2m}\right)\left[\sum_{k=0}^{\infty} c_k W_k(q)\right],$$

$$\text{with} \quad \sum_{k=0}^{\infty} |c_k|^2 = 1.$$

According to (18), the energy of $\Phi(p, q)$ is

$$E = \sum_{k=0}^{\infty} |c_k|^2 (e_k + a).$$

This last result suggests that we should focus our attention on mechanical systems S for which $V(q)$ represents the potential function of some ordinary quantum system Q (for instance, a Coulomb potential if Q is a hydrogen atom or a harmonic potential if Q is a harmonic oscillator). Now the problem is to find explicitly the potential function $F(q)$ which characterizes S once the potential function $V(q)$ which characterizes Q is given. For that purpose we return to the relationship (17).

If in that relationship we introduce the new function $f(q)$ defined by

$$F(q) = -(1/\beta) \log [f(q)], \tag{19}$$

we find

$$[-(\hbar^2/2m)\Delta + V(q)]f(q) = \mathcal{H}f(q) = -af(q).$$

Given a value of a, $f(q)$ is thus obtained by solving a time-independent Schrödinger equation. There is one particular value for a which leads to remarkable results. That is, $a = -e_0$, where e_0 is the eigenvalue of $\mathcal{H}$ corresponding to the ground state $W_0(q)$ of the quantum system Q we are considering (we assume that Q has bound states). For this value of a, we obtain $f(q) = \lambda W_0(q)$, where λ is an arbitrary constant factor. It follows then from (19) that

$$F(q) = -(1/\beta) \log [W_0(q)] - (1/\beta) \log \lambda.$$

The arbitrary additive constant $-(1/\beta) \log \lambda$ has no effect upon the dynamics of S, thus we can simply take

$$F(q) = -(1/\beta) \log [W_0(q)]$$

or

$$W_0(q) = \exp [-\beta F(q)].$$

The interesting feature of the choice $-a = e_0$ can now be made clear. The first vector Φ_0 of the basis $\{\Phi_k;\ k = 0, 1, 2, \cdots\}$ associated with the basis $\{W_k;\ k = 0, 1, 2, \cdots\}$ defined by the eigenstates of the given quantum system is

$$\Phi_0(p, q) = C^{\frac{1}{2}} \exp\left[-\beta \sum_n \frac{p_n^2}{2m}\right] W_0$$

$$= C^{\frac{1}{2}} \exp\left[-\beta \sum_n \frac{p_n^2}{2m}\right] \exp [-\beta F(q)]$$

$$= C^{\frac{1}{2}} \exp (-\beta H).$$

We see that the ground state of the quantum system Q is associated with the statistical state of thermal equilibrium of S. Furthermore, the energy of Φ_0 is

$$E_0 = e_0 + a = e_0 - e_0 = 0.$$

The other statistical states Φ_k, $k > 0$, have positive energy $E_k = e_k - e_0$. E_k can be interpreted as the energy necessary to perturb S from the statistical state of "rest" (time-independent) Φ_0 to the statistical "excited" state Φ_k.

In the Appendix we treat in detail the problem of the hydrogen atom and the problem of the harmonic oscillator.

We wish to conclude this section with two remarks.

(a) The essential reason why we were able to associate a quantum system Q with a mechanical system S having in common some physical properties is that we arrived at a Schrödinger equation. To be precise we only obtained the time-independent part of the Schrödinger equation. As to the time-dependent part, we refer the reader to an earlier report by the same authors,[2] where he could also find a different approach to the entire problem. However, it is our opinion that this specific question as to the bearing of our field equation of motion in phase space on a wave equation of quantum theory should wait until the problem is formulated in a relativistic scheme. It is more natural, then, to compare our field equation of motion with the Klein–Gordon equation since both equations are of the second order with respect to the time. Pending further investigation the discussion of this question can be postponed as to the content of the present work.

(b) The relationship between Q and S is clear from the preceding dicussion as far as statistical states and corresponding energy spectrum are concerned. But this is not so for the momentum ob-

[2] N. Wiener and G. Della Riccia, *Analysis in Function Space*, W. T. Martin and I. Segal, Eds. (Technology Press, Cambridge, Massachusetts, 1964), pp. 3–14.

servable. If we average the "probability amplitude" $\Phi(p, q)$ over the momenta coordinates $(p_1, \cdots, p_N)$ of $\mathcal{S}$, we find the usual probability amplitude $W(q)$ of quantum theory. Only then can we define the momenta of Q by the usual operator $(\hbar/i)\nabla$. It seems natural to treat the classical momenta p of $\mathcal{S}$ as hidden variables for the quantum system Q. But, vice versa, it is not clear at the moment how the observable $(\hbar/i)\nabla$ could be related to a characteristic property of $\mathcal{S}$. In any case, there will be no violation of the uncertainty principle since only the p's, and not $(\hbar/i)\nabla$, are considered to be known at the same time as the q's with the same accuracy.

III. BROWNIAN MOTION AND BORN'S STATISTICAL POSTULATE

Since we are discussing a statistical theory of a mechanical system $\mathcal{S}$, we must define in some sense a probability density function $\rho(p, q; t)$ in phase space. A natural definition, which has the advantage of justifying the concept of "probability amplitude" $\varphi(p, q; t)$ introduced previously, is $\rho(p, q; t) = \varphi\varphi^*$. In quantum theory a similar relationship between probability amplitudes (in configuration space) and probability densities is a direct consequence of Born's statistical postulate. Likewise, we could here use the same postulate to obtain the desired result. But in order to remain in a classical framework, we prefer to derive this result from ordinary probability theory. For this purpose we make use of a method already discussed by Wiener and Siegel[3] in similar circumstances. Since the main theorems have been proved many times, we only give a formal statement of the main results of the theory. For more details we refer the reader to Wiener's original work.[4]

The fundamental notion with which we start is that of a Wiener–Levy stochastic process with stationary independent increments such that

$$[x(t_2, \alpha) - x(t_1, \alpha)]$$

has a Gaussian distribution with mean 0 and variance

$$\text{Var}\,[x(t_2, \alpha) - x(t_1, \alpha)] = |t_2 - t_1|,$$

where t varies in the real line $R = (-\infty, +\infty)$ and the random parameter α takes values in the probability space $\Omega = \{[0, 1], \mathfrak{B}, \text{Lebesgue}\}$. The process is normalized by the condition $x(0, \alpha) = 0$ for all values of α. While t is the time in the case of the Brownian motion, we wish to emphasize that this will not always be the case. In fact, this variable represents for us a space variable. To avoid any confusion with time, we henceforth call the variable s instead of t and write $\{x(s, \alpha)\}$. It can be proved that for almost all values of α, $x(s, \alpha)$ as a function of s is continuous and nondifferentiable. Nevertheless, the derivative $x(ds, \alpha)/ds$ of the Wiener–Levy process can be defined in the sense of the theory of distributions. More specifically, the following stochastic integral,

$$F(\alpha) = \int_{-\infty}^{\infty} f(s)x(ds, \alpha), \quad \text{for all} \quad f(s) \in L_2(R),$$

can be defined in the sense of mean-square convergence. This integral can be generalized in different ways. First s can be a variable in R^n where n is any positive integer. Then one can define a complex Wiener–Levy process increment by

$$X(ds, \alpha) = x(ds, \beta) + ix(ds, \gamma),$$
$$\alpha \in [0, 1], \quad (\beta, \gamma) \in [0, 1] \times [0, 1],$$

where the mapping

$$[0, 1] \times [0, 1] \ni (\beta, \gamma) \to \alpha \in [0, 1]$$

is defined almost everywhere (with respect to Lebesgue measure). Thus we have, in general,

$$F(\alpha) = \int_{R^n} f(s)X(ds, \alpha), \text{ for all } f(s) \in L_2(R^n). \quad (20)$$

An important property of $F(\alpha)$ is that it is a complex Gaussian random variable (defined on the probability space $\Omega(\alpha) = \{[0, 1], \mathfrak{B}, \text{Lebesgue}\}$) with independent real and imaginary parts, mean value 0 and variance

$$\text{Var}\,\{F(\alpha)\}$$
$$= \int_0^1 F^*(\alpha)F(\alpha)\, d\alpha = \int_{R^n} f^*(s)f(s)\, ds = ||f||^2.$$

It is possible to show that, if $f_1(s)$ and $f_2(s)$ are two orthogonal functions belonging to L_2, then the corresponding random variables $F_1(\alpha)$ and $F_2(\alpha)$ are independent.

The fundamental theorem which permits us to apply these results to our problem is the following.

Theorem: Let $f_1(s)$ and $f_2(s)$ be two orthogonal functions belonging to L_2. Let S be the set of values of α such that the following relation holds:

$$\left|\int_{R^n} f_1(s)X(ds, \alpha)\right| \geq \left|\int_{R^n} f_2(s)X(ds, \alpha)\right|.$$

[3] N. Wiener and A. Siegel, Phys. Rev. 91, 1551 (1953).

[4] N. Wiener, *Nonlinear Problems in Random Theory* (Technology Press, Cambridge, Massachusetts, and John Wiley & Sons, Inc., New York, 1958).

Then the measure of S is

$$m(S) = ||f_1||^2/(||f_1||^2 + ||f_2||^2).$$

As a corollary of this theorem, we have the following.

Lemma: Let $\{c_k\}$ be a sequence of complex numbers such that $\sum_k |c_k|^2 < \infty$; let all values of k be divided into two sets I and II which are mutually exclusive. Let $\{f_k(s)\}$ be an orthonormal set of functions of s, $\sum_I c_k f_k(s)$ be the sum over I and $\sum_{II} c_k f_k(s)$ be the sum over II. Then the measure of the set S of values of α for which

$$\left|\sum_I \int c_k f_k(s) X(ds, \alpha)\right| \geq \left|\sum_{II} \int c_k f_k(s) X(ds, \alpha)\right|$$

will be

$$m(S) = \sum_I |c_k|^2/(\sum_I |c_k|^2 + \sum_{II} |c_k|^2).$$

Let us notice that the measure of sets which are contained in the segment [0, 1] has the properties appropriate for a probability. Thus, we can regard $m(S)$ as the probability for the corresponding inequality to be valid.

For our specific problem, the variable s is a phase space variable (p, q) in a $2N$-dimensional Euclidean space $R^{2N} = \Gamma$ and $f(s)$ is a "probability amplitude" function $\varphi(p, q; t) \in L_2$. Let Δ, $\Delta \subset \Gamma$, be an arbitrary measurable set and $\Gamma - \Delta$ the complement of Δ in Γ. We define two functions φ_I and φ_{II} as follows:

$$\varphi_I(p, q; t) = \begin{cases} \varphi(p, q; t), & \text{if } M(p, q) \in \Delta, \\ 0, & \text{if } M \in \Gamma - \Delta, \end{cases}$$

and

$$\varphi_{II}(p, q; t) = \begin{cases} 0, & \text{if } M \in \Delta, \\ \varphi(p, q; t), & \text{if } M \in \Gamma - \Delta. \end{cases}$$

It is clear that φ_I and φ_{II} are two orthogonal functions such that

$$||\varphi_I||^2 = \int_\Delta \varphi\varphi^* \, dp \, dq, \quad ||\varphi_{II}||^2 = \int_{\Gamma-\Delta} \varphi\varphi^* \, dp \, dq,$$

and

$$||\varphi_I||^2 + ||\varphi_{II}||^2 = \int_\Gamma \varphi\varphi^* \, dp \, dq = 1,$$

because φ, as a "probability amplitude," is normalized. [We recall that the norm of φ is time-independent since $\varphi(p, q; t) = U^t\varphi_0(p, q)$, where $U^t = \exp(i\mathfrak{L}t)$ is a unitary transformation.] The state of the mechanical system $\mathfrak{S}$, represented by a point $M(p, q)$ in phase space is a random event. We define this random variable by the following.

Postulate: If the system $\mathfrak{S}$ is in a statistical state represented by $\varphi(p, q; t)$, then its mechanical state $M(p, q)$ at time t lies in the set $\Delta \subset \Gamma$ if and only if

$$\left|\int_\Gamma \varphi_I X(dp \, dq; \alpha)\right| \geq \left|\int_\Gamma \varphi_{II} X(dp \, dq; \alpha)\right|.$$

Due to the fundamental theorem stated above, the probability that this random event occurs is

$$\text{Prob } \{M \in \Delta\} = \frac{||\varphi_I||^2}{||\varphi_I||^2 + ||\varphi_{II}||^2} = \int_\Delta \varphi\varphi^* \, dp \, dq.$$

Since Δ was chosen arbitrarily, it follows at once from the postulate that $\rho(p, q; t) \, dp \, dq$ is the probability that the state of $\mathfrak{S}$ at time t lies in the interval $[(p, q), (p + dp, q + dq)]$.

Thus we have related the probability density function ρ, in the sense of Gibbs, to the "probability amplitude" φ by the usual method of quantum theory.

As a matter of fact, we can randomize in the same fashion any other observable of the system $\mathfrak{S}$. In general, it suffices to be able to find a complete set of orthonormal functions which characterize all the values of the observable. For the "energy" observable of the fields $\Phi(p, q)$ of the form (14), we have found a proper basis $\{\Phi_k\}$ such that if we write

$$\Phi(p, q) = \sum_k c_k \Phi_k(p, q); \qquad \sum_k |c_k|^2 = 1,$$

then the energy of $\Phi(p, q)$ is $E = \sum_k |c_k|^2 E_k$, where E_k is the energy of Φ_k. Now we want to consider the energy as a random variable. We define this random variable by the same postulate as before.

Postulate: If the system $\mathfrak{S}$ is in a statistical state represented by $\Phi(p, q) = \sum_k c_k \Phi_k(p, q)$, then it will be found in the statistical state $\Phi_k(p, q)$ with energy E_k if and only if

$$\left|\int_\Gamma c_k \Phi_k X(dp \, dq; \alpha)\right| \geq \left|\sum_j{}' \int_\Gamma c_j \Phi_j X(dp \, dq; \alpha)\right|,$$

where the summation $\sum_j'$ is extended over all possible values of the index j except $j = k$.

Due to the Lemma stated above, the probability that this event occurs is

$$\text{Prob } \{E = E_k\} = |c_k|^2.$$

It is worth noting that for the "energy" observable the probabilities $\{|c_k|^2\}$ are time independent. This is because the statistical state $\varphi(p, q; t)$ at time t

is obtained from $\Phi(p, q)$ at $t = 0$ by a unitary transformation U^t. Since the probabilities depend only on the norm of various fields and these norms are preserved under U^t, the invariance property obtains.

The above Wiener–Siegel procedure of defining random events by an inequality based on Wiener's stochastic integral (20) can be used in any field theory (classical or not). It allows us to assign a probability to each possible outcome of a given experiment. It also has the remarkable feature that the values of these probabilities coincide with those predicted by Born's statistical postulate when dealing with quantum theory. When this procedure is applied to our field dynamics in the phase space of the mechanical system $\mathcal{S}$ associated with a given quantum system Q (by the methods of Sec. II), it provides an unusual connection between Gibbs statistical mechanics and the ordinary statistical interpretation of quantum mechanics.

IV. CONCLUDING REMARKS

We wish to emphasize that the point of view presented in this work does not require any change in the methods or the principles of ordinary quantum theory. On the contrary, it borrows from quantum theory some of its usual concepts such as the probability amplitude, quantized energy spectrum, probability of occurrence of values of observables, etc., and applies them to a classical system $\mathcal{S}$ which is assumed to be in interaction with its surroundings $\mathcal{O}$ as in Gibbs theory.

The present idea that a quantum dynamics can be traced back to a classical dynamics of a system in the presence of a heat bath appears in several ways sympathetic with the point of view of Bohm and Vigier.[5] These authors have postulated the existence of a "hidden thermostat" at a "subquantum" level. As a result of this hypothesis, they show that the time behavior of a quantum system is the result of a Markov process which leads to the same statistical description, in configuration space, as that predicted by the Schrödinger picture of quantum mechanics. Generally speaking, our method is the same except for the fact that we are dealing with a phase space where the use of a heat bath appears more natural. On the other hand, we were able to make use of a classical system $\mathcal{S}$ whose potential function $F(q)$ does not involve the "quantum potential" which appears in their work as a consequence of their analysis being purely in configuration space. Incidentally, the classical dynamics which they use is the one to which the quantum dynamics is asymptotic when we let Planck's constant $\hbar$ go to 0, whereas we employ a completely different dynamics for the system $\mathcal{S}$.

The physical assumption of the "hidden thermostat" of Bohm and Vigier was used by Louis de Broglie[6] in his recent extension of the theory of the "double solution" which was originated by him as early as 1927. But the most spectacular use of this assumption can be found in another recent contribution by de Broglie.[7] In this work de Broglie established a remarkable correspondence between mechanical and thermodynamic quantities. He considers a periodic mechanical system with frequency ν, in equilibrium with a heat bath with temperature T. Then he writes

$$S = k\frac{A}{h} \quad \text{and} \quad h\nu = kT,$$

where A is the action in the sense of Maupertuis, S is the entropy, k is the Boltzmann's constant, and h the Planck's constant.

With these relations de Broglie established for the first time a correspondance between the principle of least action and the principle of maximum entropy. Let us simply recall that we have found that the potential function of the system is

$$F(q) = -kT \log [W_0^2(q)],$$

which in some sense also ties up a classical dynamics with the thermodynamic concept of temperature. In fact, $F(q)$ can be interpreted as a Helmholtz "free energy"

$$F(q) = -TS$$

if the entropy is $S = k \log [W_0^2(q)]$, which agrees with Gibbs definition of entropy since $W_0^2(q)$ is indeed a probability distribution.

Finally, we would like to point out that according to our discussion the system $\mathcal{S}$ behaves in many respects like a quantum system Q after the appropriate quantities are averaged over the momenta coordinates p of $\mathcal{S}$. In this sense, $\mathcal{S}$ belongs to the "subquantum" level of Bohm and Vigier.

The ultimate validity of our work should be regarded as resting upon experimental results which, in the future, could reveal the physical existence of $\mathcal{S}$. With the present techniques of experimental

[5] D. Bohm and J. P. Vigier, Phys. Rev. 96, 208 (1954).

[6] L. de Broglie, *La Théorie de la mesure en mécanique ondulatoire* (Gauthier-Villars, Paris, 1957); *Non-linear Wave Mechanics, a Causal Interpretation* (Elsevier Publishing Company, Inc., Amsterdam and New York, 1960).

[7] L. de Broglie, *La Thermodynamique de la particule isolée* (Gauthier-Villars, Paris, 1964).

physics, the system S, if it exists, shows its presence to the observer only through those properties with which we are familiar when dealing with usual quantum systems.

APPENDIX

We wish to illustrate the methods discussed in this paper by two specific examples.

A. The Electron in a Hydrogen Atom

The Schrödinger operator related to this problem is

$$\mathcal{H} = -(\hbar^2/2m)\Delta - (e^2/r),$$

where $r = (q_1^2 + q_2^2 + q_3^2)^{\frac{1}{2}}$ and e is the charge of the electron. The ground state $W_0(q)$, defined by

$$\mathcal{H}W_0(q) = e_0 W_0(q),$$

is known to be (except for a normalization factor)

$$W_0(q) = \exp(-r/a_0),$$

where $a_0 = \hbar^2/me^2$ is the radius of the first (circular) Bohr orbit. The lowest eigenvalue of $\mathcal{H}$ is $e_0 = -e^2/2a_0$. The associated mechanical system S is characterized by the Hamiltonian function

$$H = \sum_{n=1}^{3} \frac{p_n^2}{2m} + F(q),$$

where, according to our results, we have

$$F(q) = -kT \log [W_0^2(q)] = 2kT(r/a_0).$$

Note that $F(q)$ is a central potential; however, it is not the Coulomb potential. We thus clearly see on this example the difference between the dynamics of S and the classical dynamics obtained as a limit when Planck's constant h is treated as a small quantity.

B. The Harmonic Oscillator

We treat this problem in one dimension (the case of three dimensions is very similar).

The Schrödinger operator is

$$\mathcal{H} = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial q^2} + \tfrac{1}{2}m\omega^2 q^2.$$

The ground state (except for a normalization factor) is

$$W_0(q) = \exp[-(m\omega/\hbar)(\tfrac{1}{2}q^2)],$$

and the ground energy is

$$e_0 = \tfrac{1}{2}\hbar\omega.$$

The potential function of the associated system S is

$$F(q) = -kT \log [W_0^2(q)] = (mkT\omega/\hbar)q^2.$$

Therefore, the Hamiltonian function of S is

$$H = \frac{p^2}{2m} + \tfrac{1}{2}m\nu^2 q^2,$$

where

$$\nu = (2kT\omega/\hbar)^{\frac{1}{2}}.$$

In this case we find that S is itself a harmonic oscillator but with a resonance frequency ν which is different from ω. Let us solve this problem completely. The statistical states of S are

$$\Phi_k(p, q) = C^{\frac{1}{2}} \exp(-\beta p^2/2m) W_k(q); \quad \beta = 1/2kT, \quad k = 0, 1, 2, \cdots.$$

Here we have

$$\mathcal{H}W_k(q) = e_k W_k(q); \qquad e_k = (k + \tfrac{1}{2})\hbar\omega$$

and

$$W_k(q) = \exp\left(-\frac{m\omega}{\hbar}\frac{q^2}{2}\right) h_k\left[\left(\frac{m\omega}{\hbar}\right)^{\frac{1}{2}} q\right],$$

where h_k is the Hermite polynomial of order k. Thus, using $\nu = (2kT\omega/\hbar)^{\frac{1}{2}}$, we obtain

$$\Phi_k(p, q) = C^{\frac{1}{2}} \exp(-\beta H) h_k[(\beta m)^{\frac{1}{2}}\nu q].$$

Since the general solution of the equation of motion of the classical harmonic oscillator is known to be

$$q(t) = q \cos(\nu t) + (p/m\nu) \sin(\nu t),$$

we find that

$$\varphi_k(p, q; t) = \Phi_k(T_p^{-t}, T_q^{-t}) = C^{\frac{1}{2}} \exp(-\beta H) h_k\{(\beta m)^{\frac{1}{2}}\nu[q \cos \nu t - (p/m\nu) \sin \nu t]\}.$$

This is a periodic function of the time which contains terms of frequency $0, \nu, \cdots, k\nu$. However, according to our definition of the field energy, the energy of φ_k is $E_k = e_k - e_0 = k\omega$. It is the usual quantum spectrum of energy, except for a shift in the energy scale equal to $-\frac{1}{2}\hbar\omega$.

Anisotropic Linear Magnetic Chain

JACQUES DES CLOIZEAUX AND MICHEL GAUDIN
Service de Physique Théorique, Saclay, France
(Received 24 December 1965)

The ground-state and the spin-wave states of the Hamiltonian,

$$H = \sum_j (S_j^x S_{j+1}^x + S_j^y S_{j+1}^y + \rho S_j^z S_{j+1}^z),$$

are studied for all values of ρ, and analytical expressions are given for their energies. On the other hand, by using a canonical transformation which changes $H(\rho)$ into $-H(-\rho)$, the states of highest energy can also be obtained. The ground state is ferromagnetic for $\rho \leq -1$ and antiferromagnetic for $\rho \geq -1$. For $\rho = \pm 1$, the energy has singularities, but it remains continuous. For $\rho = 1$, all its derivatives are also continuous. In the range $-1 \leq \rho \leq 1$, the spin-wave states of given momentum are degenerate but for $\rho \geq 1$; this degeneracy is removed, and an energy gap $G(\rho)$ appears.

I. INTRODUCTION

LINEAR magnetic chains are very interesting from a theoretical point of view because they provide simple nontrivial models of many-body systems. In several cases, these models can be treated exactly, and therefore they can be used to test more general approximate theories. For this purpose, we try here to obtain new exact results in analytical form concerning these magnetic chains by generalizing methods which have proved useful in the past.

Incidentally, we must note that the problem is also important from an experimental point of view. Linear antiferromagnetic chains exist in many crystals. In general, at very low temperatures these crystals have an antiferromagnetic lattice. However, for temperatures higher than the Néel point, the interactions between the chains which constitute the lattice become incoherent and, on an average, negligible. Then, the chains can be considered as isolated. Actually, specific heat measurements[1] of chain magnetism have been made at low temperature, and they agree qualitatively with the theory.

For reasons of mathematical convenience, we restrict ourselves to the study of very long chains of spins with nearest-neighbor interaction. The Hamiltonian of the system will be

$$H = \sum_{j=1}^{N} (S_j^x S_{j+1}^x + S_j^y S_{j+1}^y + \rho S_j^z S_{j+1}^z). \quad (1)$$

The spin operator of components S_j^x, S_j^y, S_j^z is associated with the site j and corresponds to local states of spin $\frac{1}{2}$. Moreover, for reasons of simplicity, it is assumed that the sites form a ring and that the site of order $(N + 1)$ coincides with the site of order 1.

The completely isotropic problem ($\rho = 1$) was investigated a long time ago by Bethe[2] and Hulthen.[3] In particular, Bethe gave a classification of all the eigenstates of the isotropic Hamiltonian by means of sets of integers (quantum numbers), and he showed that the problem of finding the eigenvalues and the eigenstates is equivalent to the resolution of a series of coupled equations. Moreover, by developing this method, he succeeded in calculating exactly the energy of the ground state which, in this case, is antiferromagnetic. Later on, Pearson and one of the authors[4] could also determine the first excited states of the isotropic Hamiltonian and calculate exactly their energies, i.e., the antiferromagnetic spin wave spectrum. Unfortunately, the statistics of these excitations does not appear very clearly. On the other hand, Orbach[5] tried to extend Bethe's treatment to the anisotorpic case ($0 < \rho < 1$), and Walker[6] gave an analytical expression of the ground state energy, in the case $\rho \geq 1$.

The same line of approach is used here. The nature of the ground state and of the first excited states is investigated for all values of ρ. Simple analytic expressions are given for the energies of these states. In Sec. II the general formalism is introduced, Sec. III is devoted to a study of the ground state, and Sec. IV to a study of the spin-wave states.

[1] T. Haseda and A. R. Miedema, Physica **27**, 1102 (1961); A. R. Miedema, H. Van Kampen, T. Haseda, and W. J. Huiskamp, Physica **28**, 119 (1962).

[2] H. Bethe, Z. Physik 71, 205 (1931).
[3] L. Hulthen, Arkiv. Mat. Astron. Fys. **26A**, 1 (1938).
[4] J. des Cloizeaux and J. J. Pearson, Phys. Rev. **128**, 2131 (1962).
[5] R. Orbach, Phys. Rev. **112**, 309 (1959).
[6] L. R. Walker, Phys. Rev. **116**, 1089 (1959).

II. GENERAL PROPERTIES OF THE EIGENSTATES OF THE HAMILTONIAN NOTATIONS

We now establish the formalism which enables one to determine the eigenstates and the eigenvalues of H, by generalizing the methods of Bethe and Orbach. First, we introduce the operators S_i^+ and S_i^-:

$$S_i^+ = S_i^x + iS_i^y, \qquad S_i^- = S_i^x - iS_i^y. \tag{2}$$

With these notations, H can be written

$$H(\rho) = \sum_{i=1}^{N} [\tfrac{1}{2}(S_i^+ S_{i+1}^- + S_i^- S_{i+1}^+) + \rho S_i^z S_{i+1}^z]. \tag{3}$$

The number N of atoms which is contained in the ring is assumed to be even. In this way, the spins of all the eigenstates of H will be integers. This assumption is not really restrictive, since the ultimate aim of our studies is the determination of macroscopic quantities which, in the limit of large N, are certainly independent of the parity of N.

The Hamiltonian H commutes with the component S^z of the total spin:

$$S^z = \sum_{i=1}^{N} S_i^z. \tag{4}$$

Therefore, it is possible to diagonalize simultaneously H and S^z; for each eigenstate $|\omega\rangle$ of H, we write

$$H\,|\omega\rangle = E\,|\omega\rangle, \tag{5}$$

$$S^z\,|\omega\rangle = M\,|\omega\rangle. \tag{6}$$

Before defining explicitly these states, we note that $H(\rho)$ and $H(-\rho)$ are related by a canonical transformation. This transformation U is the following:

$$U = \exp\left(i\pi \sum_{i=1}^{i=N} jS_i^z\right). \tag{7}$$

It commutes with S^z and conserves the cyclic boundary conditions. If $\frac{1}{2}N$ is even, it commutes also with the translation operator and thus conserves the total momentum; but, if $\frac{1}{2}N$ is odd, it transforms a state of total momentum K into a state of total momentum $K + \pi$. Moreover, we have

$$UH(\rho)U^{-1} = -H(-\rho). \tag{8}$$

Therefore, in principle, we could restrict ourselves to the study of the case $\rho > 0$. However, in order to follow the evolution of each level as a function of ρ, it is interesting to consider variations of ρ from $-\infty$ to $+\infty$. In particular, if we know the ground state of $H(\rho)$ for each value of ρ, by using transformation U, we obtain, for each value of ρ, the state of maximum energy of $H(\rho)$.

By applying this remark, it is easy to guess, for each value of ρ, the nature of the spin component M which must be associated with the ground state. We know that, for $\rho = 1$, the Hamiltonian H commutes with the total spin S. For the ground state, we have $S = 0$, $M = 0$, and this state is unique. On the contrary, the states of maximum energy correspond to $S = \frac{1}{2}N$ and are degenerate. In particular, there is a state with $M = 0$ and a state with $M = \frac{1}{2}N$ which have the same energy. For $\rho \neq 1$, the degeneracy is removed. A first-order perturbation calculation shows that, for $\rho = 1 - 0$, the state of maximum energy has a spin component $M = 0$ and that, for $\rho = 1 + 0$, the state of maximum energy is degenerate with $M = \pm\frac{1}{2}N$, a situation which remains true when $\rho \to +\infty$. Thus, by using the transformation U, we are lead to the conclusion that, for $\rho > -1$, the ground state should be unique with the value $M = 0$. But for $\rho < -1$, the ground state should be doubly degenerate with $M = \pm\frac{1}{2}N$. This view is supported by exact calculation of short chains and also by the subsequent studies of the ground-state energy.

Moreover, the study of the corresponding classical system obtained by replacing the operators $\mathbf{S}_i$ by ordinary vectors of length $\frac{1}{2}$ leads to very similar results. Thus, the state of maximum energy can be obtained from the ground state by using the transformation U which is a rotation of π, in the plane xy, for the spin vectors of odd indices. The ground state can be built immediately.

For $\rho > 1$, all the spin vectors are parallel to the oz axis but with alternating directions; in this case, $H(\rho) = -\frac{1}{4}N\rho$ and $S^z = 0$. For $-1 < \rho < 1$, the spin vectors lie in the xy plane. We assume, for instance, that there are parallel to the ox axis with alternating directions, thus $H(\rho) = \frac{1}{4}N$, with $S^z = 0$. For $\rho < -1$, all the spin vector are parallel to the oz axis and point in the same direction; thus $H(\rho) = -\frac{1}{4}N\rho$ and $S^z = \pm\frac{1}{2}N$. It is interesting to note that the same general features with singularities at $\rho = 1$ and $\rho = -1$ are found in the quantum case which is considered now.

Let $|F\rangle$ be the ferromagnetic state corresponding to $M = \frac{1}{2}N$. In this state, all spins are parallel to each other; we have, of course,

$$S_i^+\,|F\rangle = 0. \tag{9}$$

On the other hand, we may write

$$H(\rho)\,|F\rangle = E_F(\rho)\,|F\rangle, \tag{10}$$

with

$$E_F(\rho) = \tfrac{1}{4}\rho. \tag{11}$$

By flipping r spins in $|F\rangle$, we can now build any eigenstate $|\Omega\rangle$ of S^z:

$$|\Omega\rangle = \sum_{n_1<\cdots<n_r} a(n_1, \cdots, n_r) S^-_{n_1} \cdots S^-_{n_r} |F\rangle. \tag{12}$$

The corresponding eigenvalue of S^z is $M = \frac{1}{2}(N-2r)$. We must now determine the coefficients $a(n_1, \cdots, n_r)$ in order to build an eigenstate of H, of eigenvalue E:

$$H\,|\Omega\rangle = E\,|\Omega\rangle. \tag{13}$$

We define ϵ by setting

$$\epsilon = (E - E_F)/N, \tag{14}$$

and with this notation, the equations which must be satisfied by the coefficients $a(n_1, \cdots, n_r)$ can be written

$$2N\epsilon a(n_1, \cdots, n_r) = \sum [a(n'_1, \cdots, n'_r) - \rho a(n_1, \cdots, n_r)], \tag{15}$$

with $n_1 < \cdots < n_r$. In this formula, a term $a(n'_1, \cdots, n'_r)$ is obtained by changing one number n of $a(n_1, \cdots, n_r)$ of one unit. Summation must be made over all the possible $a(n'_1, \cdots n'_r)$ which can be obtained from $a(n_1, \cdots, n_r)$.

Like Bethe, we can try to express the coefficients $a(n_1, \cdots, n_r)$ in terms of r wavenumbers k_α (with $\alpha = 1, \cdots, r$) and of phases $\psi_{\alpha\beta}$ associated with each couple of wavenumbers k_α and k_β. Thus, we put

$$a(n_1, \cdots n_r) = \sum_P \exp\left(i \sum_\alpha k_{P\alpha} n_\alpha + \tfrac{1}{2} i \sum_{\alpha<\beta} \psi_{P\alpha, P\beta}\right). \tag{16}$$

By definition, P is any permutation of the numbers $(1, \cdots, r)$ and $P\alpha$ is the result obtained by permutation of the number α. We note now that this formula enables us to define coefficients $a(n_1, \cdots, n_r)$ for $n_1 \leq \cdots \leq n_r$ and that these coefficients satisfy the equation term by term:

$$\begin{aligned} 2N\epsilon a(n_1, \cdots, n_r) = \sum_\alpha [&a(n_1, \cdots, n_\alpha + 1, \cdots, n_r) \\ &+ a(n_1, \cdots, n_\alpha - 1, \cdots, n_r) \\ &- 2\rho a(n_1, \cdots, n_\alpha, \cdots, n_r)] \\ &\text{with} \quad n_1 < \cdots < n_r, \end{aligned} \tag{17}$$

where ϵ takes the value

$$\epsilon = N^{-1} \sum_\alpha (\cos k_\alpha - \rho). \tag{18}$$

Equation (17) is very similar to Eq. (15) but contains a few more terms; in fact, in the right-hand side of equation (17), there are coefficients $a(n_1, \cdots, n_r)$ for which some indices are equal. However, the phases $\psi_{\alpha\beta}$ which, until now, were arbitrary, can be chosen in such a way that the supplementary terms which appear in Eq. (17) just cancel out. By means of this trick, the eigenvalue equation (15) and Eq. (17) become identical. On the other hand, the equations which determine the phases $\psi_{\alpha\beta}$ can be written

$$a(\cdots, n_\alpha + 1, n_\alpha + 1, \cdots) + a(\cdots, n_\alpha, n_\alpha, \cdots) - 2\rho a(\cdots, n_\alpha, n_\alpha + 1, \cdots) = 0. \tag{19}$$

These equalities are equivalent to the conditions

$$\cot(\tfrac{1}{2}\psi_{\alpha\beta}) = \rho\left[\frac{\cot(\frac{1}{2}k_\alpha) - \cot(\frac{1}{2}k_\beta)}{(1-\rho)\cot(\frac{1}{2}k_\alpha)\cot(\frac{1}{2}k_\beta) - (1+\rho)}\right]. \tag{20}$$

On the other hand, as the spin system is cyclic, the coefficients $a(n_1, \cdots, n_r)$ must satisfy boundary conditions which are

$$a(n_1, n_2, \cdots, n_r) = a(n_2, \cdots, n_r, n_1 + N). \tag{21}$$

These equations imply the following relations:

$$Nk_\alpha = 2\pi\lambda_\alpha + \sum_\beta \psi_{\alpha\beta}, \tag{22}$$

where each λ_α is an integer. Incidentally, the total wave vector K is directly related to these numbers λ since we have

$$K = \sum k_\alpha = 2\pi N^{-1} \sum \lambda_\alpha. \tag{23}$$

Thus, an eigenstate $|\Omega\rangle$ is completely determined by a series of integers λ_α, and Bethe has shown (for $\rho = 1$) that all the eigenstates of H can be obtained in this way. The wave vectors k_α and the phases $\psi_{\alpha\beta}$ are solutions of the coupled Eqs. (20) and (22); thus by using these values of k_α, we get the corresponding value of ϵ given by Eq. (18) and the value of the energy E which is related to ϵ by Eq. (14). All these equations are far from simple, and it is remarkable that in the limit of large N many states may be calculated exactly.

In particular, Eq. (20) looks rather formidable, and we can write it in a less forbidding way by using auxiliary variables (i.e., θ_α for $|\rho| < 1$ and φ_α for $|\rho| > 1$) which are going to play an important role in the following sections.

For $-1 < \rho < 1$, we set

$$\rho = \cos\Theta, \qquad 0 < \Theta < \pi, \tag{24}$$

$$\tanh(\tfrac{1}{2}\theta_\alpha) = \tan(\tfrac{1}{2}\Theta)\cot(\tfrac{1}{2}k_\alpha), \tag{25}$$

$$\cot(\tfrac{1}{2}\psi_{\alpha\beta}) = \cot\Theta \tanh(\tfrac{1}{2}\theta_\alpha - \tfrac{1}{2}\theta_\beta),$$
$$-\pi < \psi_{\alpha\beta} < \pi. \quad (26)$$

For $\rho > 1$, we get

$$\Theta = i\Phi, \qquad \theta_\alpha = -i\varphi_\alpha, \quad (27)$$

which gives

$$\rho = \cosh\Phi, \qquad 0 < \Phi < +\infty, \quad (28)$$
$$\tan(\tfrac{1}{2}\varphi_\alpha) = \tanh(\tfrac{1}{2}\Phi)\cot(\tfrac{1}{2}k_\alpha), \quad (29)$$
$$\cot(\tfrac{1}{2}\psi_{\alpha\beta}) = \coth\Phi \tan(\tfrac{1}{2}\varphi_\alpha - \tfrac{1}{2}\varphi_\beta). \quad (30)$$

At last, for $\rho < -1$, we could write

$$\Theta = \pi - i\tilde{\Phi}, \qquad \theta_\alpha = -i\varphi_\alpha, \quad (31)$$

which gives

$$\rho = -\cosh\tilde{\Phi}, \qquad 0 < \Phi < +\infty, \quad (32)$$
$$\tan\tfrac{1}{2}\varphi_\alpha = \coth(\tfrac{1}{2}\tilde{\Phi})\cot(\tfrac{1}{2}k_\alpha), \quad (33)$$
$$\cot(\tfrac{1}{2}\psi_{\alpha\beta}) = -\coth\tilde{\Phi}\tan(\tfrac{1}{2}\varphi_\alpha - \tfrac{1}{2}\varphi_\beta). \quad (34)$$

These equations appear now in a form which is more tractable in the limit of large N, and the following sections are devoted to the solution of these systems of equations in a few special cases.

III. DETERMINATION OF THE GROUND STATE AND CALCULATION OF THE GROUND-STATE ENERGY FOR $N \to \infty$

The eigenstates of H corresponding to the spin component $M = 0$ are determined by a series of $\frac{1}{2}N$ integers λ_α, and Bethe has shown that, for $\rho = 1$, the antiferromagnetic ground state can be obtained by choosing for λ_α the series of numbers $(1, 3, \cdots, N-1)$. Moreover, in this case, it is assumed that the wavenumbers k_α and the phases $\psi_{\alpha\beta}$ satisfy the conditions

$$0 < k_\alpha < 2\pi, \quad (35)$$
$$-\pi < \psi_{\alpha\beta} < \pi. \quad (36)$$

These results of Bethe can be generalized for all values of ρ, by proceeding by steps.

A. $0 \leq \rho \leq 1$

As we may see, for $0 \leq \rho \leq 1$, by using for λ_α the same series $(1, 3, \cdots, N-1)$, it is possible to calculate wavenumbers k_α and phases $\psi_{\alpha\beta}$ which satisfy the conditions (35) and (36). In this way, we determine a state which must be the ground state of the system since, as shown, the energy of the state is always smaller than the ferromagnetic energy and is an analytic function of ρ in the domain $0 \leq \rho < 1$.

Thus, we set

$$\lambda_\alpha = 2\alpha - 1, \qquad \alpha = (1, \cdots, \tfrac{1}{2}N). \quad (37)$$

The total moment K of the state is according to Eq. (23):

$$K = \tfrac{1}{2}N\pi. \quad (38)$$

In the limit of large N, α can be replaced by a continuous parameter x defined by

$$x = 2\pi(2\alpha - 1)/N. \quad (39)$$

When $N \to \infty$, k_α becomes a continuous function $k(x)$, and in the same way, $\psi_{\alpha\beta}$ becomes a function $\psi(x, x')$. Consequently, the energy of the corresponding state can be expressed by the integral

$$\epsilon = \frac{1}{4\pi}\int_0^{2\pi} [\cos k(x) - \rho]\, dx. \quad (40)$$

On the other hand, with our choice of λ, the boundary equation (22) becomes for $N \to \infty$

$$k(x) = x + \frac{1}{4\pi}\int_0^{2\pi} \psi(x, x')\, dx'. \quad (41)$$

We now set

$$\rho = \cos\Theta, \qquad 0 < \Theta < \pi, \quad (42)$$

and we use the notations given at the end of Sec. II. The auxiliary variable θ_α becomes a continuous function $\theta(x)$. We may assume, and later on it is easy to verify, that $k(x)$ is an increasing function of x; on the contrary, according to (25), $\theta(x)$ must be a decreasing function of x, and it is assumed that $\theta(x)$ decreases from $+\infty$ to $-\infty$ when x increases from 0 to 2π. At the same time, the function $k(x)$ increases from Θ to $(\pi - \Theta)$, as can be checked by inspection of Eq. (25) or of Eqs. (26) and (22). On the other hand, the function $\psi(x, x')$ is nearly everywhere a decreasing function of $\theta(x)$, but it jumps from $-\pi$ to $+\pi$ when $\theta(x)$ increases from $\theta(x') - 0$ to $\theta(x') + 0$.

Now we choose the auxiliary variable θ as a new variable instead of x, and we set

$$dx/d\theta = -f(\theta). \quad (43)$$

In fact, it turns out that all the important quantities can be expressed in a rather simple way in terms of $f(\theta)$, which is the solution of an integral equation which is now established. Let us differentiate Eq. (41) with respect to θ. With our definition of $f(\theta)$, we obtain

$$\frac{dk}{d\theta} = -\tfrac{1}{2}f(\theta) + (4\pi)^{-1}\int_{-\infty}^{+\infty}\left[\frac{\partial\psi(x, x')}{\partial\theta}\right] f(\theta')\, d\theta'. \quad (44)$$

In this equation, we denote by $\partial\psi(x, x')/\partial\theta$ the continuous part of the derivative of $\psi(x, x')$. The variables $dk/d\theta$ and $\partial\psi(x, x')/\partial\theta$ are obtained immediately from the following equations [see Eqs. (24) and (25)]:

$$\tanh(\tfrac{1}{2}\theta) = \tan(\tfrac{1}{2}\Theta)\cot(\tfrac{1}{2}k), \tag{45}$$

$$\cot[\tfrac{1}{2}\psi(x, x')] = \cot\Theta\tan(\tfrac{1}{2}\theta - \tfrac{1}{2}\theta'). \tag{46}$$

Therefore, we get

$$dk/d\theta = -\sin\Theta[\cosh\theta - \cos\Theta]^{-1}, \tag{47}$$

$$\partial\psi(x, x')/\partial\theta = -\sin 2\Theta[\cosh(\theta - \theta') - \cos 2\Theta]^{-1}. \tag{48}$$

By substitution of these expressions in Eq. (44), we obtain the integral equation which gives $f(\theta)$:

$$f(\theta) + \frac{\sin 2\Theta}{2\pi}\int_{-\infty}^{+\infty}\frac{f(\theta')}{\cosh(\theta - \theta') - \cos 2\Theta}\,d\theta' = \frac{2\sin\Theta}{\cosh\theta - \cos\Theta}. \tag{49}$$

This equation can be solved by setting

$$f(\theta) = \int_{-\infty}^{+\infty} e^{i\omega\theta}a(\omega)\,d\omega. \tag{50}$$

By applying the method of residues, the following identity can be easily proved:

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{\exp(i\omega\theta')}{\cosh(\theta - \theta') - \cos\Theta}\,d\Theta = \frac{\exp(i\omega\theta)\sinh\omega(\pi - \Theta)}{\sin\Theta\sinh\omega\pi}. \tag{51}$$

With the help of this formula, the value of $a(\omega)$ can be easily deduced from the integral equation (49), where $f(\theta)$ has been replaced by its development (50),

$$a(\omega) = 1/\cosh\omega\Theta. \tag{52}$$

By substitution of this expression in Eq. (50), we get

$$f(\theta) = \pi/\Theta\cosh(\pi\theta/2\Theta). \tag{53}$$

Now, let us calculate ϵ. First, ϵ must be expressed in terms of $f(\theta)$. We start from Eq. (40), use θ as a new variable, and express $\cos k(x)$ in terms of θ by means of Eq. (45). We obtain

$$\epsilon = \frac{1}{4\pi}\int_0^{2\pi}[\cos k(x) - \cos\Theta]\,dx = \frac{\sin^2\Theta}{4\pi}\int_{-\infty}^{+\infty}\frac{f(\theta)}{\cosh\theta - \cos\Theta}\,d\theta. \tag{54}$$

This integral can be written in terms of the Fourier transform $a(\omega)$ of $f(\theta)$. By taking Eqs. (50) and (51) into account, we get

$$\epsilon = -\tfrac{1}{2}\sin\Theta\int_{-\infty}^{+\infty}\frac{\sinh\omega(\pi - \Theta)}{\sinh\omega\pi}\,a(\omega)\,d\omega. \tag{55}$$

Finally, by replacing $a(\omega)$ by the value calculated above, the energy ϵ becomes

$$\epsilon = \epsilon_1(\rho) \equiv -\sin\Theta\int_0^{+\infty}\left(1 - \frac{\tanh\omega\Theta}{\tanh\omega\pi}\right)d\omega. \tag{56}$$

Exact integrations of this expression can be made for $\rho = 0$ and $\rho = 1$. Actually, for $\rho = 0$ and $\Theta = \frac{1}{2}\pi$, we have

$$\epsilon_1(0) = -\frac{1}{2}\int_0^{\infty}[\cosh^2(\tfrac{1}{2}\omega\pi)]^{-1}\,d\omega = -1/\pi. \tag{57}$$

For $\rho = 1$, $\epsilon(1)$ is calculated by setting $\omega\Theta = x$ in Eq. (54) and taking the limit for $\Theta \to 0$.

$$\epsilon_1(1) = -\int_0^{\infty}(1 - \tanh x)\,dx = -\log 2. \tag{58}$$

This is the result of Bethe and Hülthen.

B. $1 \leq \rho < 0$.

Let us examine now the behavior of the numbers k_α and $\psi_{\alpha\beta}$ when $\rho \to +0$, in order to see how it is possible to extrapolate the results of Sec. IIIA, for $-1 \leq \rho < 0$. In this case, we have again

$$\rho = \cos\Theta, \quad \text{with} \quad \tfrac{1}{2}\pi < \Theta < \pi. \tag{59}$$

When $\rho \to +0$, according to Eq. (20), we have

$$\cot(\tfrac{1}{2}\psi_{\alpha\beta}) \to +0\,\mathrm{sgn}(k_\alpha - k_\beta), \tag{60}$$

and therefore

$$\psi_{\alpha\beta} \to -\pi\,\mathrm{sgn}(k_\alpha - k_\beta); \tag{61}$$

consequently, by setting, as we did in Sec. A,

$$\lambda_\alpha = 2\alpha - 1, \qquad \alpha = (1, \cdots, \tfrac{1}{2}N), \tag{62}$$

we obtain for k_α the following values which are deduced from Eqs. (22), (61), and (62):

$$k_\alpha = \tfrac{1}{2}\pi + (2\alpha - 1)\pi/N. \tag{63}$$

Thus, we see that for $\rho = 0$, k varies from $\frac{1}{2}\pi$ to $\frac{3}{2}\pi$ as could be expected *a priori*.

On the other hand, when $\rho \to -0$, the behavior of the phases is different, and we have

$$\cot(\tfrac{1}{2}\psi_{\alpha\beta}) \to -0\,\mathrm{sgn}(k_\alpha - k_\beta), \tag{64}$$

and therefore

$$\psi_{\alpha\beta} \to \pi\,\mathrm{sgn}(k_\alpha - k_\beta). \tag{65}$$

This discontinuity in the behavior of the phases

$\psi_{\alpha\beta}$ is purely formal. The antiferromagnetic state obtained for $\rho = +0$ can be found again for $\rho = -0$ by changing the set of integers λ_α and introducing a new set λ'_α. This operation must leave the values of k_α invariant. It is easy to verify that this condition can be fulfilled, by setting

$$\lambda'_\alpha = \tfrac{1}{2}N \tag{66}$$

for all values of α.

By using now these new values λ'_α, we can extend all the results of Sec. IIA to the domain $-1 \leq \rho \leq 0$. In particular, the limit $N \to \infty$ is considered. As previously, we set

$$x = 2\pi(2\alpha - 1)/N. \tag{67}$$

Again, k_α becomes a continuous and increasing function $k(x)$, which varies from Θ to $(\pi - \Theta)$, and in the same way, $\psi_{\alpha\beta}$ becomes a function $\psi(x, x')$. The boundary equation (22) must now be written

$$k(x) = \pi + (4\pi)^{-1} \int_0^{2\pi} \psi(x, x')\, dx', \tag{68}$$

and therefore differs from Eq. (41), but the other basic formulas (42), (45), and (46) remain unchanged. As previously, we introduce the auxiliary variable $\theta(x)$ which becomes, as before, a continuous and decreasing function of x which varies from $+\infty$ to $-\infty$ when x goes from 0 to 2π. On the contrary, now the function $\psi(x, x')$ is nearly everywhere an increasing function of $\theta(x)$, but it jumps from $+\pi$ to $-\pi$, when $\theta(x)$ increases from $\theta(x') - 0$ to $\theta(x') + 0$.

As in Sec. IIIA, we now set

$$dx/d\theta = -f(\theta) \tag{69}$$

and differentiate Eq. (68) with respect to θ. We obtain

$$\frac{dk}{d\theta} = -\tfrac{1}{2}f(\theta) + (4\pi)^{-1} \int_{-\infty}^{+\infty} \left[\frac{\partial\psi(x, x')}{\partial\theta}\right] f(\theta')\, d\theta', \tag{70}$$

where $[\partial\psi(x, x')/\partial\theta]$ represents the continuous part of the derivative of $\psi(x, x')$. This equation coincides with Eq. (44). Therefore the calculation of $f(\theta)$ can be performed exactly as in Sec. IIIA.

Equations (53), (52), and (56) which give the values of $f(\theta)$, $a(\omega)$, and ϵ remain valid in the whole range $0 \leq \Theta \leq \pi$, i.e., in the domain $-1 \leq \rho \leq 1$.

In particular, for $\rho = -1$, $\Theta = \pi$, we have according to (56)

$$\epsilon_1(-1) = 0. \tag{71}$$

This result is not surprising. In fact, by means of the unitary operator U introduced in Sec. II [Eq. (17)], it is possible to transform the antiferromagnetic state corresponding to $\rho = -1$ into the state of maximum energy corresponding to $\rho = 1$, i.e., the ferromagnetic state of quantum number $M = 0$. Therefore, we have

$$E_{AF}(-1) = -E_F(1) = E_F(-1), \tag{72}$$

since the transformation U leaves invariant the ferromagnetic state of quantum number $M = \frac{1}{2}N$. Thus, the result (71) follows immediately from the definition of ϵ [Eq. (14)] and from Eq. (72).

A straightforward calculation shows also that

$$\epsilon'_1(-1) = -\tfrac{1}{4}, \tag{73}$$

and thus, we get the result

$$[dE_{AF}(\rho)/d\rho]_{\rho=-1} = 0, \tag{74}$$

which can be obtained directly by applying first-order perturbation theory to the antiferromagnetic state of quantum number $M = 0$ for $\rho = 1$.

The fact that $\epsilon_1(\rho)$ vanishes for $\rho = 1$ shows that this value is a critical value, and this point is discussed more completely in Sec. IIID. However, we may note now that this limit is characterized by the fact that, for $\rho = -1$, all the values of k_α become equal. In fact, for $\Theta = \pi$, we have according to (25)

$$\cot(\tfrac{1}{2}k_\alpha) = 0, \tag{75}$$

which can be written

$$k_\alpha = \pi \quad \text{or} \quad k(x) = \pi. \tag{76}$$

This relation can also be obtained directly by comparing Eqs. (18) and (71).

Finally, we note that $\epsilon_1(\rho)$ remains completely analytic in the domain $-1 < \rho < 1$.

C. $\rho \geq 1$

In order to describe the ground state for $\rho \geq 1$, we may use the same set of values of λ_α as in the domain $0 \leq \rho \leq 1$, and fundamentally, the calculation of ϵ is performed as in Sec. A. However, now we must use the relation

$$\rho = \cosh\Phi \tag{77}$$

and take φ_α as an auxiliary variable.

The variable x is still defined by Eq. (39) and when $N \to \infty$, φ_α becomes a continuous function $\varphi(x)$. As in the case described in the preceding section, $k(x)$ is an increasing function of x; this function increases from 0 to 2π when x goes from

0 to 2π. For $0 < x < 2\pi$, we may also assume $-\pi < \varphi(x) < \pi$. In this case, $\varphi(x)$ is a decreasing function of x. Finally, the function $\psi(x, x')$ is nearly everywhere a decreasing function of $\varphi(x)$, but it jumps from $-\pi$ to $+\pi$ when $\varphi(x)$ increases from $\varphi(x') - 0$ to $\varphi(x') + 0$.

Now, we can choose φ as the main variable, and we set

$$dx/d\varphi = -g(\varphi). \tag{78}$$

Equation (41) relating $k(x)$ to $\psi(x, x')$ remains completely valid. Differentiation of this equation leads to the integral equation which determines $g(\varphi)$. We get

$$\frac{dk}{d\varphi} = -\tfrac{1}{2}g(\varphi) + (4\pi)^{-1}\int_{-\pi}^{+\pi}\left[\frac{\partial\psi(x, x')}{\partial\varphi}\right]g(\varphi')\,d\varphi'. \tag{79}$$

The continuous quantities $dk/d\varphi$ and $\partial\psi(x, x')/\partial\varphi$ are calculated immediately by using the following equations which have been established in Sec. II:

$$\tan(\tfrac{1}{2}\varphi) = \tanh(\tfrac{1}{2}\Phi)\cot(\tfrac{1}{2}k), \tag{80}$$

$$\cot[\tfrac{1}{2}\psi(x, x')] = \coth\Phi\tan(\tfrac{1}{2}\varphi - \tfrac{1}{2}\varphi'). \tag{81}$$

These relations give

$$dk/d\varphi = -\sinh\Phi(\cosh\Phi - \cos\varphi)^{-1}, \tag{82}$$

$$\partial\psi(x, x')/\partial\varphi = -\sinh 2\Phi[\cosh 2\Phi - \cos(\varphi - \varphi')]^{-1}. \tag{83}$$

The integral equation which determines $g(\varphi)$ is directly obtained by substitution of these expressions in Eq. (79):

$$g(\varphi) + \frac{\sinh 2\Phi}{2\pi}\int_{-\pi}^{+\pi}\frac{g(\varphi')}{\cosh 2\Phi - \cos(\varphi - \varphi')}\,d\varphi' = \frac{2\sinh\Phi}{\cosh\Phi - \cos\varphi}. \tag{84}$$

The solution $g(\varphi)$ is periodical and, consequently, can be expanded in Fourier series:

$$g(\varphi) = \sum_{-\infty}^{+\infty} a_n e^{in\varphi} \qquad (n \text{ integer}). \tag{85}$$

On the other hand, by applying the method of residues, the following identity can be easily proved:

$$\frac{1}{2\pi}\int_{-\pi}^{+\pi}\frac{\exp(in\varphi')}{\cosh\Phi - \cos(\varphi - \varphi')}\,d\varphi' = \frac{\exp(in\varphi - |n|\,\Phi)}{\sinh\Phi}. \tag{86}$$

Now by substituting the expansion of $g(\varphi)$ in Eq. (84) and by using the preceding identity, we derive easily the value of a_n,

$$a_n = 1/\cosh n\Phi. \tag{87}$$

Now in Eq. (85), we can replace a_n by this value, and by comparing with Eqs. (52) and (53), we see immediately that $g(\varphi)$ can be written

$$g(\varphi) = \sum_{-\infty}^{+\infty}\frac{\exp(in\varphi)}{\cosh n\varphi} = \sum_{-\infty}^{+\infty}\frac{\pi}{\Phi\cosh[\pi(\varphi + 2\pi n)/2\Phi]}. \tag{88}$$

This function is meromorphic and has two periods $\omega' = 2\pi$ and $\omega'' = 4i\Phi$; therefore, it is an elliptic function containing two poles in each cell. In Jacobi's notation, it is the function $2dn(\varphi)$ (with $K = \pi$ and $K' = \Phi$).

Let us now calculate the energy ϵ. First, we express ϵ in terms of $g(\varphi)$ by choosing φ as the main variable and by expressing $\cos k(x)$ in terms of φ by means of Eq. (80). After a few simplifications, we obtain

$$\epsilon = (4\pi)^{-1}\int_0^{2\pi}[\cos k(x) - \cosh\Phi]\,dx = -(4\pi)^{-1} \times \sinh^2\Phi\int_{-\pi}^{\pi}(\cosh\Phi - \cos\varphi)^{-1}g(\varphi)\,d\varphi. \tag{89}$$

This expression becomes simpler if we use Fourier transforms. By using Eqs. (85) and (87), we finally get

$$\epsilon = -\tfrac{1}{2}\sinh\Phi\sum_{-\infty}^{+\infty}a_n e^{-|n|\Phi}. \tag{90}$$

By replacing a_n by its value, we are lead to the following result:

$$\epsilon = \epsilon_2(\rho) \equiv -\sinh\Phi\left[\sum_{n=1}^{\infty}(1 - \tanh n\Phi) + \tfrac{1}{2}\right], \tag{91}$$

which is fundamentally the same as the result of Walker.

When $\Phi \to 0$, the sum can be replaced by an integral and $\epsilon_2(\rho)$ has the limit

$$\epsilon_2(1) = -\log 2, \tag{92}$$

which coincides with $\epsilon_1(1)$. On the other hand, it is not difficult to show that the states, obtained when $\rho \to 1 \pm 0$ by using respectively the methods of Sec. IIIA and IIIC, are identical. Thus for $\rho = 1$, the function $E_{AF}(\rho)$ is continuous but not analytic.

On the other hand, when $\Phi \to \infty$, $\epsilon_2(\rho) \simeq -\tfrac{1}{2}\rho$,

a result which was expected, since, for $\rho \to \infty$, the antiferromagnetic ground state is a state in which the spins point alternatively upward and downward.

D. $\rho \leq -1$

For $\rho \leq -1$, the ground state is the ferromagnetic state ($M = \pm\frac{1}{2}N$). Moreover, the antiferromagnetic state ($M = 0$) which we considered for $\rho \geq 1$ in the previous sections cannot be continued in an analytical way for $\rho < 1$. In fact, it is easy to show that, for $\rho \leq -1$, we always have

$$\epsilon(\rho) = 0 \tag{93}$$

for the antiferromagnetic state of lowest energy. For instance, let us consider the state

$$|U\rangle = S_1^- S_2^- \cdots S_{N/2}^- |F\rangle. \tag{94}$$

We see immediately that

$$\bar{E}(\rho) = \langle U| H |U\rangle = (N - 4)(\tfrac{1}{4}\rho) = E_F - \rho. \tag{95}$$

On the other hand,

$$\bar{E} > E_{AF} > E_F, \tag{96}$$

since E_{AF} is the energy of the ground state associated with the subset of states of spin $S^z = 0$ and E_F is the energy of the absolute ground state (for all values of S_z). Consequently,

$$\lim_{N\to\infty} (E_{AF} - E_F)/N = 0, \tag{97}$$

which implies Eq. (93).

E. Summary and Comments

Now, let us review briefly the discussions of the preceding sections. For $\rho \geq -1$, the ground state is antiferromagnetic ($M = 0$). For $-1 < \rho < 1$, the energy E_{AF} is an analytical function of ρ and the corresponding value $\epsilon_1(\rho)$ of ϵ is given by the integral (56). For $\rho > 1$, the energy E_{AF} is also an analytical function of ρ, and the corresponding value $\epsilon_2(\rho)$ of ϵ is given by the sum (91). The functions $\epsilon_1(\rho)$ and $\epsilon_2(\rho)$ are distinct from each other and both have an essential singularity at the point $\rho = 1$. As we show in Appendix A, they can be continued analytically everywhere in the complex plane, but on the real axis, these functions have cuts. The cuts are given respectively by the conditions $\rho \leq -1$ and $\rho \geq 1$ for $\epsilon_1(\rho)$ and $-1 \leq \rho \leq 1$ for $\epsilon_2(\rho)$. The functions $\epsilon_1(\rho)$ and $\epsilon_2(\rho)$ are also related to each other. For $\Im\rho > 0$, it is consistent to assume the conditions

$$\Theta = -i\Phi, \qquad 0 < \Re\Theta < \pi, \tag{98}$$

and with these assumptions, we can write

$$\epsilon_1(\rho) = \epsilon_2(\rho) - 2i\pi \sum_{m=1}^{\infty} \frac{1}{1 + \exp[\pi^2(2m-1)/\Phi]}, \tag{99}$$

as we show in Appendix A.

The case $\Im\rho < 0$ is obtained by taking the complex conjugate of all the equations. Thus for ρ real larger than one, we get the strange result

$$\epsilon_2(\rho) = \tfrac{1}{2}[\epsilon_1(\rho + i0) + \epsilon_1(\rho - i0)]. \tag{100}$$

On the other hand, we can calculate the derivatives of the energy on both sides of the singularity $\rho = 1$. In Appendix B we show that all the derivatives exist and are continuous for $\rho = 1$; more explicitly, we derive the equations

$$\epsilon_1^{(n)}(1 - 0) = \epsilon_2^{(n)}(1 + 0), \tag{101}$$

which are valid for all values of the integer n. This result explains why Bonner and Fisher[7] were unable to see, by machine calculations, the singularity at $\rho = 1$ predicted by Walker.[6]

Some light can be cast on the nature of the singularity occurring at the point $\rho = 1$, by considering the domain of variation of the wave vectors k_α and of the phases $\psi_{\alpha\beta}$. For $\rho = \cos\Theta$, $0 < \Theta < \pi$, we have

$$\Theta \leq k < 2\pi - \Theta, \qquad -\Theta \leq \psi < \Theta. \tag{102}$$

For $\rho = \cosh\Phi$, $0 < \Phi < \infty$, we have

$$0 \leq k < 2\pi, \qquad -\tfrac{1}{2}\pi < \psi < \tfrac{1}{2}\pi. \tag{103}$$

At last, for $\rho \leq 1$, the ground state is ferromagnetic ($M = \pm\frac{1}{2}N$), and we have always $\epsilon(\rho) = 0$. Finally, the variations of all the energies are plotted in Fig. 1.

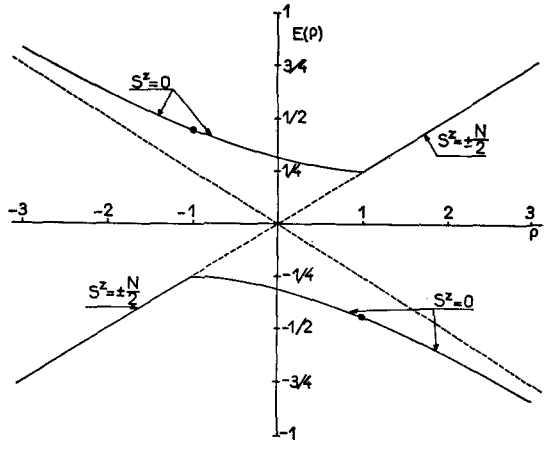


FIG 1

FIG. 1. Curves giving the energy of the ground state and the energy of the state of maximum energy, in terms of ρ. The dots for $\rho = \pm 1$ indicate the singular branch point of the two energy curves.

[7] J. C. Bonner and M. E. Fisher, Phys. Rev. 135A, 640 (1964).

IV. EXCITATIONS OF THE ANTIFERROMAGNETIC GROUND-STATE ($\rho \geq 1$) SPIN WAVE SPECTRUM ($N \to \infty$)

When $\rho = 1$, the total spin is a good quantum number, and in this case, the total spin of the ground state is zero. On the contrary, the total spin of the first excited states is equal to one. These states can be taken as eigenstates of S^z and of the translation operator T. Then, they are characterized by the quantum number M (with $M = 1, 0, -1$) and the total wavenumber K,

$$K = \tfrac{1}{2}N\pi + q, \tag{104}$$

where $\frac{1}{2}N\pi$ is the total wavenumber of the ground state and q the wavenumber of the spin wave. In a previous study, des Cloizeaux and Pearson[4] determined the quantum number λ_α which is associated with these low-lying states and calculated the corresponding spin-wave spectrum. Our aim in this section is to generalize these results in the domain $\rho \geq -1$.

For $\rho \neq 1$, the operators S^z and T commute with H, and consequently, M and q remain good quantum numbers. Therefore, for all values of $\rho \geq 1$, it is possible to determine by continuity eigenstates $|\rho, M, q\rangle$ which for $\rho = 1$ coincide with the states studied previously. For these states, we have

$$H(\rho)\,|\rho, M, q\rangle = E(\rho, M, q)\,|\rho, M, q\rangle, \tag{105}$$

$$S^z\,|\rho, M, q\rangle = M\,|\rho, M, q\rangle, \tag{106}$$

$$T\,|\rho, M, q\rangle = (-)^{N/2}\,e^{iq}\,|\rho, M, q\rangle. \tag{107}$$

In the limit $N \to \infty$, we calculate the excitation energy $\eta(\rho, M, q)$, which can be defined

$$\eta(\rho, M, q) = \lim_{N\to\infty} [E(\rho, M, q) - E_{AF}(\rho)]$$

$$= N[\epsilon(\rho, M, q) - \epsilon_{AF}(\rho)], \tag{108}$$

where $\epsilon_{AF}(\rho)$ is the value of ϵ which corresponds to the ground state and which is calculated in Sec. III. The function $\eta(\rho, M, q)$ is, of course, an even function of q, and for reasons of convenience is calculated for $-\pi < q < 0$.

The reader must realize that the derivation which is presented here is not completely rigorous but rather heuristic. This comes from the fact that the energy of an eigenstate of H is always of the order of N, whereas an excitation energy is of the order of one. For this reason, in order to calculate in a rigorous way, the difference (108) which defines $\eta(\rho, M, q)$, it is necessary to calculate the eigenstates energies with great care in the limit of large N. Calculations of this type are not difficult in principle, but involve very lengthy calculations. For this reason, we apply here a method which is much simpler and which, in spite of its lack of mathematical rigor, seems quite consistent and reliable; in fact, it agrees completely with the exact calculations which have been made for short chains.[4]

A. $0 < \rho \leq 1$

Let us first determine the excitation energy $\eta(\rho, 0, q)$ which can be calculated more easily than $\eta(0, \pm 1, q)$. As was shown in a previous paper,[4] the integers λ_α which determine the state $|\rho, 0, q\rangle$ are, for $-\pi < q < 0$,

$$\lambda_\alpha = 2\alpha - 2, \qquad 1 \leq \alpha \leq n, \tag{109}$$

$$\lambda_\alpha = 2\alpha - 1, \qquad n < \alpha \leq \tfrac{1}{2}N, \tag{110}$$

where n is an integer which is related to the spin wave vector q by

$$q = -2\pi n/N. \tag{111}$$

As in Sec. III, we introduce a continuous variable x by setting

$$x = 2\pi(2\alpha - 1)/N. \tag{112}$$

Now, when N becomes large, the function $\lambda(x)$ can be written approximately in the form

$$\lambda(x) = x + 2\pi N^{-1} \times \sum_{p=-\infty}^{+\infty} [S(x - 2\,|q| + 2\pi p) - S(x + 2\pi p)], \tag{113}$$

where p is an integer and $S(x)$ the step function

$$S(x) \equiv \tfrac{1}{2}(1 + x/|x|). \tag{114}$$

The reader may wonder why an infinite series of terms appears in the right-hand side of Eq. (113) since until now we always assumed that x belongs to the interval $(0, 2\pi)$. It seems as if the only important term is $S(x - 2\,|q|)$ and that all the other terms can be omitted (as they were in Ref. 4 for the case $\rho = 1$). However, it must be realized that this restriction $0 < x < 2\pi$ is unnecessary; the variables $\theta(x)$ or $\psi(x, x')$ must be periodic functions (of periods 2π) of x and $k(x)$. Thus, the series of small terms which appear in Eq. (113) must be introduced for reasons of consistency. Moreover, they become very important in the case $\rho \geq 1$.

The boundary equation (22) can now be written

$$k(x) = x + 2\pi N^{-1} \sum_{p=-\infty}^{+\infty} [S(x - 2\,|q| + 2p\pi) - S(x + 2p\pi)] + (4\pi)^{-1} \int_0^{2\pi} \psi(x, x')\,dx'. \tag{115}$$

As in Sec. IIIA, we express everything in terms of the auxiliary variable θ, and again set

$$dx/d\theta = -f(\theta). \tag{116}$$

Let us differentiate Eq. (115) with respect to θ. By taking Eqs. (47) and (48) into account, we obtain

$$f(\theta) + \frac{\sin 2\Theta}{2\pi}\int_{-\infty}^{+\infty}\frac{f(\theta')}{\cosh(\theta-\theta') - \cos 2\Theta}\,d\theta' = \frac{2\sin\Theta}{\cosh\theta - \cos\Theta} + \frac{4\pi}{N}[\delta(\theta-\infty) - \delta(\theta-\theta_0)], \tag{117}$$

where θ_0 corresponds to the value $x_0 = 2|q|$. This equation differs from Eq. (49) by adjunction of the last term. Therefore we set

$$f(\theta) = f_0(\theta) + \Delta f(\theta), \tag{118}$$

where $f_0(\theta)$ is the solution of Eq. (49) and is calculated in Sec. III. The term $\Delta f(\theta)$ is solution of the following equation:

$$\Delta f(\theta) + \frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{\Delta f(\theta')}{\cosh(\theta-\theta') - \cos 2\Theta}\,d\theta' = \frac{4\pi}{N}[\delta(\theta-\infty) - \delta(\theta-\theta_0)]. \tag{119}$$

In agreement with Eq. (50), we set

$$\Delta f(\theta) = \int_{-\infty}^{+\infty} e^{i\omega\theta}\,\Delta a(\omega), \tag{120}$$

and with the help of identity (51), we obtain

$$\Delta a(\omega) = -N^{-1}e^{-i\omega\theta_0} \times \sinh\omega\pi/\cosh\omega\Theta\sinh\omega(\pi-\Theta) \tag{121}$$

but

$$\Delta a(\omega) = 0 \quad \text{for} \quad \omega = 0. \tag{122}$$

Thus, the anomalous term $\delta(\theta-\infty)$ has just the effect of canceling $\Delta a(\omega)$ for $\omega = 0$. Therefore, in the present case its contribution is completely negligible.

Now, we can calculate $\eta(\rho, 0, q)$. Equation (54) remains valid, and therefore by using definition (122), we can express $\eta(\rho, 0, q)$ in terms of $\Delta\alpha(\omega)$:

$$\eta(\rho, 0, q) = \tfrac{1}{2}N\sin\Theta \times \int_{-\infty}^{+\infty}\frac{\sinh\omega(\pi-\Theta)}{\sinh\omega\pi}\,\Delta a(\omega)\,d\omega. \tag{123}$$

By replacing $\Delta a(\omega)$, by its value (121), we obtain

$$\eta(\rho, 0, q) = \tfrac{1}{2}\sin\Theta\int_{-\infty}^{+\infty}\frac{\exp(-i\omega\theta_0)}{\cosh\omega\Theta}\,d\omega = \frac{\pi}{2\Theta\cosh(\pi\theta_0/2\Theta)}. \tag{124}$$

The parameter θ_0 is a function of q, which has to be calculated now. As was noted before, θ_0 corresponds to the value $x_0 = -2q$. On the other hand, when x goes from 0 to 2π, θ varies from $+\infty$ to $-\infty$. Consequently, by integration of Eq. (116), we get

$$2q = -x_0 = -\int_{\theta_0}^{+\infty} f(\theta)\,d\theta. \tag{125}$$

In this equation, we may replace the exact value $f(\theta)$ by the function $f_0(\theta)$ which differs from it by an infinitesimal amount. This function $f_0(\theta)$ is given by Eq. (53). Thus, we may write

$$q = -\tfrac{1}{2}\pi\int_{\theta_0}^{\infty}[\Theta\cosh(\pi\theta/2\Theta)]^{-1}\,d\theta = -2\arctan[\exp(-\pi\theta_0/2\Theta)], \tag{126}$$

or more simply

$$\cosh(\pi\theta_0/2\Theta) = -1/\sin q. \tag{127}$$

Finally, we substitute this expression in Eq. (124) which gives $\eta(\rho, 0, q)$. The value of $\eta(\rho, 0, q)$ which has been computed in this way is valid for $-\pi \leq q < 0$ only, but by taking into account the parity of $\eta(\rho, 0, q)$ with respect to q, we obtain immediately, for $-\pi \leq \rho \leq \pi$, the general expression

$$\eta(\rho, 0, q) = [\pi\sin\Theta/2\Theta]\,|\sin q|. \tag{128}$$

For $\rho = 1, \Theta = 0$, we find the same result as in Ref. 4.

Now, let us calculate $\eta(\rho, \pm 1, q)$. By omitting the value $\lambda_1 = 0$ from the set of λ_α which corresponds to $|\rho, 0, q\rangle$, we obtain a new set which determines the state $|\rho, 1, q\rangle$. In the Eq. (22) for $\alpha \neq 1$, the term $\psi_{\alpha 0}$ disappears. But in the ground state, we have

$$\psi(x, 0) = -\psi(0, x) = \Theta, \tag{129}$$

and in the state $|\rho, 1, q\rangle$, the function $\psi(x, x')$ has nearly the same value as in the ground state. Consequently, by transforming the boundary equation (22), we obtain an equation which differs of (115) by the adjunction of a term Θ/N; this comes from the fact that $\psi_{\alpha 0}$ is absent from the new Eq. (22),

$$k(x) = x + 2\pi N^{-1} \times \sum_p [S(x - 2|q| + 2\pi p) - S(x + 2\pi p)] - \Theta N^{-1} + (4\pi)^{-1}\int_0^{2\pi}\psi(x, x')\,dx'. \tag{130}$$

By differentiation, this constant is eliminated, and therefore we find the same solution $f(\theta)$ as above.

Let us now consider the energy $N\epsilon(\rho, 1, q)$. This energy can be written as a sum [see Eq. (18)]

which contains $(\frac{1}{2}N - 1)$ terms, since the term corresponding to $\alpha = 1$ is absent. For the ground state, the corresponding term is

$$\cos k_0 - \rho \simeq 0, \tag{131}$$

in agreement with Eq. (102). Therefore, all the sums can be replaced by integrals as usual and the absence of the term $\alpha = 1$ in the sum giving $\epsilon(\rho, 1, q)$ does not lead to the appearence of any extra terms. On the other hand, the functions $f(\theta)$ which correspond respectively to $|\rho, 0, q\rangle$ and $|\rho, 1, q\rangle$ are identical in the limit $N \to \infty$, and therefore we must conclude that the energies corresponding to these states are identical. Thus, we have finally

$$\eta(\rho, 1, q) = \eta(\rho, -1, q) = \eta(\rho, 0, q) = (\pi \sin \Theta/2\Theta)\, |\sin q|. \tag{132}$$

B. $-1 \leq \rho \leq 0$

For $\rho = 0$, a formal discontinuity appears in the behavior of the phases which are associated with $|\rho, M, q\rangle$, in complete analogy with the case of the ground state. But, again for $\rho = 0$, the wave vectors k_α must be continuous. Consequently, in order to determine the states $|\rho, M, q\rangle$ in the range $-1 \leq \rho < 0$, it is necessary to use a new set of integers λ'_α. The treatment is the same as in Sec. IIIB, and it is easy to show that the values λ'_α which must be associated with $|\rho, 0, q\rangle$ are the following:

$$\lambda'_\alpha = \tfrac{1}{2}N - 1, \qquad 1 \leq \alpha \leq n, \tag{133}$$

$$\lambda'_\alpha = \tfrac{1}{2}N, \qquad n < \alpha < \tfrac{1}{2}N, \tag{134}$$

where n is an integer which is related to the wave vector q by

$$q = -2\pi n/N. \tag{135}$$

In the limit $N \to \infty$, the boundary equation can be written now

$$k(x) = \pi + 2\pi N^{-1} \sum_p [S(x - 2\,|q| + 2\pi p) - S(x + 2\pi p)] + (4\pi)^{-1} \int_0^{2\pi} \psi(x, x')\, dx'. \tag{136}$$

By differentiation of this expression, with respect to θ, we again find just Eq. (117). The remaining calculation can be done exactly as in Sec. IIIA; consequently, formula (128), which gives $\eta(\rho, 0, q)$ in terms of ρ, can be generalized to the whole range $-1 \leq \rho \leq 1$.

For the state $|\rho, 1, q\rangle$, the situation is quite similar. In order to ensure the continuity of the vectors k_α which correspond to this state for $-1 \leq \rho < 0$, we must give to λ'_α the following values:

$$\lambda'_\alpha = \tfrac{1}{2}N, \qquad 1 < \alpha \leq n, \tag{137}$$

$$\lambda_\alpha = \tfrac{1}{2}N + 1, \qquad n < \alpha \leq \tfrac{1}{2}N. \tag{138}$$

The relations between q and x remains the same [Eq. (127)] as can be easily verified. The new boundary equation can be written for $N \to \infty$,

$$k(x) = \pi + 2\pi N^{-1} + 2\pi N^{-1} \times \sum_{p=-\infty}^{+\infty} [S(x - 2\,|q| + 2\pi p) - S(x + 2\pi p)] + (4\pi)^{-1} \int_0^{2\pi} \psi(x, x')\, dx'. \tag{139}$$

By differentiation of this equation, we find again Eq. (116). By reasoning as in Sec. IVA, it is now easy to show that $\eta(\rho, 0, q)$ remains to equal to $\eta(\rho, \pm 1, q)$ in the whole domain $-1 \leq \rho \leq 1$.

C. $\rho \geq 1$

In this case, the states $|\rho, M, q\rangle$ are determined by the quantum numbers λ_α as in the case $0 < \rho \leq 1$ (see Sec. IVA), but of course, we must use the auxiliary variable φ instead of θ.

Let us calculate first the excitation energy $\eta(\rho, 0, q)$. Equations (113) and (115) remain valid. As in Sec. IIIC, we set

$$dx/d\varphi = -g(\varphi). \tag{140}$$

By differentiation of Eq. (115) and by taking Eqs. (82) and (83) into account, we obtain the integral equation

$$g(\varphi) + \frac{\sinh 2\Phi}{2\pi} \int_{-\pi}^{+\pi} \frac{g(\varphi')}{\cosh 2\Phi - \cos(\varphi - \varphi')}\, d\varphi' = \frac{2 \sinh \Phi}{\cosh \Phi - \cos \varphi} + \frac{4\pi}{N} \sum_p [\delta(\varphi - \pi + 2\pi p) - \delta(\varphi - \varphi_0 + 2\pi p)]. \tag{141}$$

This equation differs from Eq. (84) by adjunction of the last terms, but its solution remains a periodical function of φ. Let us now set

$$g(\varphi) = g_0(\varphi) + \Delta g(\varphi), \tag{142}$$

where $g_0(\varphi)$ is the solution of Eq. (84) for the ground state. The function $\Delta g(\varphi)$ satisfies the following equation:

$$\Delta g(\varphi) + \frac{\sinh 2\Phi}{2\pi} \int_{-\pi}^{+\pi} \frac{\Delta g(\varphi')}{\cosh 2\Phi - \cos(\varphi - \varphi')}\, d\varphi' = \frac{4\pi}{N} \sum_p [\delta(\varphi - \pi + 2\pi p) - \delta(\varphi - \varphi_0 + 2\pi p)]. \tag{143}$$

In agreement with (85), we set

$$\Delta g(\varphi) = \sum_{-\infty}^{+\infty} \Delta a_n e^{in\varphi} \qquad (n \text{ integer}). \tag{144}$$

By substitution of this expression into the integral equation (143), we obtain, with the help of identity (86),

$$\Delta a_n = -\frac{2}{N}\left[\frac{e^{-in\varphi_0} - e^{-in\pi}}{1 + e^{-2|n|\Phi}}\right]. \tag{145}$$

Formula (90) remains valid, and by using definition (108), we get

$$\eta(\rho, 0, q) = -\tfrac{1}{2}N \sinh \Phi \sum_{-\infty}^{+\infty} \Delta a_n e^{-|n|\Phi}. \tag{146}$$

Now Δa_n can be replaced by its value, and finally by comparison with Eq. (88), we obtain the result

$$\eta(\rho, 0, q) = \tfrac{1}{2}\sinh \Phi \sum_{-\infty}^{+\infty}\left(\frac{e^{-n\varphi_0} - e^{-in\pi}}{\cosh n\Phi}\right)$$
$$= \tfrac{1}{2}\sinh \Phi[g_0(\varphi_0) - g_0(\pi)], \tag{147}$$

or more explicitly

$$\eta(\rho, 0, q) = \sinh \Phi\left[\sum_{n=1}^{\infty} \frac{\cos n\varphi_0 - (-)^n}{\cosh n\Phi}\right]$$
$$= \frac{\pi \sinh \Phi}{2\Phi} \sum_{n=-\infty}^{+\infty}\left[\frac{1}{\cosh[\pi(\varphi_0 + 2\pi n)/2\Phi]}\right.$$
$$\left. - \frac{1}{\cosh[(2n+1)\pi^2/2\Phi]}\right]. \tag{148}$$

For all values of φ_0 or q, we have $\eta(\rho, 0, q) \geq 0$. This result is a trivial consequence of the relation $g_0(\varphi) \geq g_0(\pi)$, which can be proved as follows. As we noted in Sec. III, $g_0(\varphi)$ is an elliptic function (of periods 2π and $4i\Phi$) which has two simple poles in each cell. Consequently, $g'(\varphi)$ is also an elliptic function which contains two double poles in each cell. Therefore, according to a well-known theorem, $g_0'(\varphi)$ has exactly four zeros in each cell. For reasons of symmetry, these zeros are the equivalents points $\varphi = 0$, $\varphi = \pi$, $\varphi = 2i\Phi$ and $\varphi = \pi + 2i\Phi$. Therefore, on the real axis, the periodic function $g_0'(\varphi)$ vanishes only twice in the interval 2π. As the point $\varphi = 0$ is the maximum of $g_0(\varphi)$ as it appears immediately from the Fourier representation of $g_0(\varphi)$, the point $\varphi = \pi$ must be the absolute minimum of the function $g_0(\varphi)$, for all real values of φ.

Now, we must express φ_0 in terms of q. We saw that φ_0 corresponds to the value $x_0 = -2|q|$. On the other hand, when x goes from 0 to 2π, φ varies from $+\pi$ to $-\pi$. Consequently, by integration of Eq. (140), we get

$$2q = -x_0 = -\int_{\varphi_0}^{\pi} g(\varphi)\, d\varphi. \tag{149}$$

In this equation we may replace the exact value $g(\varphi)$ by $g_0(\varphi)$ which is nearly identical to it and is given by Eq. (88),

$$q = -\tfrac{1}{2}\pi + \tfrac{1}{2}\int_0^{\varphi_0} g(\varphi)\, d\varphi, \qquad -\pi < q < 0, \tag{150}$$

$$|q| = \tfrac{1}{2}\pi - \tfrac{1}{2}\varphi_0 - \sum_{n=1}^{\infty} \frac{\sin n\varphi_0}{n \cosh n\phi}$$
$$= +\tfrac{1}{2}\pi - 2\sum_{-\infty}^{+\infty}\left(\arctan\left\{\exp\left[\frac{\pi(\varphi_0 + 2\pi n)}{2\Phi}\right]\right\}\right.$$
$$\left. - \arctan\left[\exp\left(\frac{\pi^2 n}{\Phi}\right)\right]\right). \tag{151}$$

This formula is rather complicated, but we verify immediately that $q = 0$ for $\varphi_0 = \pi$, $q = -\frac{1}{2}\pi$ for $\varphi_0 = 0$, and $q = -\pi$ for $\varphi_0 = -\pi$.

As $\eta(\rho, 0, q)$ is an even function of q, it is defined by Eqs. (148) and (151) in the domain $-\pi \geq q + \pi$. Again, we verify that we have

$$\eta(\rho, 0, q \pm \pi) = \eta(\rho, 0, q), \tag{152}$$

as in the case $-1 < \rho < +1$. This function is plotted in Fig. 2 for a few values of ρ.

When $\Phi \to 0$, $\rho \to 1 + 0$, it is easy to see that we again get the limit obtained for $\rho \to 1 - 0$, i.e.,

$$\eta(1, 0, q) = \tfrac{1}{2}\pi\,|\sin q|. \tag{153}$$

On the contrary, when $\rho \to \infty$, the function $\eta(\rho, 0, q)$ reaches the limit

$$\eta(\infty, 0, q) = 1 + \cos \varphi_0 = 1 - \cos 2q. \tag{154}$$

Let us evaluate now the excitation energy $\eta(\rho, 1, q)$. By omitting the value $\lambda_1 = 0$ from the set of λ_α which is associated with $|\rho, 0, q\rangle$, we determine

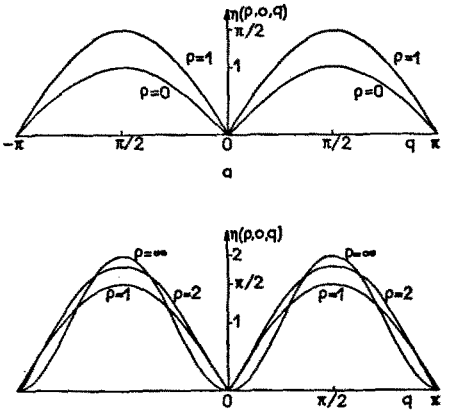


FIG. 2. (a) Spin-wave spectrum $\eta(\rho, o, q)$ for $-1 \leq \rho \leq 1$. (b) Spin-wave spectrum $\eta(\rho, o, q)$ for $\rho \geq 1$.

the state $|\rho, 1, q\rangle$. In this case, in Eq. (22) for $\alpha \neq 1$, the term $\psi_{\alpha 0}$ disappears. But in the antiferromagnetic state, we have $[\varphi(x)]_{x=0} = 0$, and consequently, we may write

$$\tan [\tfrac{1}{2}\psi(x, 0)] = -\tanh \Phi \tan [\tfrac{1}{2}\varphi(x)]. \quad (155)$$

By taking into account the absence of $\psi_{\alpha 0}$ in Eq. (22), the boundary equation corresponding to $|\rho, 1, q\rangle$ becomes, in the limit $N \to \infty$,

$$k(x) = x + 2\pi N^{-1} \times \sum_{p=-\infty}^{+\infty} [S(x - 2|q| + 2\pi p) - S(x + 2\pi p)] + (4\pi)^{-1} \int_0^{2\pi} \psi(x, x')\, dx' + 2N^{-1} \arctan \{\tanh \Phi \tan [\tfrac{1}{2}\varphi(x)]\}. \quad (156)$$

By differentiation with respect to φ, and by using the same notations as the preceding sections, we obtain the following integral equation:

$$\Delta g(\varphi) + \frac{\sinh 2\Phi}{2\pi} \int_{-\pi}^{+\pi} \frac{\Delta g(\varphi')}{\cosh 2\Phi - \cos(\varphi - \varphi')}\, d\varphi' = -\frac{4\pi}{N}\delta(\varphi - \varphi_0) + \frac{2}{N}\frac{\sinh 2\Phi}{\cosh 2\Phi + \cos\varphi}. \quad (157)$$

Finally, with the help of identity (86), we get

$$\Delta a_n = -\frac{2}{N}\left[\frac{e^{-in\varphi_0} - (-)^n e^{-2|n|\Phi}}{1 + e^{-2|n|\Theta}}\right] = -\frac{2}{N}\left[\frac{e^{-in\varphi_0} + (-)^n}{1 + e^{-2|n|\Phi}} - (-)^n\right]. \quad (158)$$

On the other hand, the excitation energy $\eta(\rho, 1, q)$ is equal to

$$\eta(\rho, 1, q) = \rho - 1 - \tfrac{1}{2}N \sinh \Phi \sum_{-\infty}^{+\infty} e^{-|n|\varphi}\, \Delta a_n. \quad (159)$$

The term $(\rho - 1)$ comes from the absence of the term $(\cos k_1 - \rho)$ in the sum (18) which gives $N\epsilon(\rho, 1, q)$ and from the fact that, in the ground state, we have $k_1 \simeq 0$ as can be easily verified. Consequently, after replacing Δa_n by its value, we get

$$\eta(\rho, 1, q) = +\tfrac{1}{2}\sinh \Phi \sum_{-\infty}^{+\infty} \frac{e^{-in\varphi_0} + (-)^n}{\cosh n\Phi} = \tfrac{1}{2}\sinh \Phi [g_0(\varphi) + g_0(\pi)] \quad (160)$$

according to definition (88). Finally, by comparing the expressions giving $\eta(\rho, 1, q)$ and $\eta(\rho, 0, q)$, we find

$$G(\rho) = \eta(\rho, 1, q) - \eta(\rho, 0, q) = \sinh \Phi\, g_0(\pi) = \sinh \Phi \sum_{-\infty}^{+\infty} \frac{(-)^n}{\cosh n\Phi} = \frac{\pi \sinh \Phi}{\Phi} \sum_{-\infty}^{+\infty} \frac{1}{\cosh [(2n + 1)\pi^2/2\Phi]}. \quad (161)$$

The fact that $G(\rho)$ is independent of q is quite remarkable. It seems that $G(\rho)$ plays the role of an energy gap appearing at the point $\rho = 1$. This gap increases very slowly at the beginning,

$$G(\rho) \simeq 4\pi \exp \{-\pi^2/2[2(\rho - 1)]^{\frac{1}{2}}\}, \quad 0 < \rho - 1 \ll 1; \quad (162)$$

but when ρ goes to infinity,

$$G(\rho) \simeq \rho - 2. \quad (163)$$

Figure 3 shows the variations of $G(\rho)$ for all values of ρ. Such a result is not surprising since the terms of the Hamiltonian H which contain ρ become dominant, when ρ is large. In this case, the ground state is made of alternating spins upwards and downwards. By reversing one spin, we create an excitation of energy ρ, i.e., equivalent to the gap $G(\rho)$. The fact that no gap appears in the expression of $\eta(\rho, 0, q)$ suggests that the state $|\rho, 0, q\rangle$ might be a collective state. For such states, in the limit $N \to \infty$, the part of H, which is independent of ρ, can never be neglected however great ρ may be. This would explain that no energy gap appears in $\eta(\rho, 0, q)$.

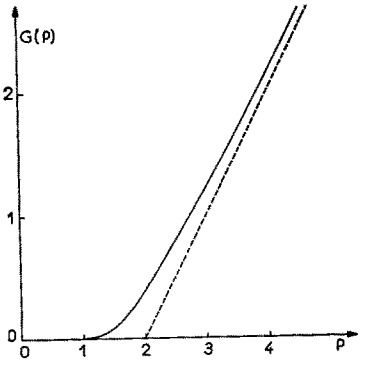


FIG. 3. Variation of $G(\rho)$ with respect to ρ, for $\rho \geq 1$.

These considerations help also to solve the following puzzle. Anderson's theory[8] of antiferromagnetism (for an Heisenberg Hamiltonian $\rho = 1$) leads only to two degenerate spin-wave states for each value of q, whereas the exact theory gives three wave states (with $S^z = 1, 0, -1$). Now, we can understand this strange behavior if we assume that

[8] P. W. Anderson, Phys. Rev. **86**, 694 (1952).

Anderson's theory is nearly correct for $\rho \geq 1$. In this way, Anderson's spin wave states should correspond to the exact spin wave states of spin components $S^z = \pm 1$. On the other hand, the spin wave state of spin component $S^z = 0$ is thought to be a collective state; accordingly, it is not surprising that Anderson's crude theory could not account for it.

V. SUMMARY AND CONCLUSIONS

The main results of this study can now be summarized briefly.

(1) In the limit of large chains, there exists a symmetry which transforms $H(\rho)$ into $-H(-\rho)$. Thus the properties of the states of maximum energy can be deduced from the properties of the ground state and conversely.

(2) The ground state is ferromagnetic ($S^z = \pm\frac{1}{2}N$) for $\rho \leq -1$ and antiferromagnetic for $\rho \geq -1 (S^z = 0)$.

(3) In the range $-1 < \rho < 1$, the ground state energy is given by an integral, it is an analytical function $E_1(\rho)$ of ρ.

(4) In the range $\rho > 1$, the ground-state energy is given by a sum; it is an analytical function $E_2(\rho)$ of ρ.

(5) These functions $E_1(\rho)$ and $E_2(\rho)$ are different and have essential singularities at the points $\rho = 1$ and $\rho = -1$; however, the ground-state energy and all its derivatives are continuous for $\rho = 1$ $[E_1^{(n)}(1 - 0) = E_2^{(n)}(1 + 0)]$.

(6) The spin states are defined by the momentum q and the spin components ($S^z = M$ with $M = -1, 0, 1$); the corresponding excitation energies $\eta(\rho, M, q)$ are given by simple expressions.

(7) In the range $-1 < \rho < 1$, the three spin states of momentum q, corresponding to the three possible values of M, are degenerate $[\eta(\rho, \pm 1, q) = \eta(\rho, 0, q)]$ and there is no gap $[\eta(\rho, M, 0) = 0]$.

(8) In the range $-1 < \rho < 1$, there is no gap for the spin state corresponding to $M = 0$, but a gap appears for the other excitations: $[\eta(\rho, \pm 1, q) = \eta(\rho, 0, q) + G(\rho)]$.

(9) The gap $G(\rho)$ vanishes for $\rho = 1$ and becomes equivalent to $(\rho - 2)$ when ρ becomes large.

All these results are not really surprising, but they are not trivial either, and we hope that they may lead to a better understanding of the many-body problem. In fact, the spin Hamiltonian can be transformed into a Hamiltonian describing a system of spinless fermions with interaction, and in a subsequent paper, we plan to examine the implications of the preceding work in this context.

Unfortunately, until now, it has not been possible to calculate exactly the partition function. For this reason, machine calculations have been performed; however, if the results obtained in this way are accurate, they do not really bring forth any new ideas, and they may leave out interesting features of the model. For instance, in the case under consideration, the singularity of the ground state energy for $\rho = 1$ cannot be detected by numerical calculation since all the derivatives of the energy are continuous at this point.

Approximate methods are also available. By using the spinless fermion representation of $H(\rho)$, Katsura[9] calculated the partition function by perturbation. In this case, the perturbing term is the term in ρ, and we can expect that the calculation is valid for $-1 < \rho < 1$ since the ground state energy is analytic in this domain, a result which was not completely obvious *a priori*.

APPENDIX A

Here we give a precise definition of $\epsilon_1(\rho)$ and $\epsilon_2(\rho)$ for complex values of ρ, and we find a relation between these functions.

For $-1 < \rho < 1$, the function $\epsilon_1(\rho)$ is given by an integral

$$\epsilon_1(\rho) = -[\sin \Theta/\Theta] I(\Theta), \tag{A1}$$

with

$$I(\Theta) = \frac{1}{2} \int_{-\infty}^{+\infty} \left[1 - \frac{\tanh x}{\tanh (\pi x/\Theta)} \right] dx \tag{A2}$$

and

$$\rho = \cos \Theta, \qquad 0 < \Theta < \pi. \tag{A3}$$

This definition can be extended to complex values of Θ. In particular, for values of Θ belonging to the strip $0 < \Re\Theta < \pi$, the integral (A2) converges and defines $I(\Theta)$ as an analytic function of Θ. Thus, in this domain, $\epsilon_1(\rho)$ can be determined by using (A1). Now to each value of Θ contained in the strip corresponds a value of ρ and conversely. Thus, we define a function $\epsilon_1(\rho)$ which is analytical with respect to ρ everywhere in the complex plane of ρ but on two cuts located on the real axis $\Im\rho = 0$ and defined respectively by the equations $\Re\rho \geq 1$ and $\Re\rho \leq -1$. These cuts correspond to the strip $0 < \Re\Theta < \pi$.

On the other hand, for $\rho \geq 1$, the function $\epsilon_2(\rho)$ is given by a sum

$$\epsilon_2(\rho) = -(\sinh \Phi/\Phi) J(\Phi), \tag{A4}$$

[9] S. Katsura, Phys. Rev. 127, 1508 (1962); S. Katsura and S. Inawashiro, J. Math. Phys. 5, 109 (1964).

with

$$J(\Phi) = \Phi\left[\sum_{n=1}^{\infty}(1 - \tanh n\Phi) + \tfrac{1}{2}\right] \qquad \text{(A5)}$$

and

$$\rho = \cosh \Phi, \qquad 0 < \Phi < \infty. \qquad \text{(A6)}$$

This definition can be extended to complex values of Φ. In particular, in the domain $\Re\Phi > 0$, $\pi > \Im\Phi > -\pi$, $J(\Phi)$ is an analytical function of Φ. Thus, $\epsilon_2(\rho)$ becomes an analytical function of ρ defined everywhere in the complex plane but on a cut located on the real axis $\Im\rho = 0$ and defined by the condition $-1 \leq \Re\rho \leq 1$. This cut corresponds to the boundary $\Re\Phi = 0$.

Now, with each value of ρ ($\Im\rho \neq 0$), we can associate a value of Θ and a value of Φ. The correspondance between these values is

$$\Theta = -i\Phi, \quad \text{if} \quad \Im\rho > 0 \quad (\Im\Theta < 0, \Im\Phi > 0), \qquad \text{(A7)}$$

$$\Theta = +i\Phi, \quad \text{if} \quad \Im\rho < 0 \quad (\Im\Theta > 0, \Im\Phi < 0). \qquad \text{(A8)}$$

In order to find a relation between $\epsilon_1(\rho)$ and $\epsilon_2(\rho)$ for complex values of ρ, we calculate $I(\Theta)$ by residue for complex values of ρ, namely, in the case $\Im\rho > 0$ and $\Im\Theta < 0$. The results in the case $\Im\rho < 0$ are obtained simply by taking the complex conjugate of all the equations.

In the expression (A2) of $I(\Theta)$, the integrand is an analytical function of x which has two series of poles which are given by the equation

$$x = il\Theta, \quad x = i(m - \tfrac{1}{2})\pi \quad (l, m \text{ integers}). \qquad \text{(A9)}$$

Now let us consider the straight line C which joins the points x_L and x_M given by

$$x_L = i(L + \tfrac{1}{2})\Theta, \qquad x_M = iM\pi, \qquad \text{(A10)}$$

where L and M are positive integers which will become large (see Fig. 4). Thus between C and the real axis C_0, there are $(L + M)$ poles. Let R_{LM} be the residues of these poles. In the expressions giving $I(\Theta)$, we can shift the path of integration from C_0 to C provided that we take these poles into account. Actually, we can write

$$I(\Theta) = R_{LM} + \frac{1}{2}\int_C \left[1 - \frac{\tanh x}{\tanh(\pi x/\Theta)}\right] dx. \qquad \text{(A11)}$$

The sum of the residues R_{LM} can be calculated easily:

$$R_{LM} = \Theta \sum_{l=1}^{L} \tan l\Theta - \pi \sum_{m=1}^{M} \cot[\pi^2(m - \tfrac{1}{2})/\Theta]. \qquad \text{(A12)}$$

In the integral appearing in Eq. (A11), we may change the integration variable by setting

$$x = \rho x_L + (1 - \rho)x_M. \qquad \text{(A13)}$$

In this way, we obtain

$$I(\Theta) = R_{LM} + \tfrac{1}{2}(x_L - x_M) \times \int_{-\infty}^{+\infty} (1 - \tanh[\rho x_L + (1 - \rho)x_M] \times \coth\{\pi\Theta^{-1}[\rho x_L + (1 - \rho)x_M]\})\, d\rho. \qquad \text{(A14)}$$

By taking Eq. (A10) into account, we can write also

$$I(\Theta) = R_{LM} + \tfrac{1}{2}(x_L - x_M) \times \int_{-\infty}^{+\infty} \{1 - \tanh[\rho(x_L - x_M)] \times \tanh[(\rho - 1)\pi\Theta^{-1}(x_L - x_M)]\}\, d\rho. \qquad \text{(A15)}$$

When L and M go to infinity, the distance $|x_L - x_M|$ goes also to infinity. For large values of $|x_L - x_M|$, we can now calculate an asymptotic value of the integral which appears in the right-hand side of Eq. (A15). For this purpose, we introduce a fixed number ρ_0 with $0 < \rho_0 < 1$. When L and M become large, we may write approximately

$$I(\Theta) \simeq R_{LM} + \tfrac{1}{2}(x_L - x_M) \times \int_{-\infty}^{\rho_0} \{1 + \tanh[\rho(x_L - x_M)]\}\, d\rho + \tfrac{1}{2}(x_L - x_M) \times \int_{\rho_0}^{+\infty} \{1 - \tanh[(\rho - 1)\pi\Theta^{-1}(x_L - x_M)]\}\, d\rho. \qquad \text{(A16)}$$

In this formula, the integration can be performed exactly, and after a straightforward calculation, we obtain

$$I(\Theta) \sim R_{LM} + (x_L - x_M), \qquad L, M \to \infty. \qquad \text{(A17)}$$

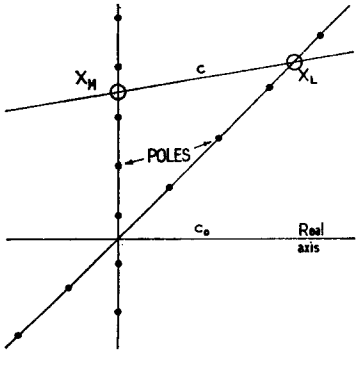


FIG. 4. In the complex x plane, we show the integration contours C_0 and C which are used to calculate $I(\Theta)$. The poles of the expression which appears by integration gives $I(\Theta)$ are represented by black dots.

Now, we replace x_L, x_M, and R_{LM} by their values [Eqs. (A10) and (A11)], and we get

$$I(\Theta) \simeq \Theta\left[\sum_{l=1}^{L}(i + \tan l\Theta) + \frac{i}{2}\right] - \pi\sum_{m=1}^{M}[i + \cot \pi^2(m - \tfrac{1}{2})/\Theta]. \quad \text{(A18)}$$

Now, by going to the limit $L \to +\infty$ $M \to \infty$, we get

$$I(\Theta) = \Theta\left[\sum_{l=1}^{\infty}(i + \tan l\Theta) + \tfrac{1}{2}\right] - \pi\sum_{m=1}^{\infty}[i + \cot \pi^2(m - \tfrac{1}{2})/\Theta]. \quad \text{(A19)}$$

Note that in this calculation, we always assumed $\Im\rho > 0$ and $\Im\Theta < 0$; therefore, in the preceding formula, we can set $\Theta = -i\Phi$, and we get by comparison with Eqs. (A5)

$$I(\Theta) = I(-i\Phi) = J(\Phi) - i\pi\sum_{m=1}^{\infty}[1 - \coth \pi^2(m - \tfrac{1}{2})/\Phi] \quad \text{(A20)}$$

or for $\Im\rho > 0$

$$\epsilon_1(\rho) = \epsilon_2(\rho) - 2i\pi(\sinh \Phi/\Phi) \times \sum_{m=1}^{\infty}\{1 + \exp[\pi^2(m - \tfrac{1}{2})/\Phi]\}^{-1}. \quad \text{(A21)}$$

We get an equivalent formula for $\Im\rho < 0$, by complex conjugation.

APPENDIX B

We want to prove the relations

$$\epsilon_1^{(n)}(1 - 0) = \epsilon_2^{(n)}(1 + 0). \quad \text{(B1)}$$

For $-1 < \rho < 1$, we have

$$\rho = \cos\Theta, \qquad 0 < \Theta < \pi, \quad \text{(B2)}$$

and $\epsilon_1(\rho)$ is defined by

$$\epsilon_1(\rho) = -[\sin\Theta/\Theta]I(\Theta), \quad \text{(B3)}$$

$$I(\Theta) = \int_0^{\infty}\left[1 - \frac{\tanh x}{\tanh(\pi x/\Theta)}\right]dx. \quad \text{(B4)}$$

For $\rho > 1$, we have

$$\rho = \cosh\Phi, \qquad 0 < \Phi, \quad \text{(B5)}$$

and $\epsilon_2(\rho)$ is defined by

$$\epsilon_2(\rho) = -[\sinh\Phi/\Phi]J(\Phi), \quad \text{(B6)}$$

$$J(\Phi) = \Phi[\sum_{n=1}^{\infty}(1 - \tanh n\Phi) + \tfrac{1}{2}]. \quad \text{(B7)}$$

Now, for complex values of ρ, we may set $\Theta = \pm i\Phi$, and we know that the function $(\sin\Theta/\Theta) = (\sinh\Phi/\Phi)$ is an analytical function of ρ around the point $\rho = 1$. Thus, in order to establish the validity of Eq. (B1), we have just to show that the derivative of $I(\Theta)$ with respect to ρ for $\rho = 1 - 0$, coincide with the derivatives of $J(\Phi)$ with respect to ρ for $\rho = 1 + 0$.

For this purpose, the functions $I(\Theta)$ and $J(\Phi)$ will be expanded into formal asymptotic series:

$$I(\Theta) = \sum_{n=0}^{\infty}\lambda_n\Theta^{2n}, \quad \text{(B8)}$$

$$J(\Phi) = \sum_{n=0}^{\infty}\mu_n\Phi^{2n}. \quad \text{(B9)}$$

Whatever the convergence of these series may be, their coefficients give the derivatives of $I(\Theta)$ and $J(\Phi)$ with respect to Θ and Φ. On the other hand, from (B2) and (B5), we can derive an expression of the form

$$\Theta^2 = -\Phi^2 = f(\rho) \quad \text{with} \quad f(1) = 0, \quad \text{(B10)}$$

where $f(\rho)$ is an analytical function of ρ for $\rho = 1$. Thus, in order to establish the validity of equation (B1), it is sufficient to show that for all values of ρ, we have

$$\lambda_n = (-)^n\mu_n.$$

For this purpose, it is convenient to use the following notations:

$$\tanh x = \sum_{n=1}^{\infty}\alpha_n x^{2n-1}, \quad \text{(B11)}$$

$$x\coth x = 1 - 2\sum_{n=1}^{\infty}(-)^n\beta_n x^{2n}, \quad \text{(B12)}$$

where β_n is given by the sum

$$\beta_n = \pi^{-2n}\sum_{p=1}^{\infty}\frac{1}{p^{2n}}, \quad \text{(B13)}$$

which is a well-known result.

First, let us derive the expression of $I(\Theta)$. From (B4), we obtain

$$I(\Theta) - I(0) = \int_0^{\infty}\tanh x[1 - \coth(\pi x/\Theta)]\,dx = (\Theta/\pi)\int_0^{\infty}\tanh(\Theta y/\pi)(1 - \coth y)\,dy. \quad \text{(B14)}$$

Now, we use the expansion (B11)

$$\tanh(\Theta y/\pi) = \sum\alpha_n(\Theta/\pi)^{2n-1}y^{2n-1}. \quad \text{(B15)}$$

By bringing back this expression into Eq. (B15), we obtain a sum of terms of the form

$$\int_0^\infty y^{2n-1}(1 - \coth y)\, dy = -2\int_0^\infty y^{2n-1}\frac{e^{-2y}}{1-e^{-2y}}$$

$$= -2^{-2n+1}(2n-1)!\sum_{p=1}^{\infty}\frac{1}{p^{2n}}$$

$$= -2(\pi/2)^{-2n}(2n-1)!\,\beta_n. \qquad \text{(B16)}$$

Moreover, we have

$$I(0) = \log 2, \qquad \text{(B17)}$$

and therefore we get the final expansion

$$I(\Theta) = \log 2 - 2\sum_{n=1}^{\infty}(2n-1)!\,(\Theta/2)^{2n}\alpha_n\beta_n. \qquad \text{(B18)}$$

Now, let us derive the expansion of $J(\Phi)$. We have

$$J(\Phi) = \Phi\left[\sum_{l=1}^{\infty}(1 - \tanh l\Phi) + \tfrac{1}{2}\right]$$

$$= \frac{\Phi}{2}\sum_{q=0}^{\infty}[2 - \tanh(q\Phi) - \tanh(q+1)\Phi]. \qquad \text{(B19)}$$

By setting

$$f(x) = -\log(1 + e^{-2x}), \qquad \text{(B20)}$$

we can write

$$J(\Phi) = \frac{\Phi}{2}\sum_{q=0}^{\infty}\{f'(q\Phi) + f'[(q+1)\Phi]\}. \qquad \text{(B21)}$$

We can calculate this expansion in terms of Φ by using Taylor's formula, which can be expressed as follows:

$$f(a+b) = \exp(b\partial_a)f(a), \qquad \text{(B22)}$$

$$f(a-b) = \exp(-b\partial_a)f(a). \qquad \text{(B23)}$$

By combining these expressions, we obtain

$$\tfrac{1}{2}[f'(a+b) + f'(a-b)] = \partial_a\cosh(b\partial_a)f(a), \qquad \text{(B24)}$$

$$\tfrac{1}{2}[f(a+b) - f(a-b)] = \sinh(b\partial_a)f(a). \qquad \text{(B25)}$$

We can now eliminate formally the function $f(a)$ which appears on the right-hand side of these equations. We get

$$b[f'(a+b) + f'(a-b)]$$

$$= (b\partial_a)\coth(b\partial_a)[f(a+b) - f(a-b)]. \qquad \text{(B26)}$$

By using the expansion (B12), we can write more explicitly

$$b[f'(a+b) + f'(a-b)] = f(a+b) - f(a-b) - 2$$

$$\times\sum_{n=1}^{\infty}(-)^n\beta_n b^{2n}[f^{(2n)}(a+b) - f^{(2n)}(a-b)]. \qquad \text{(B27)}$$

We can now use this formula to transform equation (B21). For each value of q, we set

$$a + b = (q+1)\Phi, \qquad \text{(B28)}$$

$$a - b = q\Phi. \qquad \text{(B29)}$$

Thus, we get immediately

$$\tfrac{1}{2}\Phi\,[f'(q\Phi) + f'[(q+1)\Phi]] = f[(q+1)\Phi]$$

$$- f(q\Phi) - 2\sum_{n=1}^{\infty}(-)^n\beta_n(\tfrac{1}{2}\Phi)^{2n}$$

$$\times\{f^{(2n)}[(q+1)\Phi] - f^{(2n)}(q\Phi)\}. \qquad \text{(B30)}$$

Now, we remark that all the functions $f^{(n)}(x)$ go to zero when x becomes infinite. Therefore $J(\Phi)$ can be written

$$J(\Phi) = -f(0) + 2\sum_{n=1}^{\infty}(-1)^n\beta_n(\tfrac{1}{2}\Phi)^{2n}f^{2n}(0). \qquad \text{(B31)}$$

But according to (B20) and (B11), we have

$$f'(x) = 1 - \tanh x = 1 - \sum_{n=1}^{\infty}\alpha_n x^{2n-1}. \qquad \text{(B32)}$$

Consequently,

$$f^{(2n)}(0) = -(2n-1)!\,\alpha_n. \qquad \text{(B33)}$$

Finally, the expression of $J(\Phi)$ becomes

$$J(\Phi) = \log 2$$

$$- 2\sum_{n=1}^{\infty}(-)^n(2n-1)!\,(\Phi/2)^{2n}\alpha_n\beta_n. \qquad \text{(B34)}$$

Thus we have found for $I(\Theta)$ and $J(\Phi)$ developments of the form (B8) and (B9). By comparison with Eqs. (B18) and (B31), we obtain the result

$$\lambda_0 = \mu_0 = \log 2, \qquad \text{(B35)}$$

$$\lambda_n = (-)^n\mu_n = -2(2n-1)!\,2^{-2n}\alpha_n\beta_n, \quad n \neq 0. \qquad \text{(B36)}$$

Thus, we prove relation (B10) which implies Eq. (B1).

Degeneracy of Cyclotron Motion*

VICTOR A. DULOCK† AND HAROLD V. MCINTOSH‡
Department of Physics, University of Florida, Gainsville, Florida
(Received 1 April 1965)

The classical problem of planar cyclotron motion of a charged particle in a uniform magnetic field possesses symmetries which account for the "accidental" degeneracies of the analogous nonrelativistic Schrödinger equation, as found by Johnson and Lippman. The essentially quadratic nature of the Hamiltonian is not changed by considering the particle moving in a harmonic oscillator potential, a "Zeeman effect" for the harmonic oscillator. The transitions to the limiting cases of a weak magnetic field (pure harmonic oscillator) or a strong field (pure cyclotron motion) involve the contraction of the corresponding symmetry groups, yielding Larmor precession of the oscillator orbits in the first case, and the drift of the cyclotron orbit in the second. The constants of the motion generate the unitary unimodular group SU_2 in all cases except for pure cyclotron motion, in which case one obtains the commutation rules of creation and annihilation operators. Only for certain ratios of magnetic field strength to the oscillator frequency does one obtain bounded closed orbits, and presumably only in these cases do degeneracies exist quantum-mechanically. A transition to a rotating coordinate system reduces the problem to that of a plane harmonic oscillator; however, the time dependencies of the transformation must be allowed for interpreting the constants thereby arising. Moreover, the velocity-dependent forces introduce gauge transformations which also affect the interpretation of the symmetries. There are two kinds of symmetries—inner symmetries involving the canonical coordinates and governing the shape of the orbits, and outer symmetries involving the mechanical coordinates and governing the location of the orbits.

INTRODUCTION

THE quantum-mechanical discussion of the motion of a charged particle in a uniform magnetic field shows an aspect typical of many of the familiar examples in quantum mechanics; namely, when the energy eigenfunctions are found, they show a degeneracy far beyond that required by the overt symmetry of the problem. Fock[1] gave one of the first, and perhaps one of the most elegant, explanations of this phenomenon in his 1935 paper on the hydrogen atom, in which the dynamic origin of the symmetry could be seen in the phase space of Hamiltonian mechanics, rather than in the symmetry of the configuration space.

A series of authors has discussed, not only the Coulomb problem, but other potentials such as that of the harmonic oscillator; McIntosh[2] has reviewed that work in an earlier paper. However, the field continues to be one of active interest, as evidenced by recent papers of Demkov,[3] Hudson,[4] and others.[5]

The problem of cyclotron motion has been of considerable theoretical and practical interest since its inception in the erroneous but provocative papers of Page.[6] His papers dealt with the measurement of e/m by the mass spectrograph, and at the time, in 1930, there were divergent estimates of this quantity, which he sought to resolve by giving a quantum-mechanical rather than classical analysis of cyclotron motion. Uhlenbeck and Young and others[7] finally obtained a complete solution to the Schrödinger equation for the cyclotron problem, while the discrepancy concerning the value of e/m was eventually resolved by a more careful interpretation of the experiments,[8] which showed that the value of the viscosity of air used in Millikan's oil drop experiment had to be revised.

Probably the first authors to be concerned with constants of the motion, and thereby the symmetry of the equations of motion, were Johnson and Lippmann,[9] who in 1949 applied an operator technique much used by Schwinger[10] to the solution of the cyclotron problem. Their conclusions were that the

† Supported in part by Air Force Grant AFOSR-471-64.
* Based on part of a dissertation submitted by the first author to the Graduate School of the University of Florida in partial fulfillment of the requirements for the Ph.D. degree.
‡ Present address: National Polytechnic Institute of Mexico, Mexico 14 D. F., Mexico.
[1] V. Fock, Z. Physik **98**, 145 (1935).
[2] H. V. McIntosh, Am. J. Phys. **27**, 620 (1959).
[3] Yu. N. Demkov, Zh. Eksperim. i Teor. Fiz. **26**, 757 (1954); Zh. Eksperim. i Teor. Fiz. **36**, 88 (1959); **44**, 2007 (1963) [English transls.: Soviet Phys.—JETP **9**, 63 (1959); **17**, 1349 (1963)].
[4] R. L. Hudson (private communication).
[5] G. A. Baker, Phys. Rev. **103**, 1119 (1956); A. P. Alliluev, Zh. Eksperim. i Teor. Fiz. **33**, 200 (1957) [English transl.: Soviet Phys.—JETP **6**, 156 (1958)]; E. L. Hill, "Seminar on the Theory of Quantum Mechanics" (unpublished), University of Minnesota (1954).
[6] L. Page, Phys. Rev. **36**, 444 (1930).
[7] G. E. Uhlenbeck and L. A. Young, Phys. Rev. **36**, 1721 (1930); L. D. Huff, *ibid.* **38**, 501 (1931); L. Landau, Z. Physik **64**, 629 (1930).
[8] R. T. Birge, Phys. Rev. **48**, 918 (1935).
[9] M. H. Johnson and B. A. Lippmann, Phys. Rev. **76**, 828 (1949).
[10] J. Schwinger, in *Quantum Theory of Angular Momentum*, L. C. Biedenharn and H. Van Dam, Eds. (Academic Press Inc., New York, 1965), pp. 229–279.

x and y centers of the circular cyclotron orbit were constants of the motion, but since their commutator depended upon the strength of the magnetic field, they could not be simultaneously observed. From these constants other conjugate pairs could be constructed, such as the radius and location of the center of the orbit, which were likewise not simultaneously observable.

Some of the most recent interest in the quantum mechanical solutions for the cyclotron problem has arisen in solid state physics, where one is interested in the motion of a charged particle in a uniform magnetic field, but for which there is in addition a periodic electrical field, such as occur in the de Haas, van Alphen effect.[11]

From the point of view of the study of accidental degeneracy, the cyclotron problem has several interesting features. The geometrical symmetry is cylindrical rather than spherical, and as well, there is a translational invariance due to the homogeneity of the field. Thus the geometrical symmetry group is the Euclidean group in two dimensions; moreover, one can effectively treat it as a two-dimensional problem neglecting the motion in the direction of the magnetic field. However, the velocity-dependent nature of the Lorentz force introduces a complication, since translations and rotations are thereby coupled with gauge transformations. Thus one not only has "accidental" degeneracy, but he must be rather more careful than the ordinary in predicting the expected degeneracies from the overt symmetry group.

One sees this in interpreting the eventual results of the analysis of the symmetry. Classically, the orbits are circles, of diameter inversely proportional to the strength of the magnetic field, but otherwise depending upon the initial coordinates and momenta. Quantum-mechanically, the eigenfunctions are harmonic oscillator wave functions corresponding to the classical motion as one would expect. As a result of the Euclidean group symmetry, one finds that such an orbit may have its center anywhere in the plane and that its (degenerate) axes may have any orientation. It is the "accidental" degeneracy, arising from the operators of Johnson and Lippmann, which requires the orbits to be actual circles. Thus there is an "inner" symmetry of an orbit referred to its center and an "outer" symmetry concerning the location of the orbit in the plane at large.

There are certain technical aspects to the cyclotron problem which are interesting. Classically, its Hamiltonian is

$$H = (1/2m)[\mathbf{p} - (e/c)\mathbf{A}]^2, \tag{1}$$

where, in the symmetrical gauge,

$$\mathbf{A} = (-\tfrac{1}{2}B_0 y, \tfrac{1}{2}B_0 x, 0) \tag{2}$$

yields a uniform field B_0 along the z direction. This Hamiltonian is quadratic in the coordinates and momenta, so that mathematically speaking, it is entirely equivalent to an isotropic harmonic oscillator.

Such Hamiltonians have been studied in our earlier paper,[12] the results of which we may describe concisely. With respect to the bilinear form defined by the Poisson bracket,

$$\{f, g\} = \sum \frac{\partial f}{\partial q_k}\frac{\partial g}{\partial p_k} - \frac{\partial f}{\partial p_k}\frac{\partial g}{\partial q_k}, \tag{3}$$

where p_k and q_k are the momenta and the coordinates of the $2n$-dimensional phase space

$$\Phi_n^{(1)} = \{p_1 \cdots p_n, q_1 \cdots q_n\}, \tag{4}$$

the quadratic quantity $q \in \Phi_n^{(2)}$,

$$\Phi_n^{(2)} = \{p_i q_j, p_i p_j, \cdots\}, \tag{5}$$

define linear transformations

$$T_q(f) = \{q, f\}. \tag{6}$$

Calling $\Phi_n^{(r)}$ the set of polynomials in p_k and q_k homogeneous of degree r, we see that $T_q(f)$, $f \in \Phi_n^{(r)}$ define a linear mapping of $\Phi_n^{(r)}$ into itself, and may be represented by a matrix.

It was shown that the matrices representing the transformations $T_q(f)$ would be expected to have eigenvalues which occurred in negative pairs, and that, moreover, if one found a complete set of eigenvectors in $\Phi_n^{(1)}$, they could be used to generate eigenfunctions in every other homogeneous space $\Phi_n^{(r)}$. Calling the eigenfunctions g_i, defined by

$$\{h, g_i\} = \lambda_i g_i, \tag{7}$$

where $h \in \Phi_n^{(2)}$ is the Hamiltonian, one may readily verify that

$$\{h, g_i g_j\} = (\lambda_i + \lambda_j) g_i g_j, \tag{8}$$

so that the product of two eigefunctions, with respect to the Poisson bracket, is another eigenfunction belonging to the sum of the eigenvalues.

In particular, if g_i and g_{-i} belong to a negative pair of eigenvalues λ_i and $\lambda_{-i} = -\lambda_i$, $g_i g_{-i}$ belongs to the eigenvalue 0. Otherwise said, its Poisson

[11] C. Kittel, *Quantum Theory of Solids* (John Wiley & Sons, Inc., New York, 1963).

[12] V. A. Dulock and H. V. McIntosh, Am. J. Phys. 33, 109 (1965).

bracket with h is zero, so that, when h is a quadratic Hamiltonian, we have a general method to find all the rational constants of the motion.

Expanding the Hamiltonian of Eq. (1),

$$H = \frac{1}{2m}(p_x^2 + p_y^2) + \frac{e^2B_0^2}{8mc^2}(x^2 + y^2) + \frac{eB_0}{2mc}(yp_x - xp_y), \quad (9)$$

we see that it is indeed quadratic, so that the earlier theory is applicable. Essentially, it is an isotropic harmonic oscillator Hamiltonian to which a term proportional to the angular momentum has been added. The angular momentum contains cross terms which could be eliminated by a suitable substitution and the Hamiltonian written as a sum of squares, so that one finally deals with the equivalent of an isotropic harmonic oscillator Hamiltonian.

When the eigenvalues of the Hamiltonian are calculated, it is found that not only do they occur in negative pairs, but that two of them are already zero. There are, accordingly, linear constants of the motion, as found by Johnson and Lippmann,[9] in contrast to the usual state of affairs for an oscillator, for which the constants are quadratic and generate a unitary unimodular group of symmetries. Whereas the Poisson bracket of two quadratic constants is again quadratic, the bracket of two linear constants is a constant. As a result, the symmetry group for the cyclotron problem has generators which themselves obey the commutation rules for the harmonic oscillator ladder operators.

Since Jauch and Hill[13] as well as Saenz[14] have shown how the accidental degeneracies of classical and quantum-mechanical problems are often completely equivalent, and since all the relations which we derive are linear, we discuss the classical aspects of the problem only. Also, since the pure cyclotron motion involves a quadratic Hamiltonian, we treat the more general problem of a charged particle moving in a uniform magnetic field but attracted to an origin by a harmonic force. In other words, we actually treat a classical "Zeeman" effect for a harmonic oscillator.

Having obtained the orbits as well as the constants of the motion for the harmonic oscillator in a magnetic field, we consider the two limiting cases in which we have a pure harmonic oscillator or pure cyclotron motion. In the latter case, one may see the transition from one symmetry group to another, in a manner similar to Wigner's method of contraction.[15] These limits are the weak field and strong field cases, respectively, and may be interpreted in terms of Larmor precession. However, the transformation to rotating coordinates, by which the Larmor precession is interpreted, is valid in the strong field limit as well. Rather than obtaining a slowly precessing harmonic oscillator orbit, one obtains drifting cyclotron motion, the circular loops about the field lines being a consequence of periodic phase discrepancies between orbital motion in the elliptical harmonic oscillator orbits and the rapidly but uniformly rotating coordinate system in which the magnetic effects disappear. Thus, our results permit a new perspective for problems involving weak potentials but strong magnetic fields. In a rapidly rotating coordinate system, the magnetic field appears as a harmonic oscillator potential perturbed by the actual potential present. These perturbations affect the drift of the cyclotron orbits in the rest system, transverse to the weak electrostatic field.

The commutation rules for the unitary unimodular groups may be transformed, either by a logarithm or an arctangent mapping, to the commutation rules for canonical coordinates. Thus it is possible to find a system of canonical coordinates for the classical Zeeman problem for which certain of the constants of the motion are canonical coordinates. In some instances, this permits a very useful separation of variables and has in fact been used by Goshen and Lipkin[16] for this purpose. It seems that the form of Hamiltonian which we are studying is quite useful for transformations involving angular momentum, such as to a rotating coordinate system, as well as magnetic fields; moreover, the same technique of cononical coordinates can be used to separate other harmonic oscillator constants from the Hamiltonian—constants such as the correlation or phase differences, which have also been studied by Goshen and Lipkin,[17] as well as forming the basis for much work on bands in nuclear shell structure.[18]

The physical interpretation of the constants of the motion for the classical Zeeman effect is considered, since they do not have as direct a meaning

[13] J. M. Jauch and E. L. Hill, Phys. Rev. **57**, 641 (1940).
[14] A. W. Saenz, "On Integrals of the Motion of the Runge Type in Classical and Quantum Mechanics," Ph.D. thesis, University of Michigan (1949).
[15] W. T. Sharp, "Racah Algebra and the Contraction of Groups," Ph.D. thesis, Princeton University (1960).
[16] S. Goshen and H. J. Lipkin, Ann. Phys. (N. Y.) **6**, 301 (1959).
[17] S. Goshen and H. J. Lipkin, Ann. Phys. (N. Y.) **6**, 310 (1959).
[18] J. P. Elliott, Proc. Roy. Soc. (London) **A245**, 128, 562 (1958); V. Bargmann and M. Moshinsky, Nucl. Phys. **18**, 697 (1960); **23**, 177 (1961).

as for the isotropic harmonic oscillator. As generators of infinitesimal contact transformation, they are seen to regulate the boundaries of the Lissajous figure comprising the orbits. However, upon transformation to a rotating coordinate system, they may be seen to govern the inner and outer radii as well as the rate of precession in the weak field case and the rate of drift and cyclotron radius in the high field case. One has to exercise care with constants of the motion when making a time-dependent transformation, such as the introduction of rotating coordinates; indeed, "constants" in one system become "constant rates" in the other system, and both types of "constants" may exist simultaneously.

Finally, the effect of gauge transformations is examined, since most of our analysis rests upon the choice of one particular gauge, and it is desirable to verify that it persists in other gauges. As is known, since magnetic fields affect the mechanical momentum of particles, to obtain proper conservation laws one must couple a gauge transformation with geometrical motions. When this is done, one obtains the usual symmetries of cyclotron orbits as they lie in space. However, their symmetries about their own centers and other details of the fine structure of their symmetries are only revealed by the constants of the motion responsible for the accidental degeneracies.

CLASSICAL ZEEMAN EFFECT FOR THE HARMONIC OSCILLATOR

Using the symmetric gauge of Eq. (2), the Hamiltonian for a plane isotropic harmonic oscillator in a uniform magnetic field is

$$H = \frac{1}{2m}\left(\mathbf{P} - \frac{e}{c}\mathbf{A}\right)^2 + \tfrac{1}{2}m\omega_0^2 r^2$$

$$= \frac{1}{2m}(P_x^2 + P_y^2) + \frac{m}{2}(\omega_0^2 + \omega^2)(x^2 + y^2) + \omega(yP_x - xP_y), \qquad (10)$$

which is quadratic. In Eq. (10), ω is the Larmor frequency $eB/2mc$, and ω_0 is the natural frequency of the oscillator. This Hamiltonian reduces, in the limit as $\omega \to 0$ ($B \to 0$), to that of the plane harmonic oscillator; while as $\omega_0 \to 0$, it reduces to that of cyclotron motion.

Using the methods of Ref. 12, we may find the matrix representation of this Hamiltonian considered as an operator under Poisson bracket. Calling the representation T_H, we have

$$T_H = \begin{vmatrix} 0 & \omega & m(\omega^2 + \omega_0^2) & 0 \\ -\omega & 0 & 0 & m(\omega^2 + \omega_0^2) \\ -\frac{1}{m} & 0 & 0 & \omega \\ 0 & -\frac{1}{m} & -\omega & 0 \end{vmatrix}. \qquad (11)$$

The basis of the space upon which T_H operates is composed of the monomials (x, y, P_x, P_y). The eigenvalues and eigenvectors of T_H are

Eigenvector	Eigenvalue
u	$i[(\omega^2 + \omega_0^2)^{\frac{1}{2}} + \omega]$
u^*	$-i[(\omega^2 + \omega_0^2)^{\frac{1}{2}} + \omega]$
v	$i[(\omega^2 + \omega_0^2)^{\frac{1}{2}} - \omega]$
v^*	$-i[(\omega^2 + \omega_0^2)^{\frac{1}{2}} - \omega]$

(12)

By defining

$$r^{\pm} = x \pm iy, \qquad (13a)$$

$$P^{\pm} = P_x \pm iP_y, \qquad (13b)$$

the eigenvectors u and v can be written as follows:

$$u = [m(\omega^2 + \omega_0^2)]^{\frac{1}{2}} r^+ + (i/m^{\frac{1}{2}})P^+, \qquad (14a)$$

$$v = [m(\omega^2 + \omega_0^2)]^{\frac{1}{2}} r^- + (i/m^{\frac{1}{2}})P^-, \qquad (14b)$$

where u^* and v^* are simply the complex conjugates of u and v, respectively. These four eigenfunctions satisfy the following relation:

$$\{u^*, u\} = \{v^*, v\} = 4i(\omega^2 + \omega_0^2)^{\frac{1}{2}}. \qquad (15)$$

The constants of the motion will be products of eigenfunctions, the sum of whose eigenvalues is zero. Hence one establishes the quantities uu^*, vv^*, $u^{*R}v$, and $u^R v^*$ as constants of the motion where R is a number such that

$$R\lambda_1 = \lambda_2 \qquad (16)$$

and

$$\lambda_1 = (\omega^2 + \omega_0^2)^{\frac{1}{2}} + \omega, \qquad (17a)$$

$$\lambda_2 = (\omega^2 + \omega_0^2)^{\frac{1}{2}} - \omega. \qquad (17b)$$

In order to display the symmetry group in a convenient form, the following linear combinations are taken as the constants of the motion:

$$H = [\lambda_1 uu^* + \lambda_2 vv^*]/4(\omega^2 + \omega_0^2)^{\frac{1}{2}}, \qquad (18a)$$

$$K = [u^R v^* + u^{*R} v]/R^{\frac{1}{2}}(uu^*)^{\frac{1}{2}(R-1)}, \qquad (18b)$$

$$L = i[u^R v^* - u^{*R} v]/R^{\frac{1}{2}}(uu^*)^{\frac{1}{2}(R-1)}, \qquad (18c)$$

$$D = [uu^* - Rvv^*]/R. \qquad (18d)$$

These four quantities satisfy the following relations:

$$T_H(K) = T_H(L) = T_H(D) = 0, \tag{19}$$

$$T_K(L) = \alpha D, \tag{20a}$$

$$T_L(D) = \alpha K, \tag{20b}$$

$$T_D(K) = \alpha L, \tag{20c}$$

where

$$\alpha = 8(\omega^2 + \omega_0^2)^{\frac{1}{2}}. \tag{21}$$

Since H is the Hamiltonian, the first of these equations is simply a statement of the fact that K, L, and D are constants of the motion, while the three equations (20) show that the symmetry group of the system is SU_2.

If in Eq. (16), R is a rational number, then the classical system has bounded closed orbits, and a quantum-mechanical analog exists for the operators K and L. However, if R happens to be irrational, then the orbits are space-filling.

LIMITING CASES

In taking the limit as $\omega \to 0$, i.e., as the magnetic field is turned off, Eq. (10) becomes the Hamiltonian for the two-dimensional isotropic harmonic oscillator, and the eigenvectors become

$$u = m^{\frac{1}{2}}\omega_0 r^+ + (i/m^{\frac{1}{2}})P^+, \tag{22a}$$

$$v = m^{\frac{1}{2}}\omega_0 r^- + (i/m^{\frac{1}{2}})P^-, \tag{22b}$$

while the four eigenvalues degenerate into two, namely, $\pm i\omega_0$, whence $R = 1$. Expressed in terms of the new u and v, the four constants are

$$H = \tfrac{1}{4}(uu^* + vv^*), \tag{23a}$$

$$K = uv^* + u^*v, \tag{23b}$$

$$L = i(uv^* - u^*v), \tag{23c}$$

$$D = uu^* - vv^*, \tag{23d}$$

which are the constants previously obtained for the isotropic oscillator.[12] The commutation rules for K, L, and D also still hold.

In considering the limit as $\omega_0 \to 0$, we find

$$u = m^{\frac{1}{2}}\omega r^+ + (i/m^{\frac{1}{2}})P^+ \tag{24a}$$

and

$$v = m^{\frac{1}{2}}\omega r^- + (i/m^{\frac{1}{2}})P^-. \tag{24b}$$

The eigenvalues of u and u^* approach $\pm 2i\omega$ in this limit while those belonging to v and v^* both approach zero. Hence, in order to satisfy Eq. (16), R must also approach zero. The Hamiltonian in this limit approaches that for pure cyclotron motion as in Eq. (9). From the values of the eigenvalues, one immediately has two linear constants of the motion, v and v^*, and one quadratic constant uu^*.

The constants can be explicitly derived by considering the commutation rules of K, L, and D in the limit as $\omega_0 \to 0$.

Rewriting the commutation relations explicitly gives

$$i\left\{\frac{u^R v^* + u^{*R} v}{R^{\frac{1}{2}}(uu^*)^{\frac{1}{2}(R-1)}}, \frac{u^R v^* - u^{*R} v}{R^{\frac{1}{2}}(uu^*)^{\frac{1}{2}(R-1)}}\right\} = \alpha \frac{uu^* - Rvv^*}{R}, \tag{25a}$$

$$i\left\{\frac{u^R v^* - u^{*R} v}{R^{\frac{1}{2}}(uu^*)^{\frac{1}{2}(R-1)}}, \frac{uu^* - Rvv^*}{R}\right\} = \alpha \frac{u^R v^* + u^{*R} v}{R^{\frac{1}{2}}(uu^*)^{\frac{1}{2}(R-1)}}, \tag{25b}$$

$$\left\{\frac{uu^* - Rvv^*}{R}, \frac{u^R v^* + u^{*R} v}{R^{\frac{1}{2}}(uu^*)^{\frac{1}{2}(R-1)}}\right\} = i\alpha \frac{u^R v^* - u^{*R} v}{R^{\frac{1}{2}}(uu^*)^{\frac{1}{2}(R-1)}}. \tag{25c}$$

Multiplying the first of these by R and the latter two by $R^{\frac{3}{2}}$ and then taking the limit as $\omega_0 \to 0$ ($R \to 0$) results in the equations

$$i\{(uu^*)^{\frac{1}{2}}(v^* + v), (uu^*)^{\frac{1}{2}}(v^* - v)\} = 8\omega uu^*, \tag{26a}$$

$$i\{(uu^*)^{\frac{1}{2}}(v^* - v), uu^*\} = 0, \tag{26b}$$

$$\{uu^*, (uu^*)^{\frac{1}{2}}(v^* + v)\} = 0. \tag{26c}$$

Since uu^* is simply twice the cyclotron Hamiltonian, it follows from the last two equations that both the real and imaginary parts of v are constants of the motion. Dividing Eq. (26a) by uu^* gives

$$\{\tfrac{1}{2}(v + v^*), \tfrac{1}{2}(v - v^*)\} = 8\omega. \tag{27}$$

From Eq. (9) we see that the cyclotron Hamiltonian splits into two parts, one being the harmonic oscillator Hamiltonian H_0 and the second proportional to the **B** component of the angular momentum L. Both of these terms commute with the total Hamiltonian, $H = H_0 + L$. Hence another quadratic constant of the motion is

$$D = H_0 - L. \tag{28}$$

For convenience in notation, define

$$S = (m^{\frac{1}{2}}/4i)(v - v^*) = m\omega y - P_x, \tag{29a}$$

$$Q = \tfrac{1}{4}m^{\frac{1}{2}}(v - v^*) = m\omega x + P_y. \tag{29b}$$

With these definitions the following commutation rules hold:

$$\{H, D\} = \{H, S\} = \{H, Q\} = 0, \tag{30}$$

$$\{S, Q\} = 2m\omega, \tag{31a}$$

$$\{D, S\} = 2\omega Q, \tag{31b}$$

$$\{D, Q\} = -2\omega S. \tag{31c}$$

These commutation relations coincide with those obtained by Johnson and Lippman.[9] These authors have discussed the two constants S and Q in considerable detail and have shown that they are simply related to the location of the center of the circular orbit and to its diameter, which can readily be seen in the following manner. The canonical momentum expressed in terms of the mechanical momentum is

$$\mathbf{P} = m\mathbf{v} + (e/c)\mathbf{A}. \tag{32}$$

Substituting for the canonical momenta in Eq. (29) gives

$$S = m(2\omega y - v_x), \tag{33a}$$

$$Q = m(2\omega x + v_y). \tag{33b}$$

Evaluating S when $v_x = 0$ and Q when $v_y = 0$ gives the center of the orbit as

$$(x_c, y_c) = (Q/2m\omega, S/2m\omega). \tag{34}$$

These constants also determine the diameter of the orbit. Since the orbit is a circle, only one of the constants must be considered. For example, consider Q. When v_y takes on its maximum positive value, x takes on its minimum value, and when v_y takes on its maximum negative value, x is a maximum, and hence the diameter d is

$$d = x_{\max} - x_{\min} = v_{\max}/\omega. \tag{35}$$

Because of the continuum of points available for the center for a given energy, the degeneracy of this problem is infinite.

CANONICAL COORDINATES

As in the case of the plane harmonic oscillator,[12] a set of canonical coordinates can be found such that the Hamiltonian becomes a cononical momentum. In fact, two momenta for the problem are

$$H = (\lambda_1 uu^* + \lambda_2 vv^*)/4(\omega^2 + \omega_0^2)^{\frac{1}{2}} \tag{36a}$$

and

$$D = (uu^* - Rvv^*)/R, \tag{36b}$$

while the coordinates conjugate to these momenta are

$$Q_1 = \lambda_2[\ln (u^*/u) + \lambda_1 \ln (v^*/v)]/4i\lambda_1\lambda_2, \tag{37a}$$

$$Q_2 = \lambda_2[\ln (u^*/u) - \lambda_1 \ln (v^*/v)]/16i\lambda_2(\omega^2 + \omega_0^2)^{\frac{1}{2}}, \tag{37b}$$

respectively.

A mapping similar to the Hopf mapping may also be performed where

$$u = [\lambda_2^{\frac{1}{2}} \cos \tau e^{-i\rho}] \tag{38a}$$

and

$$v = [\lambda_1^{\frac{1}{2}} \sin \tau e^{-i\sigma}]. \tag{38b}$$

Under this mapping the momenta become

$$H = \lambda_1\lambda_2/4(\omega^2 + \omega_0^2)^{\frac{1}{2}}, \tag{39a}$$

$$D = \lambda_1 \cos \theta, \tag{39b}$$

where

$$\theta = 2\tau \tag{40}$$

and the coordinates are

$$Q_1 = \Psi/2\lambda_2 \tag{41a}$$

and

$$Q_2 = \phi/8R(\omega^2 + \omega_0^2)^{\frac{1}{2}}, \tag{41b}$$

where

$$\Psi = R\rho + \sigma \tag{42a}$$

and

$$\phi = R\rho - \sigma. \tag{42b}$$

Performing the mapping on the other two constants of the motion gives

$$K = \lambda_1 \sin \theta \cos \phi \tag{43a}$$

and

$$L = \lambda_1 \sin \theta \sin \phi. \tag{43b}$$

There also exists another set of canonical coordinates which were originally defined by Goshen and Lipkin.[17] Written in terms of the Cartesian coordinates, the momenta are taken to be

$$P_q = \frac{(P_x^2 + P_y^2)/2m + m(\omega^2 + \omega_0^2)(x^2 + y^2)/2}{(\omega^2 + \omega_0^2)^{\frac{1}{2}}} \tag{44a}$$

and

$$P_\theta = xP_y - yP_x. \tag{44b}$$

The corresponding coordinates are

$$\theta = \frac{1}{2}\tan^{-1}\left[\frac{[P_xP_y/m + m(\omega^2 + \omega_0^2)xy]}{(P_x^2 + P_y^2)/2m + m(\omega^2 + \omega_0^2)(x^2 + y^2)/2}\right] \tag{45a}$$

and

$$q = \frac{1}{2}\tan^{-1}\left[\frac{(\omega^2 + \omega_0^2)^{\frac{1}{2}}(xP_x + yP_y)}{(P_x^2 + P_y^2)/2m - m(\omega^2 + \omega_0^2)(x^2 + y^2)/2}\right]. \tag{45b}$$

In terms of these variables, the Hamiltonian has the particularly simple form

$$H = (\omega^2 + \omega_0^2)^{\frac{1}{2}} P_q - \omega P_\theta, \tag{46}$$

from which it follows that both P_q and P_θ are constant in time.

ROTATING COORDINATES

The problem of the plane harmonic oscillator in a uniform magnetic field has a certain uniqueness when viewed from a rotating coordinate system. However, the problem is first solved in plane polar coordinates. An extremely lucid description and tabulation of these orbits has been given by Harrison.[19]

Assuming the direction of the magnetic field to be in the negative z direction, the Hamiltonian is

$$H = (P_r^2 + P_\theta^2/r^2)/2m + m(\omega^2 + \omega_0^2)r^2/2 + \omega P_\theta, \tag{47}$$

where

$$P_r = m\dot{r}, \tag{48a}$$

$$P_\theta = mr^2(\dot{\theta} - \omega), \tag{48b}$$

and where we have used the gauge

$$A_r = A_z = 0, \tag{49a}$$

$$A_\theta = -\tfrac{1}{2}B_0 r. \tag{49b}$$

The equations of motion are

$$\dot{P}_r = (P_\theta^2/mr^3) - m(\omega^2 + \omega_0^2)r, \tag{50a}$$

$$\dot{P}_\theta = 0, \tag{50b}$$

$$\dot{r} = P_r/m, \tag{50c}$$

$$\dot{\theta} = (P_\theta/mr^2) + \omega. \tag{50d}$$

In general, the effect of imposing a uniform magnetic field on a system with a central potential is to add two terms to the Hamiltonian, namely, a harmonic oscillator potential, which is often neglected for small fields,[20] and a term proportional to P_θ, the angular momentum.

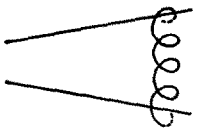


FIG. 1. Particle orbit in a harmonic oscillator potential and a strong uniform magnetic field with initial conditions $\dot{r}_0 = \theta_0 = 0$, $r_0 = \dot{\theta}_0 = 1$, and with $\omega_0 = 4$, $\omega = 63/4$.

[19] E. R. Harrison, Am. J. Phys. 27, 315 (1959).
[20] H. C. Corben and P. Stehle, *Classical Mechanics* (John Wiley & Sons, Inc., New York, 1960), 2nd ed.

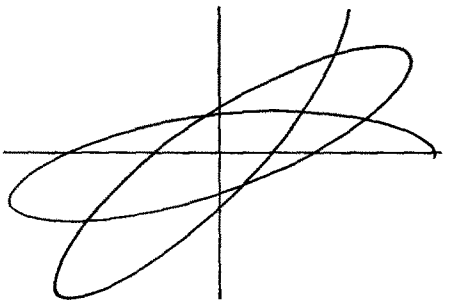


FIG. 2. Particle orbit in a harmonic oscillator potential and a weak uniform magnetic field with initial conditions $\dot{r}_0 = \theta_0 = 0$, $r_0 = \dot{\theta}_0 = 1$, and with $\omega_0 = 4$, $\omega = 3/10$.

The solutions for the orbit equations are

$$\begin{aligned}\theta - \theta_0 = {} & \frac{a}{2\,|a|}\sin^{-1}\left[\frac{br^2 - 2a^2}{r^2(b^2 - 4a^2W^2)^{\frac{1}{2}}}\right] \\ & + \frac{\omega}{2W}\sin^{-1}\left[\frac{2W^2r^2 - b}{(b^2 - 4a^2W^2)^{\frac{1}{2}}}\right] \\ & - \frac{a}{2\,|a|}\sin^{-1}\left[\frac{br_0^2 - 2a^2}{r_0^2(b^2 - 4a^2W^2)^{\frac{1}{2}}}\right] \\ & - \frac{\omega}{2W}\sin^{-1}\left[\frac{2W^2r_0^2 - b}{(b^2 - 4a^2W^2)^{\frac{1}{2}}}\right],\end{aligned} \tag{51}$$

where

$$b = (a^2/r_0^2) + W^2r_0^2 + \dot{r}_0^2, \tag{52a}$$

$$a = P_\theta/m, \tag{52b}$$

$$W = (\omega^2 + \omega_0^2)^{\frac{1}{2}}. \tag{52c}$$

The subscripts on the coordinates and velocities denote initial values. These orbits are plotted in Figs. 1–4. In all cases, units have been chosen such that $m = q = c = 1$. Figures 1 and 2 show the high and low field orbits, respectively, for the same set of initial conditions and ω_0. Figures 3 and 4 show the orbits for a fixed magnetic field, but for different values of the initial tangential velocity. In general, the orbit will be a precessing ellipse for $P_\theta > 0$ and will be a hypotrachoid similar to Fig. 1 for $P_\theta < 0$.

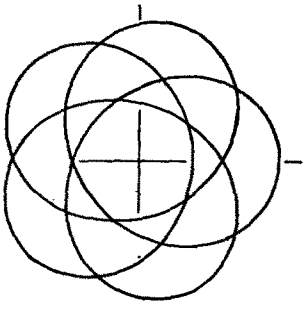


FIG. 3. Particle orbit in a fixed uniform magnetic field and harmonic oscillator potential with a large initial tangential velocity with initial conditions $\dot{r}_0 = \theta_0 = 0$, $r_0 = 1$, $\dot{\theta}_0 = 5$, and with $\omega_0 = 4$, $\omega = 3$.

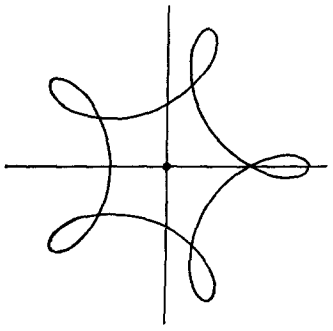


FIG. 4. Particle orbit in a fixed uniform magnetic field and harmonic oscillator with a small initial tangential velocity with initial conditions $\dot{r}_0 = \theta_0 = 0, r_0 = \dot{\theta}_0 = 1$, and with $\omega_0 = 4\ \omega = 3$.

In a rotating system defined by

$$\bar{\theta} = \theta - \omega t, \tag{53a}$$

$$\bar{r} = r, \tag{53b}$$

where ω is the Larmor frequency, the Hamiltonian is

$$\bar{H} = (P_{\bar{r}}^2 + P_{\bar{\theta}}^2/\bar{r}^2)/2m + m(\omega^2 + \omega_0^2)\bar{r}^2/2, \tag{54}$$

where

$$P_{\bar{r}} = m\dot{\bar{r}}, \tag{55a}$$

$$P_{\bar{\theta}} = m\bar{r}^2\dot{\theta}. \tag{55b}$$

Hence the equations of motion are

$$\dot{P}_{\bar{r}} = (P_{\bar{\theta}}^2/m\bar{r}^3) - m(\omega^2 + \omega_0^2)\bar{r}, \tag{56a}$$

$$\dot{P}_{\bar{\theta}} = 0, \tag{56b}$$

$$\dot{\bar{r}} = P_{\bar{r}}/m, \tag{56c}$$

$$\dot{\bar{\theta}} = P_{\bar{\theta}}/m\bar{r}^2, \tag{56d}$$

which are the equations of motion for a plane isotropic harmonic oscillator with force constant $m(\omega^2 + \omega_0^2)$. In general the transformation of a Hamiltonian with a central force from a stationary coordinate system to a rotating coordinate system will subtract a term proportional to P_θ. Generalizing then, it can be stated that a Hamiltonian with a central potential and with a uniform magnetic field present, when viewed from a rotating coordinate system, has a harmonic oscillator potential added, that is

$$\bar{H} = \bar{T} + V(\bar{r}) + \tfrac{1}{2}m\omega^2\bar{r}^2, \tag{57}$$

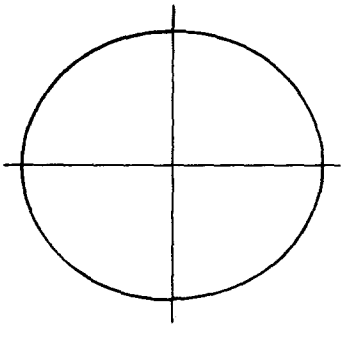


FIG. 5. Particle orbit shown in Fig. 1 as viewed from a coordinate system rotating with angular frequency $\omega = 63/4$.

where $\bar{T}$ is the kinetic energy in the rotating coordinate system and ω is the rotation frequency of the coordinate system with respect to the fixed system. In the present case, since $V(r)$ is a harmonic oscillator potential, one simply has a harmonic oscillator with a larger force constant.

The solution for the orbit defined by the Eqs. (56) is

$$\bar{\theta} - \bar{\theta}_0 = \frac{a}{2\,|a|}\sin^{-1}\left[\frac{b\bar{r}^2 - 2a^2}{\bar{r}^2(b^2 - 4a^2W^2)^{\frac{1}{2}}}\right] - \frac{a}{2\,|a|}\sin^{-1}\left[\frac{b\bar{r}_0^2 - 2a^2}{\bar{r}_0^2(b^2 - 4a^2W^2)^{\frac{1}{2}}}\right], \tag{58}$$

where W is defined in Eq. (52c) and

$$a = P_{\bar{\theta}}/m, \tag{59a}$$

$$b = a^2/\bar{r}_0^2 + W^2\bar{r}_0^2 + \dot{\bar{r}}_0^2. \tag{59b}$$

Equation (58) is the equation for an ellipse, regardless of the magnitude of ω. The orbits corresponding to Figs. 1 and 2 in a rotating coordinate system are shown in Figs. 5 and 6.

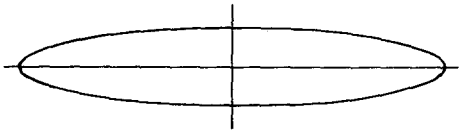


FIG. 6. Particle orbit shown in Fig. 2 as viewed from a coordinate system rotating with angular frequency $\omega = 3/10$.

DISCUSSION OF THE CONSTANTS OF THE MOTION

Of the four constants of the motion defined in Eq. (8), one is H, the energy of the system. Another constant is the angular momentum P_θ, which is a linear combination of H and D,

$$P_\theta = \left[\frac{\omega}{\lambda_1\lambda_2}\right]H - \left[\frac{1+R}{4(\omega^2 + \omega_0^2)^{\frac{1}{2}}}\right]D. \tag{60}$$

Since only three of the constants are independent, one would like to discover one other constant, independent of the two above, which has a physical meaning. This can be done by considering the constants of the motion in the rotating coordinate system and then transforming back to the original coordinate system.

In the rotating coordinate system, two constants of the motion are[12] aa^* and bb^* where

$$a = P_{\bar{x}} - i(\omega^2 + \omega_0^2)^{\frac{1}{2}}\bar{x} \tag{61a}$$

and

$$b = P_{\bar{y}} - i(\omega^2 + \omega_0^2)^{\frac{1}{2}}\bar{y}, \tag{61b}$$

and where the mass is taken to be unity. By a simple rotation, the axis can be oriented so that the semi-major axis of the ellipse lies along $\bar{x}$ and the semi-minor axis lies along $\bar{y}$. In this case then aa^* is proportional to the maximum value which $\bar{x}^2$ attains, because since

$$aa^* = P_x^2 + (\omega^2 + \omega_0^2)\bar{x}^2 \tag{62}$$

is a constant, it can be evaluated when $P_x = 0$. Similarly, bb^* is proportional to $\bar{y}^2_{\max}$. Alternatively, one can say that aa^* and bb^* give the boundaries of the orbit, i.e., aa^* gives the maximum radial distance the particle can attain and bb^* is the minimum radial distance.

These two radii are also constants of the motion in the stationary system since the transformation only involved the angle. However, it can explicitly be shown that these are constants.

By transforming from the rotating coordinate system to the stationary one, aa^* can be written in terms of the eigenvectors and their conjugates as

$$aa^* = \tfrac{1}{4}(uu^* + vv^* + uv^*e^{2i\omega t} + u^*ve^{-2i\omega t}). \tag{63}$$

In the stationary system, the time rate of change is

$$\begin{aligned}\frac{d}{dt}(aa^*) &= \{aa^*, H\} + \frac{\partial}{\partial t}(aa^*)\\ &= \tfrac{1}{4}[-2i\omega uv^*e^{2i\omega t} + 2i\omega u^*ve^{-2i\omega t}\\ &\quad + 2i\omega uv^*e^{2i\omega t} - 2i\omega u^*ve^{-2i\omega t}] = 0,\end{aligned} \tag{64}$$

and hence aa^* is a constant of the motion in the stationary system also. A similar proof also shows bb^* to be a constant. However, aa^* and bb^* are not independent, but are related to one another through H and D.

Even though a striking physical interpretation cannot be given to the constants K, L, and D, their effect on the orbit under Poisson bracket can be calculated by studying the infinitesimal change each produces.

Before calculating the effects of the constants of the orbit, it is desirable to write the orbit equation in terms of the constants of the motion and the time. This can be done by defining a complex vector

$$\mathbf{r} = x + iy, \tag{65}$$

which can be written in terms of the eigenvectors by inverting Eq. (14), where

$$\mathbf{r} = (u + v^*)/2c, \tag{66}$$

where

$$c = [m(\omega^2 + \omega_0^2)]^{\frac{1}{2}}. \tag{67}$$

Employing the polar forms of both the eigenvectors in Eq. (38) and of the constants in Eqs. (39b) and (43), u and v^* may be expressed in terms of the constants as

$$u = (\tfrac{1}{2}R)^{\frac{1}{2}}[(K^2 + L^2 + D^2)^{\frac{1}{2}} + D]^{\frac{1}{2}}e^{-i\rho}, \tag{68a}$$

$$v^* = (\tfrac{1}{2}R)^{\frac{1}{2}}[(K^2 + L^2 + D^2)^{\frac{1}{2}} - D]^{\frac{1}{2}}e^{i\sigma}. \tag{68b}$$

However, from Eq. (42),

$$\rho = (1/2R)(\Psi + \phi) \tag{69a}$$

and

$$\sigma = (1/2)(\Psi - \phi). \tag{69b}$$

Hence

$$u = (\tfrac{1}{2}R)^{\frac{1}{2}}[(K^2 + L^2 + D^2)^{\frac{1}{2}} + D]^{\frac{1}{2}}e^{-i\phi/2R}e^{-i\Psi/2R} \tag{70a}$$

and

$$v^* = (\tfrac{1}{2}R)^{\frac{1}{2}}[(K^2 + L^2 + D^2)^{\frac{1}{2}} - D]^{\frac{1}{2}}e^{-i\phi/2}e^{i\Psi/2}. \tag{70b}$$

From Eq. (43) it is apparent that

$$\phi = \tan^{-1}(L/Y). \tag{71}$$

The time dependence is in ψ since $\psi/2\lambda_2$ in Eq. (41a) is the coordinate conjugate to H, and hence from

$$d/dt\,(\Psi/2\lambda_2) = \{\Psi/2\lambda_2, H\} = 1, \tag{72}$$

it follows that

$$\Psi = 2\lambda_2(t - t_0) \tag{73a}$$

or

$$\Psi = 2R\lambda_1(t - t_0). \tag{73b}$$

Substituting Eqs. (71) and (73) into Eqs. (70) and then substituting these into Eq. (66) gives

$$\begin{aligned}\mathbf{r} = \frac{1}{2c}\left(\frac{R}{2}\right)^{\frac{1}{2}}&\{(K^2+L^2+D^2)^{\frac{1}{2}}e^{(-i/2R)\tan^{-1}(L/K)}e^{-i\lambda_1(t-t_0)}\\ &+ [(K^2 + L^2 + D^2)^{\frac{1}{2}} - D]^{\frac{1}{2}}e^{(-i/2)\tan^{-1}(L/K)}e^{i\lambda_2(t-t_0)}\}\\ &= \mathbf{r}_- + \mathbf{r}_+,\end{aligned} \tag{74}$$

where

$$\mathbf{r}_- = \frac{1}{2c}\left(\frac{R}{2}\right)^{\frac{1}{2}}[(K^2 + L^2 + D^2)^{\frac{1}{2}} + D]^{\frac{1}{2}} \times e^{(-i/2R)\tan^{-1}(L/K)}e^{-i\lambda_1(t-t_0)} \tag{75a}$$

and

$$\mathbf{r} = \frac{1}{2c}\left(\frac{R}{2}\right)^{\frac{1}{2}}[(K^2 + L^2 + D^2)^{\frac{1}{2}} - D]^{\frac{1}{2}} \times e^{(-i/2)\tan^{-1}(L/K)}e^{i\lambda_2(t-t_0)}. \tag{75b}$$

Equation (74) expresses the orbit equation as the sum of two contrarotating vectors, whose frequencies of rotation are λ_1 and λ_2, and whose amplitudes are functions of the constants of the motion.

The Poisson bracket of the constants K, L, and D with $\bar{r}$ are

$$\{K, \mathbf{r}\} = \frac{\alpha}{2}\left[\frac{L}{[(K^2+L^2+D^2)^{\frac{1}{2}}-D]} - \frac{iKD}{K^2+L^2}\right]\mathbf{r}_+ - \frac{\alpha}{2}\left[\frac{L}{[(K^2+L^2+D^2)^{\frac{1}{2}}+D]} + \frac{iKD}{R(K^2+L^2)}\right]\mathbf{r}_-, \tag{76a}$$

$$\{L, \mathbf{r}\} = -\frac{\alpha}{2}\left[\frac{K}{[(K^2+L^2+D^2)^{\frac{1}{2}}-D]} + \frac{iLD}{K^2+L^2}\right]\mathbf{r}_+ + \frac{\alpha}{2}\left[\frac{K}{[(K^2+L^2+D^2)^{\frac{1}{2}}+D]} - \frac{iLD}{R(K^2+L^2)}\right]\mathbf{r}_-, \tag{76b}$$

$$\{D, \mathbf{r}\} = \frac{i\alpha}{2R}\mathbf{r}_- + \frac{i\alpha}{2}\mathbf{r}_+, \tag{76c}$$

where α is a constant defined in Eq. (21).

The infinitesimal change induced by the constants is that each changes both the amplitude and phase of $\mathbf{r}_+$ and $\mathbf{r}_-$, while preserving the sum of the squares of their sum and difference. If the orbit were an ellipse, this last statement would be equivalent to stating that the sum of the squares of the semi axes of the ellipse is a constant. The proof of the statement goes as follows.

Before the infinitesimal change, the quantity is

$$|\mathbf{r}_+ + \mathbf{r}_-|^2 + |\mathbf{r}_+ - \mathbf{r}_-|^2 = 2(|\mathbf{r}_+|^2 + |\mathbf{r}_-|^2), \tag{77}$$

and after the change, it is

$$|\mathbf{r}_+(1+\epsilon) + \mathbf{r}_-(1+\delta)|^2 + |\mathbf{r}_+(1+\epsilon) - \mathbf{r}_-(1-\delta)|^2 = 2[|\mathbf{r}_+|^2 + |\mathbf{r}_-|^2 + (\epsilon+\epsilon^*)\,|\mathbf{r}_+|^2 + (\delta+\delta^*)\,|\mathbf{r}_-|^2], \tag{78}$$

where second-order terms have been ignored. The difference between the two terms is

$$2[(\epsilon+\epsilon^*)\,|\mathbf{r}_+|^2 + (\delta+\delta^*)\,|\mathbf{r}_-|^2],$$

which is zero in all three cases. For example, if the infinitesimal change is induced by K, one has for the difference

$$2[(\epsilon+\epsilon^*)\,|\mathbf{r}_+|^2 + (\delta+\delta^*)\,|\mathbf{r}_-|^2] = \frac{2\alpha}{2}\frac{2L\,|\mathbf{r}_+|^2}{[(K^2+L^2+D^2)^{\frac{1}{2}}-D]} - \frac{2L\,|\mathbf{r}_-|^2}{[(K^2+L^2+D^2)^{\frac{1}{2}}+D]}$$

$$= \frac{2\alpha R}{8c}\frac{L[(K^2+L^2+D^2)^{\frac{1}{2}}-D]}{[(K^2+L^2+D^2)^{\frac{1}{2}}-D]} - \frac{L[(K^2+L^2+D^2)^{\frac{1}{2}}+D]}{[(K^2+L^2+D^2)^{\frac{1}{2}}+D]} = 0. \tag{79}$$

GAUGE TRANSFORMATIONS IN UNIFORM MAGNETIC FIELDS

In uniform magnetic fields and in the absence of other external potentials, the following theorem concerning a charged particle in this field holds: The change in momentum in going from one point to another is independent of the path taken between the points. The proof is as follows: Let **P** denote the canonical momentum, π the mechanical momentum, **B** the uniform field, and **A** the vector potential such that

$$\mathbf{B} = \nabla \times \mathbf{A}, \tag{80}$$

and since B is uniform,

$$\mathbf{A} = \tfrac{1}{2}\mathbf{B} \times \mathbf{r}. \tag{81}$$

Since

$$d\pi/dt = (e/c)\mathbf{v} \times \mathbf{B} \tag{82}$$

and

$$\mathbf{v} = d\mathbf{r}/dt, \tag{83}$$

one has

$$\int_{\mathbf{r}_1}^{\mathbf{r}_2} d\pi = \frac{e}{c}\int_{\mathbf{r}_1}^{\mathbf{r}_2}\left(\frac{d\mathbf{r}}{dt} \times \mathbf{B}\right)dt = \frac{e}{c}\int_{\mathbf{r}_1}^{\mathbf{r}_2} d\mathbf{r} \times \mathbf{B} = -\frac{2e}{c}\int_{\mathbf{r}_1}^{\mathbf{r}_2} d\mathbf{A}. \tag{84}$$

Hence

$$\Delta\pi = -(2e/c)\Delta\mathbf{A} \tag{85}$$

and

$$\Delta\mathbf{P} = -(e/c)\Delta\mathbf{A}. \tag{86}$$

Equation (86) states that the difference in momenta between any two points is proportional to the difference in the vector potential evaluated at these points. From this it is inferred that the choice of a certain gauge is equivalent to picking the zero of momentum.

One can also see why a translation of coordinates must be accompanied by a gauge transformation. When the translation occurs, the zero of momentum changes, and hence this change must be subtracted from **A** to give the same zero of momentum. How-

ever, the change in **A** is simply a gauge transformation.

In order to show that a gauge transformation can induce a translation, first consider the case of a free particle in a uniform field with the following gauge:

$$\mathbf{A} = \tfrac{1}{2}B_0(-y, x, 0). \tag{87}$$

In units where m and $\omega = eB_0/2mc$ are one, the constants of the motion are

$$P_x - y = \alpha \tag{88a}$$

and

$$P_y + x = \beta. \tag{88b}$$

Using the fact that

$$\mathbf{P} = \mathbf{v} + (e/c)\mathbf{A}, \tag{89}$$

the location of the center of the orbit is found to be

$$(x_c, y_c) = (\tfrac{1}{2}\beta, \tfrac{1}{2}\alpha). \tag{90}$$

However, with the gauge,

$$\mathbf{A} = \tfrac{1}{2}B_0[-(\lambda + y), \mu + x, 0], \tag{91}$$

even though the two constants are the same, one finds the center of the orbit to be

$$(x_c, y_c) = [\tfrac{1}{2}(\beta - \mu)\tfrac{1}{2}(\alpha - \lambda)]. \tag{92}$$

Hence the center of the orbit has been translated $-\frac{1}{2}\mu$ units in the x direction and $\frac{1}{2}\lambda$ units in y.

The center of the cyclotron orbit may also be rotated; however, this is not accomplished by a simple gauge transformation. Instead, one must rotate the vector **A** and also rotate the coordinate system. For the vector potential defined in Eq. (87), the center of the orbit is given in Eq. (90). A rotation of the vector field defined by **A** gives

$$\mathbf{A}_R = \frac{\mathbf{B}_0}{2}\begin{vmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{vmatrix}\begin{vmatrix} -y \\ x \\ 0 \end{vmatrix}$$

$$= \tfrac{1}{2}B_0(-y\cos\theta + x\sin\theta, y\sin\theta + x\cos\theta, 0). \tag{93}$$

The rotation of the coordinates defined by

$$x = x'\cos\theta - y'\sin\theta, \tag{94a}$$

$$y = x'\sin\theta + y'\cos\theta, \tag{94b}$$

transforms $\mathbf{A}_R$ into $\mathbf{A}'$, where

$$\mathbf{A}' = \tfrac{1}{2}B_0(-y', x', 0). \tag{95}$$

Hence, the center of the orbit is

$$(x_c', y_c') = (\tfrac{1}{2}\beta', -\tfrac{1}{2}\alpha') \tag{96}$$

and

$$\tfrac{1}{2}\beta = \tfrac{1}{2}\beta'\cos\theta + \tfrac{1}{2}\alpha'\sin\theta, \tag{97a}$$

$$\tfrac{1}{2}\alpha = -\tfrac{1}{2}\beta'\sin\theta + \tfrac{1}{2}\alpha'\cos\theta, \tag{97b}$$

which shows that the center of the orbit has been rotated with respect to the original coordinate system, while the diameter of the orbit has remained unchanged.

The vector field can also be rotated without rotating the coordinates and still preserve the curl if at the same time the vector is dilated by a factor of sec θ. The new vector potential is

$$\mathbf{A} = \tfrac{1}{2}B_0\sec\theta \times (-y\cos\theta + x\sin\theta, y\sin\theta + x\cos\theta, 0). \tag{98}$$

The transformation leaves the center of the orbit invariant and, in fact, is equivalent to a gauge transformation where the vector $\mathbf{A}_D$,

$$\mathbf{A}_D = \tfrac{1}{2}B_0\tan\theta(x, y, 0), \tag{99}$$

whose curl is zero, has been added to the original vector potential **A**.

SUMMARY

It appears that the problem of cyclotron motion in a uniform magnetic field must join a list of many others—the harmonic oscillator, the Kepler problem, the rigid rotator, the particle in a box—which are said to possess "accidental" degeneracy. As our understanding of these systems and their associated symmetry grows, we find a shifting emphasis upon the role of "accident" in the explanation of the symmetry. Demkov[3] has recently proposed a classification of symmetry types in classical and especially quantum mechanics, which shows a rather interesting trend. Once, when "accidental" degeneracy was a rarity, one was content to find some kind of "hidden" symmetry, generally a symmetry of phase space which was not at all evident when one considered only configuration space. But now, when the mechanism of symmetry in phase space is more apparent, we are faced with the problem not of explaining why there is so much symmetry, but why there is so little. For instance, we must not ask why the isotropic harmonic oscillator has so much degeneracy, since we know the answer. Rather, the question is, why quantum-mechanically that the anisotropic oscillator with incommensurable frequencies has no degeneracy at all?

The analysis of the present paper has a bearing

on this latter question because it shows that there are no defects in the symmetry group in the classical realm whatever obstacles may exist to its extension to the quantum mechanical realm.

When the constant R which appears in Eq. (18) is not equal to unity, or in other words, when the oscillator is not isotropic, the constants of the motion are transcendental functions of the coordinates and momenta. Regarded as functions of the complex variables $(p \pm iq)$, they have an inherent multiple valuedness. Nevertheless, the group which they generate is well defined, as is their effect upon the orbits as described by Eq. (76). Of course, caution must be exercised in the application of such formulas, but if an initial choice of one of the many values of the functions is made consistent power series expansions may be made.

There are two hazards in such a procedure. On the one hand, there is ample evidence from the theory of adiabatic invariants that convergence questions must be scrutinized carefully lest one obtain only an asymptotic series. On the other hand, there may arise very complicated problems of connectivity—not in the group manifold, but in the orbit space on which the group operates. It is clear that neither of these two questions is adequately treated by formula (76) which merely describes the infinitesimal variation in a point on the orbit under the action of one of the three constants.

It is for the resolution of questions such as these that the introduction of the canonical coordinates of Eqs. (36) and (37) is particularly useful. They were visualized, for the isotropic oscillator, in terms of the Hopf mapping, in Ref. 12. The advantage of these canonical coordinates is that according to Eqs. (39) and (43) the constants of the motion, K, L, and D, are actually the Cartesian coordinates of a three-dimensional space for which there is no further doubt concerning the nature of the symmetry group which they generate. All ambiguities are confined to the transformation to these canonical coordinates, and it is there that all instances of multiple values may be identified. It is also this transformation which is nearly impossible to describe in terms of quantized operators, quantum-mechanically.

In addition to clarifying the meaning of symmetry for a quantum-mechanical system, there are many points of the paper which are interesting from a purely classical point of view. These have included the transferral of a constant of the motion from a static to a rotating coordinate system, and the removal of a magnetic field by the introduction of a rotating coordinate system studied in both limits of a strong and weak magnetic field. These latter two techniques are applicable for an arbitrary potential although discussed for a particular case. Finally, the symmetries and constants of the motion appropriate to cyclotron motion in a uniform field have been explicitly shown, in isolation as well as a limiting case for a vanishing harmonic potential.

The Definition of States in Quantum Statistical Mechanics*

GÉRARD G. EMCH
University of Maryland, College Park, Maryland
(Received 22 December 1965)

The controversy relative to the use of normal states in quantum physics is discussed in the light of ergodic theory. The nature of the spectrum of the Hamiltonian is shown to play a central role in the decision to enlarge the ordinary frame provided by the traditional density-matrix formalism. The connection of these considerations with the infinite-time, infinite-volume limits in nonequilibrium statistical mechanics is pointed out.

I. INTRODUCTION

THE ergodic theorems play an important role in the foundations of statistical mechanics. First of all, it is generally taken for granted that an equilibrium state (when it exists!) is also an ergodic state in a sense to be made precise hereafter. Therefore, ergodic theory lies in the basis of *equilibrium* statistical mechanics. Second, the ergodic theorems help to formulate necessary (although in general not sufficient) conditions on a physical system to approach equilibrium. Consequently, ergodic theory is also relevant for the study of the principles of *nonequilibrium* statistical mechanics.

In this paper, our principal tool is an extension of von Neumann's ergodic theorem (valid in case of groups of unitary operators acting on a Hilbert space) to certain groups of operators acting on some class of nonreflexive Banach algebras.

Our aim is to direct the reader's attention to the *form* of ergodic states (and hence of equilibrium states) when one insists on certain assumptions currently made in most of the contemporary papers on quantum nonequilibrium statistical mechanics. By way of introduction, let us recall here that it was recognized a long time ago that the classical Poincaré recurrence theorem has an obvious analog in quantum statistical mechanics: the states of the system are almost periodic functions of the time t when the Hamiltonian has a discrete spectrum. As a consequence, the limit as $t \to \infty$ does not exist for most of the quantities of physical interest, even if their ergodic values (and therefore their equilibrium values) are well defined. The recurrences are not of so much concern for the equilibrium theory. They are, however, much more puzzling as soon as one wants to describe the actual approach to equilibrium. In order to bypass this difficulty, it is common to assume that the Hamiltonian has a continuous spectrum. It is precisely for this last property that one introduces (in a more or less rigorous way) a limiting procedure in which the volume of the system is allowed to become infinite, whereas the density is kept fixed. This procedure is intrinsically intended to be the remedy for some mathematical difficulty: the appearance of recurrences. We want to point out here an unexpected difficulty linked to this "remedy." The precise statement of this difficulty allows to give the mathematical frame of the "correct" infinite-volume–infinite-time limiting procedure.

II. MATHEMATICAL PRELIMINARIES

Since some of the mathematical tools we intend to use are not familiar to all physicists, we define and state some of the properties of the objects with which we deal. This section also serves to introduce the notation. For further details, the reader is referred to any one of the textbooks listed in Ref. 1.

Let $\mathfrak{H}$ be a separable infinite-dimensional Hilbert space. We denote by $\mathfrak{B}$ the Banach algebra of all bounded linear operators on $\mathfrak{H}$ under the usual algebraic operations and the operator bound $|\cdots|$ as a norm. The following two-sided *-ideals of $\mathfrak{B}$ are used hereafter.

The subset $\mathfrak{F}$ of $\mathfrak{B}$ is defined as the set of all bounded operators of *finite rank*:

$$\mathfrak{F} \equiv \{F \in \mathfrak{B} \mid \dim \overline{\mathfrak{R}(F)} < \infty\}.$$

$\mathfrak{A}$ denotes the set of all *compact* (or completely continuous) operators:

$$\mathfrak{A} \equiv \{A \in \mathfrak{B} \mid Af_n \to Af \quad \text{whenever} \quad f_n \rightharpoonup f\}.$$

* This research was carried out at the Institute for Fluid Dynamics and Applied Mathematics of the University of Maryland under support from the Office of Naval Research, Contract NONR 595(22).

[1] M. A. Naimark, *Normed Rings* (P. Noordhoff Ltd., Groningen, The Netherlands, 1964). C. E. Rickart, *General theory of Banach Algebras* (D. Van Nostrand Company, Inc., Princeton, New Jersey, 1960). J. Dixmier, *Les algèbres d'opérateurs dans l'espace Hilbertien* (Gauthier-Villars, Paris, 1957), referred to as Dixmier I. J. Dixmier, *Les C*-algèbres et leurs représentations* (Gauthier-Villars, Paris, 1964), referred to as Dixmier II. F. Riesz and B. Sz.-Nagy, *Leçons d'analyse fonctionnelle* (Gauthier-Villars, Paris, 1955). N. Dunford and J. T. Schwartz, *Linear Operators* (Interscience Publishers, Inc., New York, 1964), Pt. I.

Two properties of $\mathfrak{A}$ are used later on. First, $\mathfrak{A}$ is the closure of $\mathfrak{F}$ in the $|\cdots|$-norm, and, in fact, $\mathfrak{A}$ is the only *closed* two-sided *-ideal in $\mathfrak{B}$. Second, any self-adjoint element A of $\mathfrak{A}$, considered as an operator acting on $\mathfrak{H}$, has a completely discrete spectrum and all its eigenprojectors are finite-dimensional, with the possible exception of the eigenprojector corresponding to the eigenvalue zero.

Let us next denote by $\mathfrak{L}$ the subset of $\mathfrak{B}$ consisting of all *Hilbert–Schmidt* operators:

$$\mathfrak{L} \equiv \{L \in \mathfrak{B} | \sum_{\mu} |L\phi_\mu|^2 < \infty \text{, for any basis } \{\phi_\mu\} \text{ in } \mathfrak{H}\}.$$

The above defining property for $\mathfrak{L}$ induces on $\mathfrak{L}$ a structure of normed algebra with respect to the norm $||\cdots||$ which derives from the scalar product $(A, B) = \tau(AB^*)$ where τ is the usual trace function, defined without ambiguity for all pairs (A, B) of elements of $\mathfrak{L}$. Equipped with this structure, $\mathfrak{L}$ becomes a (closed) Hilbert algebra. It is, moreover, the closure of $\mathfrak{F}$ in the $||\cdots||$-norm. The $||\cdots||$-norm majorizes the $|\cdots|$-norm and $\mathfrak{L}$ is a subset of $\mathfrak{A}$. The Hilbert–Schmidt class has also been referred to as the *Liouville space*, and appears to play an important role both in quantum[2] and classical[3] physics.

The subset $\mathfrak{T}$ of $\mathfrak{A}$, called the *trace class*, can be conveniently defined here as the set of all bounded operators formed as the product of at least two elements of $\mathfrak{L}$. The trace function τ is defined (and finite) on each element of $\mathfrak{T}$ (hence the name of this ideal). τ can be used to define a third norm on $\mathfrak{T}$, namely,

$$|||T||| \equiv \tau([T^*T]^{\frac{1}{2}}),$$

which majorizes the $||\cdots||$-norm. $\mathfrak{T}$ becomes a (closed) Banach algebra under this norm and is, in fact, the $|||\cdots|||$-closure of $\mathfrak{F}$ in $\mathfrak{B}$.

We furthermore need to know that $\mathfrak{T}$ is *isomorphic* to the dual space $\mathfrak{A}^*$ of $\mathfrak{A}$, and that the dual space $\mathfrak{T}^*$ of $\mathfrak{T}$ is *isomorphic* to $\mathfrak{B}$, the isomorphic mapping being provided in both cases by the trace function τ.

Let us now turn our attention to some more detailed properties of the dual $\mathfrak{B}^*$ of $\mathfrak{B}$, since this paper is intended to clarify the role played in physics by the elements of this space. Let us denote by $\mathfrak{B}^*$, $\mathfrak{B}^{(1)}$, $\mathfrak{B}^{(2)}$, $\mathfrak{B}^{(3)}$, and $\mathfrak{B}^{(4)}$ the sets of linear functionals on $\mathfrak{B}$ which are, respectively, continuous for the uniform, ultrastrong, strong, ultraweak, and weak topologies. We have

$$\mathfrak{B}^* \supset \mathfrak{B}^{(1)} = \mathfrak{B}^{(3)} \supset \mathfrak{B}^{(2)} = \mathfrak{B}^{(4)}.$$

For simplicity we denote by $\mathfrak{B}_*$ either $\mathfrak{B}^{(1)}$ or $\mathfrak{B}^{(3)}$. We, moreover, know that $\mathfrak{B}_*$ is the closure (in the sense of the norm!) of $\mathfrak{B}^{(2)}$, so that $\mathfrak{B}_*$ is not only a linear manifold of $\mathfrak{B}^*$, but also a (closed) subspace of $\mathfrak{B}^*$.

The subspace $\mathfrak{B}_*$ of $\mathfrak{B}^*$ has another important property: it coincides with the set of all normal functionals on $\mathfrak{B}$, and it is isomorphic to $\mathfrak{A}^*$ and therefore to $\mathfrak{T}$. To give the physicist a better idea of the importance of this rather profound mathematical theorem, we proceed along the following path (which is by no means intended as a proof).

Any "simple" quantum system can be described[4] through the set of all bounded self-adjoint operators acting on an appropriate Hilbert space $\mathfrak{H}$. Let us define $\mathfrak{B}$ from $\mathfrak{H}$ as above, and denote by $\mathfrak{B}^+$ the set of all positive self-adjoint elements of $\mathfrak{B}$. Physicists are certainly willing to describe the "states" on the considered system as bounded linear functionals on $\mathfrak{B}$, i.e., as elements of $\mathfrak{B}^*$. Some further restrictions are also imposed.

Let us denote by $\mathfrak{B}_*^+$ (resp. $\mathfrak{B}_*^+$) the set of all positive elements of $\mathfrak{B}^*$ (resp. $\mathfrak{B}_*$):

$$\mathfrak{B}_*^+ \equiv \{\phi \in \mathfrak{B}^* \mid \phi(B) \geq 0 \quad \text{for all} \quad B \in \mathfrak{B}^+\}$$

$$\mathfrak{B}_*^+ \equiv \{\phi \in \mathfrak{B}_* \mid \phi(B) \geq 0 \quad \text{for all} \quad B \in \mathfrak{B}^+\}.$$

A positive linear functional $\phi \in \mathfrak{B}_*^+$ is said to be *normal* if and only if

$$\phi(\sum_i E_i) = \sum_i \phi(E_i)$$

for *any* family $\{E_i\}$ of mutually orthogonal projectors of $\mathfrak{B}$. There are several other definitions of the normality of a (positive) linear bounded functional. The definition given above has the advantage of emphasizing why these functionals are also called *completely additive*. This property becomes quite important when we want the expectation-values theory to satisfy at least the axioms of the ordinary theory of probabilities when restricted to any classical subset of the considered quantum system. The assumption that states are not only representable as elements of $\mathfrak{B}_*^+$, but also as elements of $\mathfrak{B}_*^+$ seems, therefore, quite natural and convenient. *This is, however, the assumption we challenge in the present paper.* Incidentally, we should

[2] G. Emch, *Lectures in theoretical Physics, Boulder 1965* (to be published); see also for details: G. Emch, Helv. Phys. Acta **37**, 270, 532 (1964); *ibid.* **38**, 164 (1965); G. Emch and C. Favre, *Coarse-Graining in Liouville Space and Ergodicity* (preprint, Geneva, 1965).

[3] J. C. T. Pool, *Mathematical Aspects of the Weyl Correspondence* (preprint, Brandeis University, 1965).

[4] G. Emch *Lectures in theoretical Physics, Boulder 1964* (University of Colorado Press, Boulder, Colorado, 1965), Vol. VIIa; C. Piron, thesis University of Lausanne (1964).

remark that this controversy is not new. We simply show how ergodic theory can help to argue this case. Still, to make our point clear, we want to spell out that the above-mentioned isomorphism between $\mathfrak{B}_*$ and $\mathfrak{T}$ leads to the conclusion that any normal positive functional ϕ can be represented uniquely by a density operator $W \in \mathfrak{T}^+$ (i.e. by a self-adjoint, positive element of the trace class) via the canonical isomorphism:

$$\phi(B) = \tau(WB) \quad \text{for all} \quad B \in \mathfrak{B}.$$

The belief that these states were the only states of importance for a physical theory led the author[2] to his Liouville space formulation of statistical mechanics. The present paper shows how careful one should be when one uses this formalism.

The last information we need about $\mathfrak{B}^*$ is a theorem due to Dixmier.[5] This theorem states that there is a direct sum decomposition of $\mathfrak{B}^*$ as

$$\mathfrak{B}^* = \mathfrak{A}^* \oplus \mathfrak{A}^\perp,$$

where

$$\mathfrak{A}^\perp \equiv \{\phi \in \mathfrak{B}^* \mid \phi(A) = 0 \quad \text{for all} \quad A \in \mathfrak{A}\}.$$

This means that any functional $\phi \in \mathfrak{B}^*$ can be written in an *unique way* as the *sum* of a *normal* linear functional $\psi \in \mathfrak{B}_*$ and a linear functional $\chi \in \mathfrak{A}^\perp$ *which reduces to zero when restricted to* $\mathfrak{A}$.

For the theorems relative to the ergodicity of one-parameter semigroups of operators acting on a nonreflexive Banach space, the reader is referred to the book of Hille and Phillips,[6] Chap. XVIII. We renounce the summarization of this theory here, since even the most truncated and still consistent exposition of it would extend far beyond the limits of this paper. In this connection, it has been thought more elegant to collect all the relevant lemmas in an Appendix, and to mention in the main text only the precise reference to the ergodic theorems we intend to use.

III. THE TIME EVOLUTION

For any simple quantum mechanical system, the time evolution is described by a continuous one-parameter group $\{\mathfrak{U}_t\}$ of *-algebraic automorphisms of the algebra $\mathfrak{B}$ of all bounded linear operators on an appropriate (separable) Hilbert space $\mathfrak{H}$. Throughout this paper, $\mathfrak{H}$ is supposed to be infinite dimensional. (The case of finite-dimensional Hilbert spaces is trivial since, there, $\mathfrak{F}$, $\mathfrak{T}$, $\mathfrak{L}$, $\mathfrak{A}$, and $\mathfrak{B}$ coincide.) We know (see, for instance, Ref. 1 or, for another point of view which, moreover, justifies the above definition of the time evolution, see Ref. 7) that for each t the automorphism $\mathfrak{U}_t$ can be implemented by any one of the members of a family $[U_t]$ of unitary elements of $\mathfrak{B}$: i.e.,

$$\mathfrak{U}_t B = U_{-t} B U_t \quad \text{for all} \quad B \in \mathfrak{B}.$$

The first consequence of this property is that $\{\mathfrak{U}_t\}$ is still a continuous one-parameter group of automorphisms of $\mathfrak{T}$, $\mathfrak{T}^+$, $\mathfrak{L}$, $\mathfrak{L}^+$, $\mathfrak{A}$, and $\mathfrak{A}^+$. Since $\{\mathfrak{U}_t\}$ is a representation of the real line $\mathbf{R}$, there exists[8] a continuous one-parameter group $\{U_t\}$ of unitary elements of $\mathfrak{B}$ which implements $\{\mathfrak{U}_t\}$. We can, therefore, use Stone's theorem. Let H be the Hermitian "generator" of $\{U_t\}$. H is the *Hamiltonian* of the system. We shall hereafter assume for simplicity that H also belongs to $\mathfrak{B}$. Whether this assumption is not too restrictive for the purpose of nonequilibrium statistical mechanics has been discussed in Ref. 2. We can also define the corresponding *Liouville operator* L as the *Hermitian inner derivation* of $\mathfrak{B}$:

$$LB \equiv [H, B] \quad \text{for all} \quad B \in \mathfrak{B},$$

which "generates" $\{\mathfrak{U}_{-t}\}$.

Let us denote by $\{\mathfrak{U}_t^*\}$ the continuous one-parameter group induced on $\mathfrak{B}^*$ by $\{\mathfrak{U}_t\}$ in the following way:

$$(\mathfrak{U}_t^*\phi)(B) \equiv \phi(\mathfrak{U}_t B)$$
$$\equiv \phi_t(B) \text{ for all } (\phi, B) \in \mathfrak{B}^* \times \mathfrak{B}, \text{ and all } t \in \mathbf{R}.$$

$\{\mathfrak{U}_t^*\}$ is also a continuous group of automorphisms of $\mathfrak{B}_*^*$, $\mathfrak{B}_*$, and $\mathfrak{B}_*^+$. Restricted to $\mathfrak{B}_*$, $\{\mathfrak{U}_t^*\}$ coincides with $\{\mathfrak{U}_{-t}\}$ via the canonical isomorphism of $\mathfrak{B}_*$ and $\mathfrak{T}$.

IV. ERGODICITY

Our problem is now to study the time evolution of the various functions $\phi_t(B)$ and, in particular, to determine the properties of

$$\lim_{t\to\infty} \phi_t(B) \equiv \phi_\infty(B) [\text{with } (\phi, B) \in \mathfrak{B}^* \times \mathfrak{B}]$$

whenever it exists.

We note that, if $\phi_\infty(B)$ exists for a given pair (ϕ, B), then the following limit also exists and coincides with $\phi_\infty(B)$:

$$\lim_{t\to\infty} \left(\frac{1}{t}\right) \int_0^t ds \phi_s(B) \equiv \phi_C(B).$$

[5] J. Dixmier, Ann. Math. 51, 387 (1950), referred to as Dixmier III.

[6] E. Hille and R. S. Phillips, *Functional Analysis and Semi-groups* (American Mathematical Society Colloquium Publications, Providence, Rhode Island, 1957), Vol. XXXI.

[7] G. Emch and C. Piron, J. Math. Phys. 4, 469 (1963).

[8] V. Bargmann, Ann Math. 59, 1 (1954).

It turns out that in some cases it is easier to discuss $\phi_C(B)$ than $\phi_\infty(B)$. We therefore direct most of our attention to $\phi_C(B)$, keeping in mind, however, that it is the asymptotic value $\phi_\infty(B)$ of the expectation value $\phi_t(B)$ which is the principal concern of a physical theory. This procedure leads us to several of our conclusions using *ab absurdo* reasonings.

Let us now introduce the following notation:

(i) For each element ϕ in $\mathfrak{B}^*$, $\mathfrak{C}_\phi$ is the subset of $\mathfrak{B}$ on which $\phi_C(B)$ exists.

(ii) ϕ_C is the linear functional defined *on* $\mathfrak{C}_\phi$ by $\phi_C(B)$.

(iii) $\mathfrak{C}^*$ is the subset of $\mathfrak{B}^*$ defined as the set of all $\phi \in \mathfrak{B}^*$ for which $\mathfrak{C}_\phi$ coincides with $\mathfrak{B}$.

(iv) $\mathfrak{C}$ is the subset of $\mathfrak{B}$ defined as the set of all $B \in \mathfrak{B}$ for which $\phi_C(B)$ exists for all ϕ in $\mathfrak{B}^*$.

We note incidentally that

$$\mathfrak{C} = \bigcap_{\phi \in \mathfrak{B}^*} \mathfrak{C}_\phi.$$

One of the purposes of this section is to discuss some properties of $\mathfrak{C}_\phi$, $\mathfrak{C}^*$, and $\mathfrak{C}$ in terms of the properties of the spectrum of H. Mathematically, this program amounts to the study of the *(C,1)-ergodicity*, in the weak-operator topology, of a given one-parameter group $\{\mathfrak{U}_t\}$ of automorphisms of $\mathfrak{B}$, the continuity properties of the group being determined by the fact that it can be implemented by the group $\{U_t\}$ generated by H in $\mathfrak{B}$. $\{\mathfrak{U}_t\}$ is said to be weakly $(C, 1)$-ergodic in the case where $\mathfrak{C}^* \times \mathfrak{C}$ coincides with $\mathfrak{B}^* \times \mathfrak{B}$, i.e., when $\phi_C(B)$ exists for all pairs (ϕ, B) in $\mathfrak{B}^* \times \mathfrak{B}$. In view of the physicist's interest in normal states, special emphasis is put on the normality questions.

We first want to point out that the above program can be carried out with elementary tools when the pairs (ϕ, B) are restricted to belong to $\mathfrak{B}_* \times \mathfrak{A}$. Incidentally, this will exhibit one limitation of the Liouville space formalism introduced by the author[2] for a rigorous treatment of some problems in non-equilibrium statistical mechanics.

Let us consider the restriction $\{\mathfrak{V}_t\}$ of $\{\mathfrak{U}_t\}$ to $\mathfrak{L}$. $\{\mathfrak{V}_t\}$ is now a one-parameter, continuous group of *unitary* operators acting on a *Hilbert space*. One can therefore make use of von Neumann's ergodic theorem which asserts that the following limit exists in the strong topology (i.e., here in the topology induced by the above $||\cdots||$-norm):

$$s - \lim_{t\to\infty} \left(\frac{1}{t}\right) \int_0^t ds \mathfrak{V}_s = \mathfrak{E}_0$$

and, moreover, that $\mathfrak{E}_0$ is the projector on the subspace of $\mathfrak{L}$ invariant under $\{\mathfrak{V}_t\}$. One can even evaluate[2] the effect of $\mathfrak{E}_0$ in terms of the spectral family of H:

$$\mathfrak{E}_0 A = \sum_i P_i A P_i \equiv \bar{A}_0,$$

where $\{P_i\}$ is the set of all eigenprojectors of H, i.e., the P_i are the discontinuous jumps in the spectral family of H. We remark that $\bar{A}_0$ belongs to $\mathfrak{T}$ (resp. to $\mathfrak{T}^+$) whenever A does.

Let ϕ be any normal functional on $\mathfrak{B}$, T the corresponding element in $\mathfrak{T}$, and $\bar{\phi}_0$ the normal functional on $\mathfrak{B}$ corresponding to $\bar{T}_0$ ($\bar{\phi}_0$ is positive whenever ϕ is positive). Since the existence of $\mathfrak{E}_0$ was established in the strong topology of $\mathfrak{L}$ (considered as an Hilbert space), $\phi_C(B)$ exists for any pair

$$(\phi, B) \in \mathfrak{B}_* \times \mathfrak{L} \sim \mathfrak{T} \times \mathfrak{L} \subseteq \mathfrak{L} \times \mathfrak{L}.$$

Since $\mathfrak{L}$ contains $\mathfrak{F}$, the closure of which in the $|\cdots|$ norm is $\mathfrak{A}$, the above result can be extended to $\mathfrak{A}$:

$$\phi_C(A) = \bar{\phi}_0(A) \text{ for all } (\phi, A) \in \mathfrak{B}_* \times \mathfrak{A} \sim \mathfrak{A}^* \times \mathfrak{A}.$$

We therefore established the weak $(C, 1)$ ergodicity of $\{\mathfrak{U}_t\}$ restricted to $\mathfrak{A}$. We moreover calculated $\phi_C(B)$ in terms of the spectral resolution of H.

We now prove that this result cannot, in general, be extended to $\mathfrak{B}_* \times \mathfrak{B}$. We proceed by a counter-example which looks so unexpected from the usual physical point of view that we refer to it as the *first ergodic paradox*.

We want to show that there exist cases in which the linear functional ϕ_C, defined from a normal functional ϕ, is no more normal. Let H be any self-adjoint element of $\mathfrak{B}$ which, when considered as an operator acting on $\mathfrak{H}$, has a purely continuous spectrum. As usual, let us denote by $\{\mathfrak{U}_t\}$ the group of automorphisms of $\mathfrak{B}$ defined by

$$\mathfrak{U}_t B \equiv e^{iHt} B e^{-iHt} \text{ for all } (t, B) \in \mathbf{R} \times \mathfrak{B}.$$

$\{\mathfrak{U}_t\}$ satisfies all our previous hypothesis. Let ϕ be any element of $\mathfrak{B}_*$. Since $\{P_i\}$ is empty, $\bar{\phi}_0$ defined as above is the zero functional. If ϕ_C were normal it would have to coincide with $\bar{\phi}_0$ at least on $\mathfrak{A}$ and therefore on $\mathfrak{B}$, since $\mathfrak{B}_*$ is isomorphic to $\mathfrak{A}^*$. Therefore, ϕ_C, if normal, could only be zero (this result is valid for any ϕ in $\mathfrak{B}_*$). To prove that ϕ_C is not normal in general, even when $\phi \in \mathfrak{B}_*$, it is thus sufficient to exhibit a pair (ϕ, B) of elements of $\mathfrak{B}_* \times \mathfrak{B}$ for which $\phi_C(B)$ can be calculated directly and is not zero. There is a class of elements of $\mathfrak{B}$ for which $\phi_C(B)$ can be calculated directly whatever $\phi \in \mathfrak{B}^*$ could be: the commutant $\{H\}'$ of H in $\mathfrak{B}$. One has

$$\phi_c(K) = \phi(K) \text{ for all } K \in \{H\}'.$$

There always exists some element $\phi \in \mathfrak{B}_*$ for which $\phi(K) \neq 0$ for at least some $K \in \{H\}'$. Consequently, for these pairs $\phi_c(K) \neq 0$.

With this counterexample we proved that, in the case where H has a purely continuous spectrum, there always exist some functionals ϕ for which ϕ_c is *not* normal. The reader will convince himself, moreover, that these normal functionals ϕ are not at all pathological, although they give rise to an unexpected result: they include, for instance, all the normal states for which the expectation value of the energy is nonzero. Incidentally, we noticed that for any normal ϕ, $\mathfrak{C}_\phi$ at least contains $\mathfrak{A}$ *and* $\{H\}'$.

The above result can also be expressed in the following form: The assumption that *all* physically relevant states can be represented as *density operators* is self-contradictory as soon as one wants to include *ergodic*, *equilibrium*, or *asymptotic states* in the microscopic description of a simple quantum system, the Hamiltonian of which has a *purely continuous* spectrum. We comment further on this point later in this paper.

V. FURTHER REMARKS ON ERGODICITY

Expressed in mathematical terms, the counterexample presented in the above section shows that $\{\mathfrak{U}_t\}$, restricted to $\mathfrak{T}$, is *not* weakly $(C, 1)$ ergodic in cases where the spectrum of H is purely continuous. As usual, the advantage of a counterexample is that it points out precisely where things are going wrong. Incidentally, the asserted result could also have been derived *ab absurdo* from the general theorems known for semigroups of operators acting on nonreflexive Banach spaces. We used, for this purpose, the sequence of theorems 18.7.3, 18.6.2, 18.5.2, and 18.4.3 of Hille and Phillips.[6] As emphasized by the counterexample, the fact that the spectrum of H was purely continuous played a central role.

The next question is whether things are going better in the opposite case, namely when the spectrum of H is completely discrete. In this case one can again use Theorem 18.7.3 of Ref. 6, and complete its conclusion with Theorem 18.7.4 of the same reference. The answer to our new question is, then, that $\{\mathfrak{U}_t\}$ restricted to $\mathfrak{T}$ *is* $(C, 1)$ ergodic (in either the strong or the weak topologies). We can even prove more, namely that this group is (C, α) ergodic (see, for instance, definition 18.4.3 of Ref. 6) for all strictly positive α, but not for $\alpha = 0$, in agreement with the general recurrence theorem for ϕ_t, which can be proved either on $\mathfrak{L}$ or on $\mathfrak{T}$. In this case, moreover, ϕ_c (which is normal) coincides with $\bar{\phi}_0$ on the whole of $\mathfrak{B}$. (Remember that this result has been established for the case where the initial state ϕ is supposed to be a normal functional.)

The physical consequence of the last result is that there is no *internal* contradiction in the assumption that all physical states are representable by density operators as long as one is dealing with systems, the evolution of which is governed by a Hamiltonian with a purely discrete spectrum.

The last problem in this connection seems to be somewhat more delicate than the rather crude considerations made up to now. For the moment, we do not see any satisfying answer to it besides the bypass limiting procedure proposed in the next section. The question is whether the above results can be extended from $\mathfrak{T}$ to $\mathfrak{B}$. The most simple form in which this problem can be reduced seems to be the following: What are the conditions on H (and in particular on its spectrum) so that the inner Hermitian derivation L, naturally induced on $\mathfrak{B}$ by H, is such that the closure of the sum of its range and of its kernel coincides with $\mathfrak{B}$? The answer to this question would give a criterion for the ergodicity of the time evolution of simple quantum systems. Contrary to the assertions commonly encountered in the current literature, this last problem has by no means been solved in the appropriate generality. The only known ergodic theorems for these systems apply only to systems with a discrete Hamiltonian and only under the assumption that all physical states are representables by density operators.[2]

VI. CONCLUSIONS

In spite of the seemingly paradoxical nature of some of the considerations developed in the present paper, the last section confirmed an expected result, namely: When the spectrum of the Hamiltonian is purely discrete, there are no ergodic difficulties connected with a systematic use of normal functional to describe physical states. Stated more loosely, this last result is: For any simple quantum mechanical system, the discreteness of the spectrum of the Hamiltonian implies that the ergodic average of an ensemble can always be replaced by an ensemble average. (By ensemble we mean here the object that the physicists call either mixture or density matrix, and that the more mathematically inclined minds recognize as a positive normal functional defined on the algebra of all bounded linear operators on a separable Hilbert space.) In statistical mechanics, this is known as the "linear" ergodic

theorem and is the basis of the microscopic theory of equilibrium, since ϕ_C always exists in this case, even when ϕ_∞ does not exist.

This theorem was proved here in such a way as to emphasize the limitations of its validity. The importance of these limitations becomes apparent when one recalls that, in all cases where the above ergodic theorem is valid, no nontrivial asymptotic state exists and no definitive approach to equilibrium is possible in the frame of a purely mechanistic description of the microscopic evolution. This difficulty has been recognized for a long time and is known as the recurrence paradox. It has been proposed to bypass this difficulty by the use of some limiting procedure the aim of which is to ensure the continuity of the spectrum of the Hamiltonian in order to make obsolete the recurrence theorem. In this case, however, the analysis carried out in the main part of this paper (namely Sec. IV) pointed out another difficulty: In the case where the Hamiltonian has a purely continuous spectrum, one can no more work consistently in the traditional frame provided by a description of the physical states as density operators. A consistent treatment would, indeed, require the introduction of states which are not normal as soon as one is interested in the behavior of the system for t going to infinity. This analysis seems, therefore, to generate a dilemma, the terms of which are the following.

(a) One works systematically in the finite volume case in order to be able to rely consistently on the orthodox tools provided by the density matrix formalism. One has, however, to face in this case all the difficulties connected with the recurrence paradox. In particular, probabilistic statements have to be avoided, as they turn out in most cases to be totally irrelevant (see for instance Kac's discussion[9] of the mathematical dogs-and-fleas model due to Ehrenfest).

(b) One takes some kind of infinite volume limit to ensure the continuity of the spectrum of the Hamiltonian. Doing so one avoids the occurrence of the recurrence paradox only to see another paradox rising. This new paradox, which we referred to as the first ergodic paradox, manifests itself by the fact that equilibrium and (*a fortiori*) nontrivial asymptotic states are no more representable in general as density operators even if all the initial states are bona fide normal states.

This dilemma seems, therefore, to throw some new light on the long-lasting controversy "normal states vs generalized states."

Besides this last point, which could seem to be of rather academic interest, this dilemma has also some practical advantage. It emphasizes why one of the most important procedures in nonequilibrium statistical mechanics, namely, the combined infinite-time, infinite-volume limit, has to be carried out in a definite order and in a definite topology—at least if we want to proceed by successive steps all of which are mathematically well defined. The procedure we are about to discuss is latent in most of the modern papers dealing with the foundations of *nonequilibrium statistical mechanics*. It should also be of some importance in *scattering theory*. Most of the time, however, it is not clearly formulated and barely respected. Since it is the way out of the above-mentioned dilemma, we cannot refrain from stating it explicitly.

(i) Calculate the time-dependent expectation values of the observables of interest, *prior* to any infinite-volume limit. These real-valued functions of time are the actual quantities of interest. They are, in general, volume dependent. So also are their ergodic averages, which, however, always exist if the initial states are normal (so that they can be interpreted as orthodox ensembles).

(ii) In a case where this volume dependence is not relevant (i.e., when one is willing to disregard the finite-size effects), take the infinite-volume limit of these expectation values keeping the time finite. Do the same for their ergodic averages. The later can justifiably be called the equilibrium values of the observables of interest. Whether these expectation values (either time-dependent or time-averaged) can be expressed as the restriction of some normal functional is now of purely academic interest. In fact, this is not possible in most of the cases of interest. For instance, the microcanonical distribution does not exist, in general, as a normal functional and the microcanonical equilibrium value exists for the macroscopic observables. (Incidentally, given a microscopic system, the question whether there exists a set of natural macroscopic observables has never been answered by any of the axiomatic discussions of statistical mechanics. It should be pointed out that the possible relation between these observables and the set $\mathfrak{C}$ defined in Sec. IV might turn out to be worth studying. This is, however, beside the point here.)

(iii) Discuss the long-time behavior of the time-dependent expectations values obtained under (ii).

[9] M. Kac, *Probability and Related Topics in Physical Sciences* (Interscience Publishers, Inc., New York, 1959).

If they exist as t goes to infinity, they must approach the corresponding ergodic values calculated directly as in (ii). They show, then, a definitive approach to equilibrium, which is now in no contradiction with the principles of reversible quantum mechanics. One has, however, to remember at this point that this approach to equilibrium is nothing but *an approximation* of the true time-behavior of the finite system under investigation in the laboratory. The *time scale* in which this approximation is valid depends (for a given interaction Hamiltonian) on the size of the sample under investigation, and is hopefully, in general, many orders of magnitude greater than the time available to follow the evolution of the system.

For the sake of completeness, we could mention that this program has been carried out completely in a case[10] where the Hamiltonian of the system was simple enough to allow *exact* calculations involved in each step. In particular, the time scale in which the approximation involved in (ii) is valid was also determined in this particular case. The nonnormality of the ergodic state is also apparent in this prototype model.

In closing, it is useful to emphasize again that the above discussion was motivated by a paradox. This paradox was a consequence of the too-restrictive definition of states usually admitted in statistical mechanics. It reflected, mainly, some property of the spectrum of the Hamiltonian, namely its continuity. It is very true that the thermodynamical limit, as described above, was intended to obtain this property in order to avoid unwanted recurrences. In the process, the number of particles actually present in the system obviously goes to infinity with the volume. To say, however, as one might be tempted to, that the paradox discussed in this paper arises alone from the fact that the number of particles goes to infinity would simply amount to a confusion between causes. We might seem to overemphasize this point, but we think that, although this paper was written with statistical mechanics in mind, its conclusions might also turn out to be useful in other fields of physics where asymptotic time limits are considered.

So far, for the physical implications of the considerations developed in this paper, the mathematical problem mentioned at the end of Sec. V still remains to be solved, however, mainly for its intrinsic mathematical interest.

ACKNOWLEDGMENTS

The author would like to express his appreciation for the comments he received from Professor G. J. Maltese who kindly read the manuscript. This research started from a discussion held with Professor R. Haag during the 1964 Boulder Summer Institute for Theoretical Physics. It was carried out in part while the author was at Princeton University, where he had the opportunity to benefit from the advice of Professor V. Bargman, Professor A. S. Wightman, Dr. M. Guenin, and Dr. W. Hunziker.

APPENDIX

The purpose of this appendix is to establish various lemmas, the knowledge of which might help to reconstruct the details of some of the reasonings sketched in the main part of this paper.

Lemma 1: Let L be the Liouville operator defined on $\mathfrak{B}$ in Sec. III. Let $\mathfrak{R}(L)$ and $\mathfrak{Z}(L)$ be, respectively, its range and its kernel in $\mathfrak{B}$. Then $\mathfrak{R}(L)$ and $\mathfrak{Z}(L)$ have only the zero element $0 \in \mathfrak{B}$ in common.

Proof: This lemma can be seen as a corollary of Putnam's theorem. In the generalized form given by Miles,[11] this theorem states that, for any self-adjoint derivation L on a B^*-algebra $\mathfrak{B}$, $L^2B = 0$ implies $LB = 0$ (where B denotes any element of $\mathfrak{B}$). If B belongs to $\mathfrak{R}(L)$, there exists an element C in $\mathfrak{B}$ such that $LC = B$. If, moreover, B belongs to $\mathfrak{Z}(L)$, $LB = 0$ which implies $L^2C = 0$ and, therefore, by Putnam's theorem, $B = 0$. Q.E.D.

(This lemma is also valid when $\mathfrak{B}$ is replaced by $\mathfrak{T}$ in both statement and conclusion.)

Lemma 2: $\{U_t\}$, as well as its restriction to $\mathfrak{T}$, is of class (E) and of type $\omega_0 = 0$, and, moreover, of class (A) in the classification of Hille and Phillips.[6]

This lemma readily results from the definitions given by Hille and Phillips[6] (beginning of paragraph 10.6. and definition 18.4.1).

Lemma 3: Let H be any Hermitian element of $\mathfrak{B}$, $\mathrm{Sp}(H)$ its spectrum, L the derivation of $\mathfrak{B}$ generated by H, and $L|_{\mathfrak{A}}$ the restriction of L to $\mathfrak{A}$. Then the derivation $L|_{\mathfrak{A}}$ of $\mathfrak{A}$ cannot be an *inner* derivation when $\mathrm{Sp}(H)$ is purely continuous.

H Hermitian and $\mathrm{Sp}(H)$ continuous imply that H does not belong to $\mathfrak{A} + C$, where C is the set of all the scalar multiples of the identity in $\mathfrak{B}$. The lemma results from Dixmier II,[1] exercise 1.9.11e.

[10] G. Emch, J. Math. Phys. 7, 1198 (1966).

[11] P. Miles, Pacific J. Math. 14, 1359 (1965).

Lemma 4: With the same notation as in Lemma 3, the kernel $\mathfrak{Z}(L|_{\mathfrak{x}})$ of $L|_{\mathfrak{x}}$ reduces to the zero element $0 \in \mathfrak{B}$ when $\mathrm{Sp}(H)$ is purely continuous.

Proof: If A belongs to $\mathfrak{Z}(L|_{\mathfrak{x}})$ so does A^*. It is therefore sufficient to prove that all the self-adjoint elements of $\mathfrak{Z}(L|_{\mathfrak{x}})$ are 0. Since every Hermitian element of $\mathfrak{A}$ has discrete spectrum, all the points of which (with the possible exception of 0) have a finite multiplicity, it is furthermore sufficient to prove that any finite-dimensional projector in $\mathfrak{Z}(L|_{\mathfrak{x}})$ is zero. All these projectors commute with H by definition of the kernel. Consequently, H maps on itself any of the subspaces of $\mathfrak{H}$ corresponding to these projectors. Purely continuous $\mathrm{Sp}(H)$ thus implies that all these spaces reduce to zero. Q.E.D.

We want to mention at this point that the above reasoning also applies to all the ideals of $\mathfrak{B}$ contained in $\mathfrak{A}$. In particular, this kind of argument was used[2] in the determination of the form of $\mathfrak{E}_0$ in terms of the spectral family of a general Hermitian, bounded H.

Let us now denote by $\mathfrak{R}(L|_{\mathfrak{x}})$ and $\mathfrak{Z}(L|_{\mathfrak{x}})$, respectively, the linear manifolds in $\mathfrak{T}$ defined as the range and the kernel of $L|_{\mathfrak{x}}$. From the remark following the proof of Lemma 1 above, we know that these two linear manifolds in $\mathfrak{B}$ have only 0 in common.

Lemma 5: The closure in $\mathfrak{T}$ of the direct sum $\{\mathfrak{R}(L|_{\mathfrak{x}}) + \mathfrak{Z}(L|_{\mathfrak{x}})\}$ is properly contained in $\mathfrak{T}$ when $\mathrm{Sp}(H)$ is purely continuous, and coincides with $\mathfrak{T}$ when $\mathrm{Sp}(H)$ is purely discrete.

Let us first treat the case where $\mathrm{Sp}(H)$ is purely continuous. From the first remark following the proof of Lemma 4 we know that in this case $\mathfrak{Z}(L|_{\mathfrak{x}})$ reduces to 0. It is therefore sufficient to prove that the closure of $\mathfrak{R}(L|_{\mathfrak{x}})$ is properly contained in $\mathfrak{T}$. We know[1] that if $L|_{\mathfrak{x}}$ is any bounded linear operator on a Banach space $\mathfrak{T}$, the closure of its range in $\mathfrak{T}$ is the set of all vectors T in $\mathfrak{T}$ such that $\phi(T) = 0$ for all ϕ in $\mathfrak{T}^*$ which satisfy the equation $L|_{\mathfrak{x}}^*\phi = 0$. We here have the advantage of knowing that $\mathfrak{T}^*$ is isomorphic to $\mathfrak{B}$ and that the isomorphic mapping is provided by the trace function. Moreover $(-L|_{\mathfrak{x}}^*)$ coincides with L as defined originally on $\mathfrak{B}$. Consequently $L|_{\mathfrak{x}}^* = 0$ implies $LB = 0$ for the corresponding B. Therefore, $\mathfrak{R}(L|_{\mathfrak{x}})^\perp$ is isomorphic to the commutator $\{H\}'$ of H in $\mathfrak{B}$. If we finally remember that $\mathfrak{T}$ is isomorphic to $\mathfrak{B}_*$, we conclude that the closure of $\mathfrak{R}(L|_{\mathfrak{x}})$ is isomorphic to the set of all linear normal functionals on $\mathfrak{B}$ which annihilate $\{H\}'$. Should this subspace coincide with $\mathfrak{T}$ itself, we had that any linear normal functional on $\mathfrak{B}$ annihilates $\{H\}'$, which is false. Consequently, the closure in $\mathfrak{T}$ of $\mathfrak{R}(L|_{\mathfrak{x}})$ is always properly contained in $\mathfrak{T}$. This suffices to prove the lemma for the case $\mathrm{Sp}(H)$ purely continuous, because of the first remark made in this proof. Let us now turn our attention to the case where $\mathrm{Sp}(H)$ is purely discrete. The orthogonal complement (in $\mathfrak{T}^*$) of our direct sum is the intersection of the respective orthogonal complements of $\mathfrak{R}(L|_{\mathfrak{x}})$ and $\mathfrak{Z}(L|_{\mathfrak{x}})$.

We have

$$\mathfrak{Z}(L|_{\mathfrak{x}}) = \{T \in \mathfrak{T} \mid LT = 0\} = \{H\}' \cap \mathfrak{T},$$

$$\mathfrak{Z}(L|_{\mathfrak{x}})^\perp \equiv \{\phi \in \mathfrak{T}^* \mid \phi(T) = 0 \ \forall\ T \in \{H\}' \cap \mathfrak{T}\},$$

$$\sim\tilde{\mathfrak{Z}}^\perp = \{B \in \mathfrak{B} \mid \tau(TB) = 0 \ \forall\ T \in \{H\}' \cap \mathfrak{T}\}.$$

We remember

$$\mathfrak{R}(L|_{\mathfrak{x}})^\perp \sim \{H\}'.$$

We now want to prove that $\mathfrak{R}(L|_{\mathfrak{x}})^\perp \cap \mathfrak{Z}(L|_{\mathfrak{x}})^\perp$ only contains the zero functional on $\mathfrak{T}$. This is equivalent to proving that $\{H\}'$ and $\tilde{\mathfrak{Z}}^\perp$ have only the zero element of $\mathfrak{B}$ in common. For any basis $\{\psi_n\}$ of eigenvectors of H, let us form the operators $T_n \equiv \psi_n \otimes \psi_n$. Each of these T_n belongs to $\{H\}'$ and to $\mathfrak{T}$. Therefore, for any B in $\tilde{\mathfrak{Z}}^\perp$, the trace of (T_nB) exists and is zero. This implies that, for any eigenbasis $\{\psi_n\}$ of H, $(B\psi_n, \psi_n)$ vanishes. If, moreover, B belongs to the commutant of H, B maps every eigensubspace of H into itself. Together with the preceding statement, this implies that B vanishes identically. We therefore proved that the orthogonal complement (in $\mathfrak{T}^*$) of $\mathfrak{R}(L|_{\mathfrak{x}}) + \mathfrak{Z}(L|_{\mathfrak{x}})$ is zero. Consequently, the closure of this manifold in $\mathfrak{T}$ coincides with $\mathfrak{T}$ itself. This achieves the proof of the second part of our lemma. Q.E.D.

This result, together with Lemma 2, was central to the derivation of our conclusions about the ergodicity of $\{U_t\}$ restricted to $\mathfrak{T}$.

Wave-Packet Derivation of Field-Theoretic Scattering Amplitude*

LEE M. FRANTZ
TRW Systems, Redondo Beach, California
(Received 3 January 1965)

A formal expression is derived for the field-theoretic scattering amplitude in a Brillouin–Wigner perturbation expansion. Wave packets are used to introduce the initial conditions, thereby avoiding the necessity of adiabatically switching the coupling constant. The field of incident particles is second-quantized, and the target is first-quantized. The principal improvement on previous derivations is that the number of incident quanta, although finite, is otherwise unrestricted. The result is thus applicable, for example, to the nonrelativistic description of the scattering of a photon beam of arbitrary intensity by an atom or a charged particle.

1. INTRODUCTION

IN conventional field theory, where scattering problems are solved by following the time development of solutions to the Schrödinger equation, it is customary to insert the boundary conditions by involking the adiabatic hypothesis. That is, the initial and final separation of the scattering particles is *simulated* by slowly switching the coupling constant on and off. Although this procedure is manifestly artificial, it is not so repugnant in potential scattering, since there, at least, the interaction term actually does vanish when the incident particles are separated from the target. However, in quantum field theory the interaction is present at all times, even when the scattering bodies are too far apart to influence each other, so that switching the coupling constant on and off is inconsistent with the physical boundary conditions.

In the following derivation of the field-theoretic scattering amplitude, the statement of the boundary conditions is made by using localized wave packets to isolate the scattering particles in the remote past and future. In this way the need for adiabatic switching is eliminated, and the scattering bodies are treated as fully "dressed," rather than "bare," even asymptotically.

There is no conflict found here with the results obtained by using the adiabatic hypothesis; they are in fact confirmed for events involving finite but otherwise arbitrary numbers of scatterers. Nevertheless, the use of the hypothesis in deriving the scattering amplitude weakens the logical connection between the latter and the physical boundary conditions. The purpose of this paper is to strengthen and clarify that connection.

Wick[1] has shown how wave packets may be used to obviate turning off the interaction in potential scattering. He makes use of the fact that one can produce simple formal expressions for the scattering eigenfunctions of the complete Hamiltonian. These, of course, are the same well-known eigenfunctions as those obtained by adiabatic switching,[2]

$$\psi_n = \phi_n + (E_n - H + i\epsilon)^{-1} V \psi_n, \qquad (1.1)$$

but no claim need be made as to their significance in terms of boundary conditions. They are taken only to be mathematical quantities that formally satisfy the Schrödinger equation. Wick is able to show that a weighted superposition of these (with their oscillatory time dependences included) tends in the remote past to a superposition of "free" wavefunctions (eigenfunctions of the unperturbed Hamiltonian) with the same weight factors. Since the latter can be chosen to be a localized wave packet isolated from the scatterer, the initial conditions are established. The asymptotic behavior in the remote future is again that of a superposition of free waves, with the coefficients this time being identified as the transition amplitudes into the various free-particle states.

The asymptotic behavior is obtained by observing that the integral over the energy E of the superposed wavefunctions is caused to vanish for large $|t|$ by the rapid oscillation of the time factor e^{-iEt}, except at places where the integrand varies rapidly enough. The only such place is at the pole on the energy shell, where the energy denominator vanishes, and the contribution is readily extracted. In field theory the wavefunctions are not so simple as (1.1), and the singularities are not so explicit. Wick is concerned only with the scattering of a single boson coupled to a target, and for this case is able to develop a satisfactory field-theoretic version of wave-

* Supported in part by the U. S. Army Research Office, Durham, North Carolina.
[1] G. C. Wick, Phys. Rev. 80, 268 (1950).

[2] ϕ_n is the eigenfunction of the unperturbed Hamiltonian H-V corresponding to the eigenvalue E_n.

packet scattering. This version, however, is not adequate for many-quantum scattering, such as the interaction of an intense photon beam with an atom or charged particle, and does not seem to be general enough in form to be readily extendable. In the present derivation, the (finite) number of incident quanta, as well as the nature of their coupling to the target, is completely arbitrary. We use the general idea developed for potential scattering by Wick, but our approach to the field-theoretic aspect is quite different from his. The derivation applies to a second-quantized field of particles interacting with a singly quantized target. It applies, therefore, to quantum electrodynamics, provided the theory is restricted to the nonrelativistic domain, where there is no pair creation.

The outline of the derivation is as follows. First, we produce formal expressions for the eigenvectors $\psi_k(E_k)$ of the total Hamiltonian, satisfying

$$(E_k - H)\psi_k(E_k) = 0. \tag{1.2}$$

Then, from a linear superposition of these, we construct a general, time-dependent solution $\psi(t)$ of the Schrödinger equation,

$$\psi(t) = \sum_k C_k e^{-iE_k t}\psi_k(E_k), \tag{1.3}$$

where k is a composite variable which stands for the momenta and internal variables of the incident particles and the target. In the continuum limit, the summation over k implies an integration over all the momenta, and therefore over the total energy E_k of the system. Again, for very large $|t|$ the rapid oscillations of the exponential factor $e^{-iE_k t}$ cause the integral over E_k to vanish, except for certain terms in $\psi_k(E_k)$ which do not actually depend on E_k, and also except for places where the remainder of the integrand varies sharply enough. The sharp variations occur at the poles of certain modified propagators, and the contributions from these two types of quantities become the leading terms in an asymptotic expansion of $\psi(t)$ in reciprocal powers of the time. From the asymptotic expansion, we obtain expressions for $\psi(t)$ in the remote past and future, and by comparing them we obtain the scattering amplitude.

To study the asymptotic behavior of $\psi(t)$, we must be able to describe the physical field particles and the physical target when they are not interacting. The description is done with product wavefunctions obtained by taking the state vector for the physical target with all of its virtual field particles, and simply tacking onto it creation operators for the plane-wave field particles,

$$\phi_k = a^\dagger(k_1) \cdots a^\dagger(k_N)\psi\,(\text{target}). \tag{1.4}$$

Then, from an appropriate superposition of these, we construct wave packets,

$$\phi(t) = \sum_k C_k e^{-iE_k t}\phi_k. \tag{1.5}$$

Since $\phi(t)$ is a product of wave packets, it provides a proper description of the physical field particles and the physical target when they are separated.

By means of the asymptotic expansion mentioned above, we find that in the remote past $\psi(t)$ approaches the free wave packet $\phi(t)$ with correction terms going as $1/t$,

$$\psi(t) \sim \phi(t), \qquad t \to -\infty, \tag{1.6}$$

and in the remote future $\psi(t)$ approaches $\phi(t)$ plus a scattered wave packet, again with correction terms going as $1/t$,

$$\psi(t) \sim \phi(t) + \phi_{sc}(t), \qquad t \to +\infty. \tag{1.7}$$

The scattered packet, like the incident packet $\phi(t)$, is found to be a superposition of the product wavefunctions used to describe the physical target and the plane-wave field particles,

$$\phi_{sc}(t) = \sum_p A_p e^{-iE_p t}\phi_p, \tag{1.8}$$

and we identify the weight factors A_p as the probability amplitudes for finding the system in the corresponding states at large times. That is, A_p is the scattering amplitude into state p.

The expression for the scattering amplitude A_p is obtained in terms of a Feymann diagram prescription. It may be seen that the wave-packet nature of the initial state of the system is explicit in the expression for A_p, i.e., the weight factors C_k used to construct the incident packet $\phi(t)$ appear directly.

The explicit appearance of the incident localized packet makes it possible to obtain the cross section in a manner which has somewhat more intuitive appeal than that commonly employed in formal scattering theory. It is customary to calculate the cross section by first calculating the rate of change of the probability of finding the system in some group of final states, and then dividing this by the incident flux. In so doing, one interprets the rate of change of probability as the scattering rate, and the probability current as the particle flux. To avoid attributing special significance to these quantities at some particular instant of time during the scattering event, one calculates average values of them over an infinite time interval. One must then visualize what is essentially a steady-state picture with a constant incident and scattered flux. In

Appendix A, we describe an alternative approach, which adheres more closely to the intuitive dynamical picture involving a single scattering event and localized particles. The cross section can be obtained directly from the probability of finding the system in a given state in the remote future, and it is demonstrated that the expression obtained in this way reduces to the conventional steady-state expression in the limit of quasi-monochromatic, quasi-plane wave incident particles.

2. NOTATION AND DEFINITIONS

The unperturbed, or free-field Hamiltonian is denoted by H_0, and the interaction Hamiltonian by H_1, so that the total Hamiltonian is $H = H_0 + H_1$. The eigenvalues of H_0 are denoted by script $\mathcal{E}$'s and the corresponding eigenvectors by lower case italic letters in Dirac bras and kets, as

$$H_0 \, |k\rangle = \mathcal{E}_k \, |k\rangle. \tag{2.1}$$

Eigenvalues of H will be denoted by capital italic E's. We employ a Brillouin–Wigner perturbation expansion; i.e., energy denominators evaluated in any eigenstate of H_0 serve as bare propagators for the system in that state,

$$S_k(E) = \langle k| \, (E - H_0)^{-1} \, |k\rangle = (E - \mathcal{E}_k)^{-1}. \tag{2.2}$$

As a simple example of how matrix elements of operators are to be described in terms of the Feynman diagrams that characterize them, consider the quantity $\langle l|H_1(E - H_0)^{-1}H_1|k\rangle$. By using the unit operator expanded in the complete set of eigenstates of H_0,

$$1 = \sum_q |q\rangle\langle q|, \tag{2.3}$$

we express this as

$$\langle l| \, H_1(E - H_0)^{-1}H_1 \, |k\rangle = \sum_q \langle l| \, H_1 \, |q\rangle S_q(E)\langle q| \, H_1 \, |k\rangle. \tag{2.4}$$

In this form the matrix element is clearly the sum of contributions from all possible two-vertex diagrams for going from state k to state l. We use the convention that the ordering of vertices in the diagrams, as in the matrix elements, is from right to left.

We make considerable use of the operator

$$G(E) = (E - H)^{-1}, \tag{2.5}$$

which is to be thought of as expanded in powers of the perturbation H_1,

$$G(E) = (E - H_0 - H_1)^{-1} = (E - H_0)^{-1} \sum_{n=0}^{\infty} [H_1(E - H_0)^{-1}]^n. \tag{2.6}$$

Both of these forms are equivalent to

$$G(E) = (E - H_0)^{-1} + (E - H_0)^{-1}H_1G(E). \tag{2.7}$$

The generalization of (2.4) is then

$$\begin{aligned} &\langle l| \, H_1G(E)H_1 \, |k\rangle \\ &= \langle l| \, H_1(E - H_0)^{-1}H_1 \, |k\rangle \\ &\quad + \langle l| \, H_1(E - H_0)^{-1}H_1(E - H_0)^{-1}H_1 \, |k\rangle + \cdots \\ &= \sum_q \langle l| \, H_1 \, |q\rangle S_q(E)\langle q| \, H_1 \, |k\rangle \\ &\quad + \sum_{qq'} \langle l| \, H_1 \, |q\rangle S_q(E)\langle q| \, H_1 \, |q'\rangle S_{q'}(E) \\ &\quad \times \langle q'| \, H_1 \, |k\rangle + \cdots, \end{aligned} \tag{2.8}$$

which is evidently the sum of contributions from all possible diagrams for going from k to l with two *or more* vertices.

The modified propagator $S'_k(E)$ for the system in any unperturbed state k is defined in terms of the bare propagator and the self-energy function $\Sigma_k(E)$ for that state by

$$\begin{aligned} S'_k(E) &= S_k(E) + S_k(E) \, \Sigma_k(E)S_k(E) \\ &\quad + S_k(E) \, \Sigma_k(E)S_k(E) \, \Sigma_k(E)S_k(E) + \cdots \end{aligned} \tag{2.9}$$

$$= [S_k^{-1}(E) - \Sigma_k(E)]^{-1}$$

$$= [E - \mathcal{E}_k - \Sigma_k(E)]^{-1}. \tag{2.10}$$

This may also be written in the form

$$S'_k(E) = S_k(E) + S_k(E) \, \Sigma_k(E)S'_k(E). \tag{2.11}$$

The self-energy function $\Sigma_k(E)$ is the sum of contributions from all diagrams in which the system goes from state k back to state k without k appearing as in intermediate state. It may be expressed as

$$\Sigma_k(E) = \langle k| \, H_1 + H_1G(E)H_1 \, |k\rangle_0, \tag{2.12}$$

where the subscript zero on the ket $|k\rangle_0$ means that we are to discard all diagrams in which k appears as an intermediate state.

Finally, we note that the modified propagator may be expressed as

$$S'_k(E) = \langle k| \, G(E) \, |k\rangle. \tag{2.13}$$

This is easily verified by expanding $G(E)$ as in Eq. (2.6) and comparing the diagram interpretation of the result with that of Eq. (2.9). In both cases we have a propagator line for state k, plus the sum of all diagrams for modifying this line by going from k back to k.

3. THE STATE VECTOR

In this section we obtain formal expressions for the eigenvectors of H. We start with the ansatz

$$\psi_k(E) = Z_k^{\frac{1}{2}}\{|k\rangle + \sum_{l\neq k} |l\rangle S_l(E)\langle l| H_1 + H_1 G(E) H_1 |k\rangle_0\}, \quad (3.1)$$

and prove that $\psi_k(E)$ satisfies

$$(E - H)\psi_k(E) = 0 \quad (3.2)$$

for some appropriate E. In (3.1) the subscript zero on the ket $|k\rangle_0$ means that in the perturbation expansion obtained by means of (2.3) and (2.6) we are to include only contributions from diagrams having no self-energy modifications of the k line, i.e., we are to discard all diagrams in which k appears as an intermediate state. The matrix element $\langle l|\cdots|k\rangle_0$ is then the sum of contributions from all possible diagrams for going from state k to state l with no end modifications of the k line. The summation over discrete values of l reflects the fact that the system is considered to be periodically quantized in a finite volume V. However, V is to be considered sufficiently large so that any summation can be well approximated by an integral. Eventually, we wish to take the limit of infinite quantization volume, where only the integral form is appropriate, but for the time being we must keep V finite so we can keep track of the volume dependences. The energy variable E is to be thought of as complex, so there is no danger of vanishing energy denominators until it becomes necessary to approach the real axis. Finally, Z_k is the wave-function renormalization constant defined by requiring

$$1 = \langle\psi_k(E_k) \mid \psi_k(E_k)\rangle, \quad (3.3)$$

where E_k is the actual eigenvalue. We are not concerned here with the renormalization procedure. Z_k is considered finite, and we assume implicitly that all infinite integrals are handled by imposing an energy cutoff or by using a regularization procedure.

From (3.1) we have

$$H_1\psi_k(E) = Z_k^{\frac{1}{2}} \sum_m |m\rangle\{\langle m| H_1 |k\rangle + \sum_{l\neq k}\langle m| H_1 |l\rangle S_l(E)\langle l| H_1 + H_1 G(E) H_1 |k\rangle_0\}. \quad (3.4)$$

The second term in curly brackets is the sum of contributions from all possible diagrams for going from k to m with at least two vertices, with l specified as the state occurring just before m, and with no end modifications of the k line. Consequently, the summation over l (with $l \neq k$) yields the sum of all diagrams for going from k to m with at least two vertices, and no end modifications of the k line. The term for which m takes on the value k is then the self-energy function $\Sigma_k(E)$, so we may write

$$H_1\psi_k(E) = Z_k^{\frac{1}{2}}\{|k\rangle \Sigma_k(E) + \sum_{m\neq k} |m\rangle\langle m| H_1 + H_1 G(E) H_1 |k\rangle_0\}. \quad (3.5)$$

Using Eq. (3.1) to obtain $(E - H_0)\psi_k(E)$ and (3.5) for $H_1\psi_k(E)$, and using the definition (2.2) of the bare propagator, we now find

$$(E - H)\psi_k(E) = Z_k^{\frac{1}{2}}[E - \varepsilon_k - \Sigma_k(E)]\,|k\rangle. \quad (3.6)$$

The right-hand side of (3.6) vanishes and the Schrödinger equation is satisfied, if and only if $E = E_k$, where E_k is the solution to

$$E_k - \varepsilon_k - \Sigma_k(E_k) = 0. \quad (3.7)$$

Equation (3.7) is evidently a factor of the secular equation. The appropriateness of the term "self-energy" is evident here from the manner in which $\Sigma_k(E_k)$ is displayed as the difference between the energy eigenvalue E_k, and the unperturbed energy ε_k. If (3.7) is satisfied, then E_k is an eigenvalue of the Hermitian operator H, and is consequently real. However, (3.7) clearly cannot be satisfied for real E_k unless Im $\Sigma_k(E_k)$ vanishes. It is shown in Appendix B that Im $\Sigma_k(E_k)$ is proportional to the transition rate out of the state k, and is therefore nonzero for states capable of decaying. As a consequence, there is no solution to Eq. (3.7) unless k represents a stable internal state of the scatterer. For example, if we are studying the scattering of photons by an atom, then k must represent either the atom in its ground level, or an unexcited ion plus free electrons. The transition rate out of a state representing an unexcited ion plus free electrons vanishes as the quantization volume V tends to infinity, since the flux corresponding to a finite number of electrons is proportional to $1/V$. Such a state is thus stable. However, excited atomic states, for which the electron is localized, have finite decay rates, and therefore qualify as unstable. Thus, although each eigenstate of H can be characterized by some eigenstate of H_0, as $\psi_k(E_k)$ is characterized by $|k\rangle$, the reverse is not necessarily true. That is, there are no eigenstates of the total Hamiltonian that correspond to *excited* states of the unperturbed system; the latter show up instead as scattering resonances.

Comparison of (3.7) with expression (2.10) for the modified propagator makes it clear that the

latter has a singularity at a characteristic value of the energy. Most of the modified propagators that are implicit in the definition (3.1) of $\psi_k(E)$ contain energy variables of virtual photons which are integrated over. For these, the singularities can be made well-defined by taking $\psi_k(E_k)$ to be the limit of $\psi_k(E)$ as E approaches E_k from either above or below the real axis. However, some of the singular propagators represent discrete states, which are not part of an integration. One way to eliminate the resultant infinities is first to construct a superposition of state vectors, thereby providing an integral over the (complex) energies of the "discrete" propagators, then let the integration path approach the real axis. The result is a well-defined wave packet, even though the state vectors are individually somewhat ill defined. The limiting procedure is most simply accomplished with the usual $\pm i\epsilon$ device used to define integration contours for ingoing and outgoing waves,

$$\psi(t) = \lim_{\epsilon\to 0} \sum_k C_k e^{-iE_k t}\psi_k(E_k \pm i\epsilon). \tag{3.8}$$

We show below that if the plus sign is chosen, $\psi(t)$ takes the form of a "free" wave packet at large negative times.

The conclusion that $\psi_k(E_k)$ is an eigenvector is subject to the criticism that, although the right-hand side of (3.6) vanishes for $E = E_k$, the factor $\psi_k(E_k)$ is itself poorly defined. However, no such difficulty exists with the quantity $\psi(t)$. It is well defined, and the same manipulations that led to (3.6) can be used to prove that it satisfies the time-dependent Schrödinger equation,

$$[i(\partial/\partial t) - H]\psi(t) = 0. \tag{3.9}$$

4. THE QUANTIZATION VOLUME DEPENDENCES

Our goal is to show that a time-dependent solution to the Schrödinger equation of the form of $\psi(t)$ in (3.8) is a superposition of "free" field particle (e.g., photon) states at sufficiently large $|t|$. We wish to identify the transition amplitudes by a comparison of the asymptotic forms in the remote past and future. Therefore, we are concerned with the detailed mathematical structure of $\psi(t)$. It simplifies the analysis considerably if, from the outset, we recognize certain types of terms occurring in $\psi(t)$ that vanish in the limit as the quantization volume V tends to infinity.

In periodic quantization the field operators, e.g., the vector potential in quantum electrodynamics, are expanded in a set of plane waves of the form $V^{-\frac{1}{2}}e^{i\mathbf{k}\cdot\mathbf{x}}$, so that, to each external photon line in a Feynman diagram, there corresponds a factor $V^{-\frac{1}{2}}$. There is an additional factor $V^{-\frac{1}{2}}$ coming from the weight factor (or wavefunction) C_k in the following way. Suppose C_k is a function of N independent wave-vector variables $\mathbf{k}_1$ to $\mathbf{k}_N$, so that $\sum_k$ represents a multiple summation,

$$C_k = \tilde{C}(\mathbf{k}_1, \cdots \mathbf{k}_N), \quad \sum_k = \sum_{\mathbf{k}_1} \cdots \sum_{\mathbf{k}_N}. \tag{4.1}$$

The wavefunction must be normalized by

$$\sum_k |C_k|^2 = \sum_{\mathbf{k}_1} \cdots \sum_{\mathbf{k}_N} |\tilde{C}(\mathbf{k}_1, \cdots, \mathbf{k}_N)|^2 = 1. \tag{4.2}$$

The transition from a summation over discrete wave vectors to an integration over a continuum is accomplished by introducing the density of states,

$$\sum_{\mathbf{k}} \to [V/(2\pi)^3] \int d^3k, \tag{4.3}$$

so the normalization condition becomes

$$[V/(2\pi)^3]^N \int d^3k_1 \cdots d^3k_N \, |\tilde{C}(\mathbf{k}_1 \cdots \mathbf{k}_N)|^2 = 1. \tag{4.4}$$

It follows that $\tilde{C}$ is proportional to a factor $V^{-\frac{1}{2}}$ for each independent variable, so we define a V-independent wavefunction by

$$C_k \equiv [(2\pi)^3/V]^{\frac{1}{2}N} C(\mathbf{k}_1, \cdots \mathbf{k}_N), \tag{4.5}$$

$$\int d^3k_1 \cdots d^3k_N \, |C(\mathbf{k}_1, \cdots \mathbf{k}_N)|^2 = 1. \tag{4.6}$$

In the diagrammatic expansion of the expression (3.8) for $\psi(t)$ there is, then, for each external photon line representing an absorption one factor of $V^{-\frac{1}{2}}$ from the vertex function, and another from the wavefunction. This net factor of $1/V$ is exactly canceled by the V in the density of states when the summation over the wave vector of the absorbed photon is converted to an integral via (4.3). For each external photon line representing an emission there is a factor of $V^{-\frac{1}{2}}$ from the vertex function, and therefore a net factor of $1/V$ when we square the transition amplitude, which is eventually extracted from $\psi(t)$. This too is exactly canceled by the V in the density of states when we obtain the transition probability by integrating over a range of final-state values of the wave vector corresponding to the emission line. Similarly, for each virtual photon there is a $V^{-\frac{1}{2}}$ from the emission vertex, a $V^{-\frac{1}{2}}$ from the absorption vertex, and a compensating V from the integration over the wave vector.

Now, if two or more wave-vector variables are constrained to be identical (meaning that the cor-

responding photons are constrained to occupy the same modes of the radiation field), as for example in a forward-scattering diagram (one in which a photon is absorbed, then re-remitted into exactly the same mode), the balance of volume factors will be destroyed. There will be fewer integrations, since the number of independent wave-vector variables will be smaller, with a resultant loss in V's from the densities of states. As a consequence, all such terms vanish in the limit of infinite quantization volume.

There is another type of term that vanishes for similar reasons. The occupation numbers corresponding to the mode variables $\mathbf{k}_1$ to $\mathbf{k}_N$ need not be unity. If we denote them by n_1 to n_N, then k can be

$$|k\rangle = |n_1(\mathbf{k}_1), \cdots, n_1(\mathbf{k}_N)\rangle, \tag{4.7}$$

so that in the summation over k there are n_1 photons constrained to have the same mode variable $\mathbf{k}_1$, and so forth. By virtue of the argument above, we know that no more than one photon corresponding to a single mode variable can be absorbed. However, it is also true that if a given occupation number is greater than unity, not even one photon having the corresponding mode variable can be absorbed. The reasoning is as follows. If we specify that a single photon is absorbed from a particular mode having occupation number n greater than one, then we are simultaneously specifying that the $n - 1$ remaining photons will continue to occupy that mode in the final state of the system. Consequently, the mode variable of the absorbed photon must be the same in the initial state as in the final state. This in turn means that, since we must specify a discrete final state in calculating the transition amplitude, we cannot integrate over the wave vector of the absorbed photon, and there is a resulting extra factor of $1/V$. The square of the amplitude is proportional to $1/V^2$, so the factor of V from the integration over final states cannot salvage the term, and it must vanish. It follows that wave packets in which more than one photon are constrained to occupy the same mode do not interact, and therefore do not describe physical phenomena. In the remaining discussion each photon is considered to have its own wave-vector variable, and therefore its own wave packet.

5. THE INITIAL CONDITIONS

We have indicated that $\psi(t)$ must satisfy the boundary condition that, at large negative times, it tends to a superposition of free photon states, which we denote by $\phi(t)$,

$$\psi(t) \to \phi(t), \quad \text{as} \quad t \to -\infty. \tag{5.1}$$

To simplify the description of the free photon states we have taken only the photon field to be second-quantized, and not the scatterer; this procedure is justified in a nonrelativistic theory where there is no pair creation. The physical photon states are then identical to the bare photon states. The physical state of the scatterer is, of course, much more complicated than its bare state. For concreteness, let us think of the scatterer as being an atom and denote any stable state by $|A\rangle$. The corresponding physical state of the isolated atom (no real photons present) is, according to Eq. (3.1),

$$\psi_A = Z_A^{\frac{1}{2}}\{|A\rangle + \sum_{l \neq A} |l\rangle S_l(E_A) \times \langle l| H_1 + H_1 G(E_A) H_1 |A\rangle_0\}, \tag{5.2}$$

where E_A is the solution to

$$E_A - \varepsilon_A - \Sigma_A(E_A) = 0. \tag{5.3}$$

Now, let us take the state $|k\rangle$ to be one with the atom in its unperturbed stable level A plus a group of N photons in specified modes

$$|k\rangle = |\mathbf{k}_1, \cdots, \mathbf{k}_N, A\rangle, \tag{5.4}$$

and let $|l\rangle$ be a state with the atom in some arbitrary unperturbed level B plus another group of photons in specified modes

$$|l\rangle = |\mathbf{l}_1, \cdots, \mathbf{l}_M, B\rangle. \tag{5.5}$$

We use the notation $|l + k\rangle$ to designate a state with the atom in level B and the modes of both k and l occupied,

$$|l + k\rangle = |\mathbf{l}_1, \cdots, \mathbf{l}_M, \mathbf{k}_1, \cdots, \mathbf{k}_N, B\rangle. \tag{5.6}$$

With this notation, we define

$$\phi_k = Z_A^{\frac{1}{2}}\{|k\rangle + \sum_{l \neq A} |l + k\rangle S_l(E_A) \times \langle l| H_1 + H_1 G(E_A) H_1 |A\rangle_0\}, \tag{5.7}$$

which describes a physical atom plus N additional photons in modes k_1 to k_N. This is the same state vector as that indicated in Eq. (1.4) in a creation operator notation. ϕ_k is not an eigenvector of H. However, it follows from Eq. (1.6), which is proved in Sec. 7, that the superposition

$$\phi(t) = \textstyle\sum_k C_k e^{-iE_k t}\phi_k$$

does satisfy the time-dependent Schrödinger equation for large negative times. This behavior should

not be surprising, since $\phi(t)$ is essentially a product wavefunction with one factor describing the photon packet, i.e., the Fourier transform of C_k, which in the remote past does not overlap the position of the scatterer.[3]

$\phi(t)$ properly describes the time development at large negative times of a system containing a physical atom and one or more physical photon packets, which are far enough separated so that they do not interact. It is thus the appropriate "free photon" state vector for the initial condition (5.1). If the atom is in an ionized state, then C_k must also contain a wave packet for the continuum-state electron, and $\phi(t)$ satisfies the Schrödinger equation for times such that there is no overlap of the photons with either the ion or the electron.

For comparison with the asymptotic form of $\psi(t)$ which we obtain later, it is convenient to cast the expression (5.7) for ϕ_k into a slightly different form. First we write

$$\phi_k = Z_A^{\frac{1}{2}}\{|k\rangle + \sum_{l\neq A} |l+k\rangle S_{l+k}(E_k) \times \langle l+k| H_1 + H_1 G(E_k) H_1 |k\rangle_0^0\},$$

where the superscript zero on the ket $|k\rangle_0^0$ means that all diagrams involving interactions of the incident photons designated by k are to be discarded. We have replaced E_A by $E_k = E_A + \omega_k$, where ω_k denotes the total energy of the photons in state k. This was necessitated by the fact that, in each energy denominator, the expectation value H_0 is increased by ω_k, because of the presence of the incident photons. By virtue of the restrictions imposed by the rule indicated by the superscript zero, we may exchange the dummy variable $l + k$ for a more general one and write

$$\phi_k = Z_A^{\frac{1}{2}}\{|k\rangle + \sum_{m\neq k} |m\rangle S_m(E_k) \times \langle m| H_1 + H_1 G(E_k) H_1 |k\rangle_0^0\}. \quad (5.8)$$

The matrix element $\langle m|\cdots|k\rangle_0^0$ is the sum of contributions from all diagrams for going from k to m with no interactions of the external photons and no end modifications of the k line. This may be separated into two groups of diagrams. The first, which we denote by $T_{mk}^0(E_k)$, is the sum of all diagrams for going from k to m with no interactions of the incident photons, and no end modifications of either the k line *or* the m line,

$$T_{mk}^0(E_k) = {}_0\langle m| H_1 + H_1 G(E_k) H_1 |k\rangle_0^0. \quad (5.9)$$

The second group then contains all the diagrams with at least one self-energy correction of the m line. That is, it is the sum of all ways of going from k to m without end modifications, followed by all possible end modifications, so the corresponding contribution can be written as

$$\Sigma_m(E_k) S_m'(E_k) T_{mk}^0(E_k). \quad (5.10)$$

In view of the definition (2.11) of the modified propagator we may finally write ϕ_k in the form

$$\phi_k = Z_A^{\frac{1}{2}}\{|k\rangle + \sum_{m\neq k} |m\rangle S_m'(E_k) T_{mk}^0(E_k)\}. \quad (5.11)$$

6. THE ASYMPTOTIC EXPANSION IN TIME

In the expression (3.8) for $\psi(t)$ the rapid oscillations of the factor $e^{-iE_k t}$ for large $|t|$ cause all contributions to the integral over E_k to vanish except for those terms that do not depend on E_k and except at those places where the remainder of the integrand varies rapidly enough with E_k. The only such rapid variations take place where the (simple) poles of the modified propagators in $\psi_k(E_k \pm i\epsilon)$ occur. To find the behavior of $\psi(t)$ in the remote past and future we make an asymptotic expansion in powers of $1/t$. To obtain the general mathematical form of the expansion theorem that is to be used, it is sufficient to know that the singularities we must deal with are simple poles. Specifically, we develop an asymptotic expansion in $1/t$ for an integral of a function $f(z)$ which is of bounded variation in the range of integration along the real axis, except for a finite number of simple poles,

$$\int_c f(z) e^{-izt}\, dz.$$

The integration contour is shown in Fig. 1. The poles are located at $x = (x_1, x_2, \cdots, x_N)$, and the contour goes above the poles, which is equivalent to the choice of a positive infinitesimal imaginary term $+i\epsilon$ in the propagators. The derivation of the theorem is outlined in Appendix C. The result is

$$\int_c f(z) e^{-izt}\, dz \sim -2\pi i \theta(t) \sum_{n=1}^{N} e^{-ix_n t} R(x_n), \quad (6.1)$$

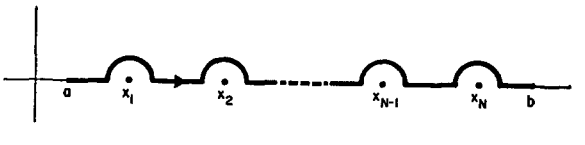


FIG. 1. Integration contour for the integral $\int_c f(z)e^{-izt}dz$, showing the path going above the poles.

[3] The atomic nucleus is thought of throughout as being fixed at a point in space, but one could just as well make a wave packet for it, too, by introducting a wave-vector variable for it in C_k.

plus corrections going as $1/t$, where $\theta(t)$ is the unit step function,

$$\theta(t) = \begin{cases} 0, & t < 0, \\ 1, & t > 0, \end{cases} \tag{6.2}$$

and $R(x_1), \cdots, R(x_N)$ denote the residues at the poles of $f(z)$.

If we are to use this asymptotic expansion for $\psi(t)$, it is first of all necessary that the terms integrated over E_k actually depend on E_k. Those terms corresponding to diagrams with no external photon lines involving the wave vectors of the incident photons do *not*, since the variables of the incident photons cancel out of the propagators. That is, these variables appear in both E_k and the expectation value $\langle H_0 \rangle$, and therefore cancel out of the difference $E_k - \langle H_0 \rangle$ in all energy denominators. We now extract the terms independent of E_k and show that their sum is simply the incident wave packet $\phi(t)$.

The expression for the time-dependent state vector $\psi(t)$ is from (3.1) and (3.8),

$$\psi(t) = \sum_k C_k e^{-iE_k t} Z_k^{\frac{1}{2}} \{ |k\rangle + \sum_{l \neq k} |l\rangle S_l(E_k) \langle l| H_1 + H_1 G(E_k) H_1 |k\rangle_0 \}. \tag{6.3}$$

The imaginary infinitesimal $i\epsilon$ is taken to be positive, but we no longer explicitly indicate it. By an argument identical to the one used in Sec. 5, we can divide the diagrams contributing to $\langle l| \cdots |k\rangle_0$ into two groups, one in which all diagrams have *no* end modifications of the l line, and one in which all have *at least one* end modification of the l line, and then use the definition of the modified propagator to obtain

$$\psi(t) = \sum_k C_k e^{-iE_k t} Z_k^{\frac{1}{2}} \{ |k\rangle + \sum_{l \neq k} |l\rangle S_l'(E_k) \times {}_0\langle l| H_1 + H_1 G(E_k) H_1 |k\rangle_0 \}. \tag{6.4}$$

Now, we separate out the E_k-independent terms as follows,

$$\psi(t) = \sum_k C_k e^{-iE_k t} Z_k^{\frac{1}{2}} \{ |k\rangle + \sum_{l \neq k} |l\rangle S_l'(E_k) \times [T_{lk}^0(E_k) + T_{lk}(E_k)] \}. \tag{6.5}$$

Except for the ket $|k\rangle$, they are all contained in $T_{lk}^0(E_k)$ which has already been defined by (5.9). $T_{lk}(E_k)$ is a similar quantity. It denotes the sum of contributions from all diagrams for going from k to l with no end modifications of either the k line or the l line, and with *at least one* external photon line involving the variables of the incident photons

$$T_{lk}(E_k) = {}_0\langle l| H_1 + H_1 G(E_k) H_1 |k\rangle_0^1, \tag{6.6}$$

where the superscript "one" on the ket $|k\rangle_0^1$ symbolizes the above rule that at least one of the incident photons of state k must interact in each diagram. Comparison of (6.5) with the expression (5.11) for ϕ_k shows that, in extracting terms that are independent of E_k, we have extracted $\phi(t)$ from $\psi(t)$,

$$\psi(t) - \phi(t) = Z_A^{\frac{1}{2}} \sum_l |l\rangle I_l(t), \tag{6.7}$$

$$I_l(t) = \sum_{k \neq l} S_l'(E_k) T_{lk}(E_k) C_k e^{-iE_k t}. \tag{6.8}$$

We have incidentally used the fact that, in the limit $V \to \infty$, the renormalization constant depends only on the state of the atom and not on the state of the radiation field, so that $Z_k = Z_A$.

Note that $\phi(t)$ does not vanish for large $|t|$. Not only is it independent of E_k, but it can also be seen by using the definition (5.7) of ϕ_k that the integration apparently implied by $\sum_k$ in (1.5) is not a c-number integration, but rather a superposition of orthogonal basis vectors. We can now restrict our considerations to the asymptotic properties of the c-number function $I_l(t)$. In the perturbation expansion of $I_l(t)$, the vertex functions are well behaved, so that any singularities contributing to the integration over E_k must occur in the modified propagators. It follows from the definition (2.10) of the modified propagator that it has a pole if and only if the secular equation (3.7) is satisfied. Therefore, the modified propagator for a bare state characterized by an excited atom (unstable state) has no poles along the real axis, while the modified propagator for a stable state has a pole at the corresponding eigenenergy of the total Hamiltonian. To show that the singularity in the latter case is a simple pole, and to determine its residue, we use an analog of Lehmann's theorem[4] which is appropriate to the Brillouin–Wigner form of perturbation theory being used here. We close this section with a review of the theorem.

First recall that, in the limit of infinite quantization volume, no diagrams with external photons contribute to $S_k'(E)$ since these must all be forward-scattering diagrams, so that

$$S_k'(E) = S_A'(E - \omega_k), \tag{6.9}$$

where, as before, ω_k is the energy of the photons in state k, and the change in the argument results

[4] H. Lehmann, Nuovo Cimento 11, 342 (1954); B. S. DeWitt, *The Operator Formalism in Quantum Perturbation Theory*, UCRL-2884 (1955), Chap. 10, p. 155. The latter reference contains a quite detailed discussion of the Brillouin–Wigner perturbation formalism used in the present paper, as well as numerous further references.

from the fact that, in each energy denominator, the expectation value of H_0 has been diminished by ω_k. We use the representation shown in (2.13)

$$S'_A(E) = \langle A | (E - H)^{-1} | A \rangle, \tag{6.10}$$

and invoke the closure property be inserting the identity operator

$$1 = \sum_n |\psi_n\rangle\langle\psi_n| \tag{6.11}$$

into (6.10). The ψ_n are the complete set of eigenstates of H corresponding to the eigenvalues E_n, so

$$S'_A(E) = \frac{Z_A}{E - E_A} + \sum_{n \neq A} \frac{|\langle A | \psi_n \rangle|^2}{E - E_n}. \tag{6.12}$$

We have used the fact, evident from (5.2), that the projection of the physical state vector ψ_A onto the bare state vector $|A\rangle$ is just the renormalization constant $Z_A^{\frac{1}{2}}$. The second term is an integral over a continuum of values of E_n, and is therefore finite for all values of E, provided it is defined on the real axis as a limit approached from above or below. Consequently, the behavior of the modified propagator for a stable state in the vicinity of its characteristic energy is

$$S'_k(E) \to Z_k/(E - E_k), \quad \text{as} \quad E \to E_k. \tag{6.13}$$

7. THE SCATTERING AMPLITUDE

Now, we are in a position to find the asymptotic forms of $I_l(t)$, as given by (6.8). For very large V, the summation over the discrete states k may be replaced by an integral,

$$\sum_k \to \int_k, \tag{7.1}$$

where $\int_k$ is a compact notation for integration over all mode variables,

$$\int_k = \frac{V}{(2\pi)^3} \int d^3k_1 \cdots \frac{V}{(2\pi)^3} \int d^3k_N$$

$$= \frac{V}{(2\pi)^3} \int k_1^2 \, dk_1 \, d\Omega_1 \cdots \frac{V}{(2\pi)^3} \int k_N^2 \, dk_N \, d\Omega_N. \tag{7.2}$$

The integral signs also imply a summation over any discrete variables, such as those for polarization, which have been suppressed. The energy E_k of the system in state k is

$$E_k = E_A + k_1 + \cdots + k_N,$$

where $k_i = |\mathbf{k}_i|$ is the photon energy corresponding to wave vector $\mathbf{k}_i$, and we have taken $\hbar = c = 1$. In the integrations over $k_1, \cdots, k_N$, the poles appear in the modified propagators of the intermediate or final stable states. These states, however, must be not only stable, but also discrete, because a propagator does not provide a pole if it represents a state containing a virtual photon. This is because there is an integration over the energy of the virtual photon with either an imaginary infinitesimal $i\epsilon$ or a prescribed integration contour which keeps the integral finite regardless of the value of E_k. In Fig. 2(a) the state between the second and third vertices is discrete, but in Fig. 2(b) the state in the same relative position is part of a continuum because of the presence of a virtual photon.

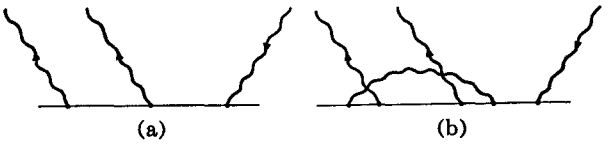


FIG. 2. (a) Feynman diagram illustrating a discrete state between the second and third vertices. (b) Feynman diagram in which the state between the second and third vertices is part of a continuum because of the presence of a virtual photon.

The quantity $T_{lk}(E_k)$ contained in $I_l(t)$ can be written in the following form, which explicitly displays certain of the modified propagators,

$$T_{lk}(E_k) = \sum_n \sum_Q A_{lk}(q_1, \cdots, q_n) \times S'_{q_1}(E_k) \cdots S'_{q_n}(E_k), \tag{7.3}$$

where $\sum_Q$ means sum over q_1 to q_n,

$$\sum_Q = \sum_{q_1} \cdots \sum_{q_n}. \tag{7.4}$$

The quantity

$$A_{lk}(q_1, \cdots, q_n) S'_{q_1}(E_k) \cdots S'_{q_n}(E_k)$$

is the sum of contributions from all diagrams for going from k to l without end modifications, and with precisely n intermediate stable, discrete states (SD states), $q_1, \cdots, q_n$, occurring *after* (to the left of) the last vertex involving the incident mode variables $k_1, \cdots, k_N$ (for brevity, we refer to this vertex as the "last incident-photon vertex", or the LIV). The indicated ordering of the states $q_1, \cdots, q_n$ is the same as the ordering of the propagators in the perturbation term, and of the corresponding atom line in the diagram. The summation $\sum_Q$ can be over all possible values of the q's, since by virtue of its definition, A_{lk} vanishes if any q assumes a value for which the above requirements are not satisfied.

We have already indicated the reason for isolating only the propagators of SD states. To understand the restriction to states occurring after the LIV, one must remember that the integration is

being considered with respect to E_k, the coefficient of t in e^{-iE_kt}. Since the integration in (6.8) is with respect to $k_1, \cdots, k_N$, to take advantage of the asymptotic expansion theorem (6.1), we must first make a transformation so that E_k is one of the integration variables. $I_l(t)$ can be expressed in the form

$$I_l(t) = \int dk_1 \cdots dk_N\, e^{-iE_kt} \sum_n \sum_Q f_Q(k_1, \cdots, k_N) \times S_l'(E_k) S_{q_1}'(E_k) \cdots S_{q_n}'(E_k), \tag{7.5}$$

where

$$f_Q(k_1, \cdots, k_N) = \frac{V}{(2\pi)^3} \int k_1^2\, d\Omega_1 \cdots \frac{V}{(2\pi)^3} \int k_N^2\, d\Omega_N \times \tilde{C}(k_1, \cdots, k_N) A_{lk}(q_1, \cdots, q_n), \tag{7.6}$$

and $\tilde{C}$ is defined by (4.1). We make the transformation of variables,

$$\begin{aligned} x_1 &= E_A + k_1 + \cdots + k_N, \\ x_2 &= k_2, \\ &\vdots \\ x_N &= k_N, \end{aligned} \tag{7.7}$$

to obtain

$$I_l(t) = \sum_n \sum_Q \int dx_1\, e^{-ix_1t} g_Q(x_1) \times S_l'(x_1) \cdots S_{q_N}'(x_1), \tag{7.8}$$

where

$$g_Q(x_1) = \int dx_2 \cdots dx_N\, f_Q(y, x_2, \cdots, x_N) \tag{7.9}$$

and

$$y = x_1 - E_A - x_2 - \cdots - x_N. \tag{7.10}$$

The assertion that poles with respect to E_k appear only in propagators for states occurring after the LIV is equivalent to the assertion that $g_Q(x_1)$ contains no poles with respect to x_1. Any poles of g_Q must be traced to the propagators of f_Q, and these must depend on x_1 alone, otherwise the integration over the other variables according to the prescribed contour will remove the singularity. However, the propagators cannot depend on x_1 alone, unless the variables $x_1, \cdots, x_N$ happen to occur in the linear combination

$$y + x_2 + \cdots + x_N = x_1 - E_A = k_1 + \cdots + k_N. \tag{7.11}$$

The crucial point here is that propagators occurring before the LIV are independent of those mode variables not yet involved in vertices; therefore they cannot depend on the linear combination (7.11), which contains *all* the incident mode variables. This is so because the same quantity, say k_i, appears in both E_k and the expectation value $\langle H_0 \rangle$, and therefore cancels out of the difference $E_k - \langle H_0 \rangle$ in all energy denominators appearing to the right of the first vertex involving k_i.

The residue of the pole at any q_i for a fixed set $Q = (q_1, \cdots, q_n)$ is now found to be

$$R_Q(E_{q_i}) = Z_{q_i} \int_k C_k e^{-iE_kt}\, \delta(E_k - E_{q_i}) \times A_{lk}(q_1, \cdots, q_N) S_l'(E_k) K_Q^i(E_k), \tag{7.12}$$

$$C_k \equiv \tilde{C}(\mathbf{k}_1, \cdots, \mathbf{k}_N), \tag{7.13}$$

$$K_Q^i(E) = [S_{q_1}'(E) \cdots S_{q_{i-1}}'(E)] \times [S_{q_{i+1}}'(E) \cdots S_{q_n}'(E)]. \tag{7.14}$$

We can now utilize the asymptotic expansion theorem to obtain

$$I_l(t) \sim -2\pi i \theta(t) \int_k C_k e^{-iE_kt} \times \Big\{ \epsilon(l) Z_l\, \delta(E_k - E_l) T_{lk}(E_k) + S_l'(E_k) \sum_q Z_q\, \delta(E_k - E_q) \times \sum_n \sum_{j=1}^{n} B_{lk}(q \mid j-1, n-j) \Big\}, \tag{7.15}$$

where

$$\epsilon(l) = \begin{cases} 1, & \text{if } l \text{ is stable,} \\ 0, & \text{if } l \text{ is unstable,} \end{cases} \tag{7.16}$$

and where we have defined the quantity

$$B_{lk}(q \mid j-1, n-j) = \sum A_{lk}(q_1, \cdots, q_{j-1}, q, q_{j+1}, \cdots, q_n) K_Q^i(E_k); \tag{7.17}$$

the summation in (7.17) is over all q_1 to q_n with the exclusion of q_i.

It follows from the definition of A_{lk} that B_{lk} is the sum of contributions from all diagrams without end modifications for going from k to l in which q appears as an intermediate SD state, preceded by precisely $n - j$ intermediate SD states and followed by precisely $j - 1$ intermediate SD states, all following the LIV, and with the modified propagator for state q deleted from the contribution.

Since there is no propagator for q we adopt the convention of drawing no scatterer line for it in diagrams. The diagram is then "closed" at q. For example, Fig. 3(a) shows a standard diagram in which there is a line for state q, but the diagram is closed at k and l. Figure 3(b) shows the corresponding diagram, which is closed at q, indicating that there is no propagator for q. The q line can have no end modifications, since all self-energy terms for state q have already been absorbed into the modified propagator $S_q'(E_k)$, which was later deleted.

The energy-conserving delta function in (7.15) requires that state q have the same energy as state k, i.e., q must be on the energy shell. Furthermore, q must be stable, so that it has all the properties needed to characterize a "real" or physical state like the initial state k, as contrasted to the intermediate virtual states. We refer to such states as "intermediate real" (IR) states. Their mathematical purpose is to serve as the final states in the expression for the scattering amplitude.

The double sum appearing in the second term of (7.15) is

$$D_{lqk} = \sum_{n=1}^{\infty} \sum_{j=1}^{\infty} B_{lk}(q \mid j-1, n-j), \qquad (7.18)$$

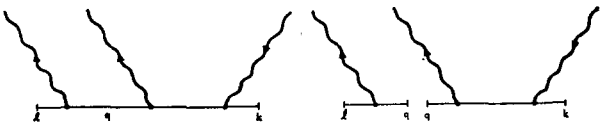
where it is understood that $B_{lk}(q \mid m, n)$ vanishes by definition if $m + n$ is greater than the number of possible intermediate SD states following the LIV. To evaluate the sum, we first invert the order of summation, then make the successive substitutions $r = n - j$, and $s = j - 1$. The result is

$$D_{lqk} = \sum_{s=0}^{\infty} \sum_{r=0}^{\infty} B_{lk}(q \mid s, r). \qquad (7.19)$$

In this form, it is clear from the definition of B_{lk} that D_{lqk} is the sum of contributions from all diagrams going from k to l in which q appears as an IR state following the LIV; in these diagrams the k, q, and l lines have no end modifications, and there is at least one incident-photon vertex. D_{lqk} may also be written in the following form, as can be seen by comparing the diagrammatic definitions,

$$D_{lqk} = T^0_{lq}(E_k) T_{qk}(E_k) \epsilon(q)(1 - \delta_{lq})(1 - \delta_{qk}). \qquad (7.20)$$

The rule that none of the photons of state q may interact in the diagrams of $T^0_{lq}(E_k)$ prevents the possible reabsorption of photons emitted between k and q, so that q is guaranteed to be discrete in spite of the generality of the definition of $T_{qk}(E_k)$. It also guarantees that q occurs after all vertices for absorption of the incident photons of state k,



(a) (b)

FIG. 3. (a) Feynman diagram with a line for state q for which there is a propagator, but "closed" at states k and l for which there are no propagators. (b) Feynman diagram which is closed at state q, indicating that there is no propagator for q.

i.e., that q occurs after the LIV. The presence of the factor $\epsilon(q)$ [defined by (7.16)] guarantees that q is a SD state, and the absence of a propagator for q then makes it an IR state. The factors $(1 - \delta_{lq})(1 - \delta_{qk})$ reflect the fact that q would have had to occur as part of an end modification of either the l or k line if T^0_{ll} or T_{kk} were to appear. The asymptotic expression for $I_l(t)$ now finally reduces to

$$I_l(t) \sim -2\pi i \theta(t) \Big[\epsilon(l) e^{-iE_l t} \times \int_k Z_l \, \delta(E_k - E_l) T_{lk}(E_k) C_k + \sum_{q \neq l} \epsilon(q) e^{-iE_q t} S_l'(E_q) T^0_{lq}(E_q) \times \int_k Z_q \, \delta(E_k - E_q) T_{qk}(E_k) C_k \Big]. \qquad (7.21)$$

Recalling that $I_l(t)$ is related to the time-dependent state vector by

$$\psi(t) - \phi(t) = Z_\Lambda^{\frac{1}{2}} \textstyle\sum_l |l\rangle I_l(t), \qquad (6.7)$$

we see that because of the factor $\theta(t)$ above, $\psi(t)$ satisfies the appropriate boundary condition,

$$\psi(t) \sim \phi(t) \quad \text{for} \quad t \to -\infty. \qquad (7.22)$$

We denote by $\phi_{sc}(t)$ the asymptotic limit of $\psi(t) - \phi(t)$ for large positive t. $\phi_{sc}(t)$ is the scattered wave,

$$\psi(t) - \phi(t) \sim \theta(t) \phi_{sc}(t). \qquad (7.23)$$

Inserting the expression (7.21) for $I_l(t)$ into (6.7), reversing the order of summation over l and q, and utilizing the definition (5.11) of ϕ_k, we find

$$\phi_{sc}(t) = \textstyle\sum_l e^{-iE_l t} A_l \phi_l, \qquad (7.24)$$

where

$$A_l = \int_k (-2\pi i)(Z_l Z_A)^{\frac{1}{2}} \, \delta(E_k - E_l) T_{lk}(E_k) C_k. \qquad (7.25)$$

$\phi_{sc}(t)$ is evidently a free wave packet, and A_l is the probability amplitude for finding the outgoing photons in state l. A_l is a weighted superposition of elements of the T matrix expressed in a Brillouin-

Wigner perturbation expansion, each of which is obtained by taking the sum of contributions from all diagrams for going from initial state k to final state l with no end modifications of the k or l line.

8. SUMMARY

A formal expression has been derived for the field-theoretic scattering amplitude in a Brillouin–Wigner perturbation expansion. Wave packets have been used to introduce the boundary conditions, thereby avoiding the need for invoking the adiabatic hypothesis. To simplify the description of the free wave packets, we have second-quantized only the field of incident particles, but not the scatterer. The principal innovation is the removal of all restrictions other than finiteness on the number of incident quanta. The resulting expression is thus appropriate, for example, to the nonrelativistic description of the scattering of a photon beam of arbitrary intensity by an atom.

The derivation consisted first of producing a set of formally exact scattering eigenvectors, then forming a time-dependent solution to the Schrödinger equation from a weighted superposition of them. An asymptotic expansion of the wave function in reciprocal powers of the time was developed, and it was shown to have the form, in the remote past and future, of a free wave packet of quanta spatially isolated from the scatterer. The scattering amplitude was then identified by a comparison of the initial and final-state wave packets. It was found to be a superposition of elements of the usual Brillouin–Wigner T matrix, each weighted by the corresponding weight factor in the initial-state wave packet.

APPENDIX A

By using wave packets to localize the incident quanta in describing a scattering event, one may develop an approach alternative to the usual one with a steady-state stream of incident and scattered probability flux. In the steady-state picture, the cross section is obtained from the rate of change of probability for finding the system in some group of final states, which is interpreted to be the scattering rate. In the present picture, it is obtained directly from the probability itself for finding the system in the chosen states in the remote future. Thus, it does not matter if the probability and its derivative change radically during the course of the event. One is concerned only with the ultimate disposition of the system.

Actually, it is the final-state probability itself which is most closely related to the data in a scattering experiment. That is, one sets up counters, and by observing the fractional number of events in which quanta are scattered into them one learns the corresponding final-state probabilities. The cross section σ is then defined by the statement that the probability P of a scattering is equal to the number per unit area $1/A$ of incident quanta times the effective scattering area (cross section) presented by the target to these quanta,

$$P = \sigma/A. \tag{A1}$$

To find the probability P, we start with the expression derived in the text for the transition amplitude from state k to state p,

$$A_p = -2\pi i(Z_p Z_A)^{\frac{1}{2}} \int_k \delta(E_p - E_k) T_{pk}(E_k) C_k, \tag{A2}$$

and consider the case of a single incident photon, described by the wave packet

$$C_k = [(2\pi)^3/V]^{\frac{1}{2}} C(\mathbf{k}), \tag{A3}$$

scattering into a final state in which it has wave vector $\mathbf{p}$. The total contribution to the T matrix has two external photon lines in each diagram, so it has the form

$$T_{pk} = [(2\pi)^3/V] W(\mathbf{p} \mid \mathbf{k}). \tag{A4}$$

We are not concerned with renormalization, so we assume the two constants Z_p and Z_A to be absorbed into the T matrix, and write for the transition amplitude,

$$A(\mathbf{p}) = -2\pi i[(2\pi)^3/V]^{\frac{1}{2}} \times \int d^3k\, \delta(p - k) W(\mathbf{p} \mid \mathbf{k}) C(\mathbf{k}). \tag{A5}$$

The probability P that the final-state wave vector $\mathbf{p}$ will occupy a point in some chosen volume Δ of wave-vector space is

$$P = \frac{V}{(2\pi)^3} \int_\Delta d^3p\, |A(\mathbf{p})|^2. \tag{A6}$$

Note that the quantization volume V is canceled when (A5) is inserted into this expression. However, the area A of the photon packet in the plane normal to the propagation direction is still implicit in the description. To make it explicit we write $C(\mathbf{k})$ in a product form,

$$C(\mathbf{k}) = a(k_i, k_j) b(k_\kappa), \tag{A7}$$

$$k_i = \mathbf{k}\cdot\mathbf{i}, \qquad k_j = \mathbf{k}\cdot\mathbf{j}, \qquad k_\kappa = \mathbf{k}\cdot\boldsymbol{\kappa}, \tag{A8}$$

where $\boldsymbol{\kappa}$ is a unit vector in the propagation direction, and $\mathbf{i}$ and $\mathbf{j}$ are mutually orthogonal unit

vectors in the plane normal to κ. We now define

$$2\pi/L = \int dk_i\, dk_j\, a(k_i, k_j), \tag{A9}$$

and interpret $|L|$ to be transverse dimension of the packet. This interpretation is easily justified by evaluating the integral in (A9) for a packet having uniform probability amplitude over an area L^2, zero everywhere else.

The cross section is now given, according to (A1), by

$$\sigma = 4\pi^2\, |L|^2 \int d^3p \times \left| \int d^3k\, \delta(p - k) W(\mathbf{p} \mid \mathbf{k}) C(\mathbf{k}) \right|^2. \tag{A10}$$

In the limit of a quasi-plane, quasi-monochromatic incident packet this reduces to the expression obtained in the steady-state picture. To make the wave quasi-plane we take $a(k_i, k_j)$ to be sharply peaked about $k_i = 0$, and $k_j = 0$. If it is peaked sharply enough we may evaluate the remainder of the integrand at $k_i = k_j = 0$, and remove the latter from the integral over k_i and k_j to obtain

$$\sigma = (2\pi)^4 \int_\Delta d^3p\, |W(\mathbf{p} \mid p\kappa) b(p)|^2. \tag{A11}$$

Finally, we may make the wave packet quasi-monochromatic, too, by taking $b(p)$ to be highly peaked about some energy $p = q$. We can now evaluate the remainder of the integrand at $p = q$, and remove it from the integration over $p = |\mathbf{p}|$. Since $b(p)$ is normalized to unity,

$$\int dp\, |b(p)|^2 = 1, \tag{A12}$$

the result is the same as that obtained by replacing $|b(p)|^2$ by a delta function of energy,

$$\sigma = (2\pi)^4 \int_\Delta d^3p\, |W(\mathbf{p} \mid \mathbf{q})|^2\, \delta(p - q), \qquad \mathbf{q} = q\kappa. \tag{A13}$$

Equation (A13) is the conventional steady-state, plane-wave expression for the cross section. Its validity requires only that the incident packet be sufficiently plane and sufficiently monochromatic for W to be insensitive to the variation with respect to k_i, k_j, and k_κ. The packet is still localized and the scattering probability $P = \sigma/|L|^2$ still depends on the *local* flux, rather than on an assumed uniform flux pervading the entire quantization volume.

APPENDIX B

In this appendix we derive a prescription for evaluating Im $\Sigma_\alpha(E)$, where $\Sigma_\alpha(E)$ is the self-energy function for the system (scatterer plus radiation field) in an eigenstate α of the unperturbed Hamiltonian H_0. In so doing, we show that Im $\Sigma_\alpha(E)$ is proportional to the transition rate out of state α. Described in terms of Feynman diagrams, $\Sigma_\alpha(E)$ is the sum of all possible diagrams for going from state α back to state α without α appearing as an intermediate state. It can be written formally as

$$\Sigma_\alpha(E) = \langle\alpha|\, H_1 + H_1 D(E) H_1\, |\alpha\rangle. \tag{B1}$$

The quantity $D(E)$ operates in the subspace obtained by removing the basis vector $|\alpha\rangle$ from the complete set of eigenvectors of H_0. It is given by

$$D(E) = [f^{-1}(E) - \Lambda H_1 \Lambda]^{-1}, \tag{B2}$$

where Λ is a projection operator off the state α,

$$\Lambda = 1 - |\alpha\rangle\langle\alpha|, \tag{B3}$$

and

$$f(E) = (E - H_0 + i\epsilon)^{-1}\Lambda, \qquad f^{-1}(E) = (E - H_0 + i\epsilon)\Lambda. \tag{B4}$$

The product of reciprocal operators is unity only within the subspace, and is zero on $|\alpha\rangle$,

$$DD^{-1} = ff^{-1} = \Lambda. \tag{B5}$$

To compare (B1) with the diagrammatic definition of $\Sigma_\alpha(E)$, $D(E)$ may be thought of in terms of a perturbation expansion in $\Lambda H_1 \Lambda$,

$$D = f + f\Lambda H_1 \Lambda f + f\Lambda H_1 \Lambda f \Lambda H_1 \Lambda f + \cdots. \tag{B6}$$

The presence of the projection operator Λ guarantees that α will not appear as an intermediate state in the perturbation expansion of $\Sigma_\alpha(E)$. The expectation value of $f(E)$ in any eigenstate of H_0 within the subspace is just the bare propagator for that state. We also make use of the implicit relation

$$D(E) = f(E) + f(E)\Lambda H_1 \Lambda D(E), \tag{B7}$$

which can be proved by formal algebra using the definition (B2), or simply by summing the geometric series in (B6).

It follows from the formal expression (B1) that

$$2i\ \text{Im}\ \Sigma_\alpha(E) = \langle\alpha|\, H_1 [D(E) - D^*(E)] H_1\, |\alpha\rangle. \tag{B8}$$

Furthermore, from the definition (B2) of $D(E)$ it follows that

$$D - D^* = D^*(D^{*-1} - D^{-1})D = -2i\epsilon D^* D, \tag{B9}$$

where, of course, the limit $\epsilon \to 0$ is not to be taken until it serves to define a pole in an integral. Combining (B8) and (B9) leads to

$$\text{Im } \Sigma_\alpha (E) = -\epsilon \langle \alpha | H_1 D^*(E) D(E) H_1 | \alpha \rangle. \quad \text{(B10)}$$

Now, we use the implicit definition (B7) of $D(E)$ to obtain

$$\text{Im } \Sigma_\alpha (E) = -\epsilon \langle \alpha | H_1 [1 + D^*(E) \Lambda H_1 \Lambda] f^*(E) f(E) \times [1 + \Lambda H_1 \Lambda D(E)] H_1 | \alpha \rangle. \quad \text{(B11)}$$

Then, we invoke the closure property by inserting the unit operator $1 = \sum_n |n\rangle\langle n|$ between $f^*(E)$ and $f(E)$. This operation yeilds the expression

$$\text{Im } \Sigma_\alpha (E) = -\epsilon \sum_{n \neq \alpha} |S_n(E) \langle n | H_1 + H_1 D(E) H_1 | \alpha \rangle|^2. \quad \text{(B12)}$$

Since the quantity $\langle n | \cdots | \alpha \rangle$ is the sum of contributions from all possible diagrams for going from α to n (with no intermediate α), *including* diagrams with end modifications of the n line, it may be related as follows to ${}_0\langle n | \cdots | \alpha \rangle$, which is defined in the same way except that diagrams with end modifications of the n line are *excluded*,

$$\langle n | H_1 + H_1 D(E) H_1 | \alpha \rangle = [1 + \Sigma_n (E) S_n'(E)] \times {}_0\langle n | H_1 + H_1 D(E) H_1 | \alpha \rangle. \quad \text{(B13)}$$

Therefore, in view of the implicit definition (2.11) of the modified propagator, (B12) may be written as

$$\text{Im } \Sigma_\alpha (E) = -\sum_{n \neq \alpha} |{}_0\langle n | H_1 + H_1 D(E) H_1 | \alpha \rangle|^2 \epsilon |S_n'(E)|^2. \quad \text{(B14)}$$

Now, according to the definition (2.10) of the modified propagators, we have

$$\epsilon |S_n'(E)|^2 = \epsilon \{ |E - \mathcal{E}_n - \Sigma_n (E) + i\epsilon|^2 \}^{-1}. \quad \text{(B15)}$$

From the discussion of Sec. 3 it follows that $E - \mathcal{E}_n - \Sigma_n (E)$ does not vanish for any E if n is an unstable state. Therefore, we conclude that

$$\lim_{\epsilon \to 0} \epsilon |S_n'(E)|^2$$

vanishes if n is unstable. Furthermore, if n is stable, then from the behavior (6.13) of $S_n'(E)$ near its pole, combined with the definition of the delta function

$$\lim_{\epsilon \to 0} \frac{\epsilon}{x^2 + \epsilon^2} = \pi \delta(x), \quad \text{(B16)}$$

we conclude

$$\lim_{\epsilon \to 0} \epsilon |S_n'(E)|^2 = \begin{cases} 0, & n \text{ unstable}, \\ Z_n \pi \delta(E - E_n), & n \text{ stable}. \end{cases} \quad \text{(B17)}$$

Finally, we have

$$\text{Im } \Sigma_\alpha (E) = -\pi \sum_n |Z_n^{\frac{1}{2}} {}_0\langle n | H_1 + H_1 D(E) H_1 | \alpha \rangle|^2 \delta(E - E_n), \quad \text{(B18)}$$

where n runs *only over stable states* (not including α), as defined by $\text{Im } \Sigma_n (E_n) = 0$. $\text{Im } \Sigma_\alpha (E)$ is seen to be, except for a factor $(-2Z_\alpha)$, the usual expression for the transition rate out of the state α.

For the discussion of Eq. (3.17) the stable and unstable states can actually be distinguished without identifying $-2Z_\alpha \text{ Im } \Sigma_\alpha (E)$ as a transition rate. This can be done by using (B18) for $\text{Im } \Sigma_\alpha (E_\alpha)$, and considering the quantization volume dependences due to the wavefunctions of the particles represented by α. For example, if α represents a ground-state atom plus plane-wave photons (recall that α is an individual eigenstate of H_0), then since the vector potential is proportional to $V^{-\frac{1}{2}}$, $\text{Im } \Sigma_\alpha (E_\alpha)$ is proportional to a factor $1/V$ for each absorbed photon, and therefore vanishes as V tends to infinity. Alternatively, if α represents an excited atom, with or without photons present, then there is at least one possible transition involving no photon absorptions. In this case only the atomic wavefunction, which is normalized in a finite volume, is involved. The term is then independent of V, and does not vanish. Note that the quantization volume dependences of the final states are irrelevant, since they are always canceled by the volume factor in the density of final states.

APPENDIX C

To obtain an asymptotic expansion of

$$I(t) = \int_c f(z) e^{-izt} \, dz, \quad \text{(C1)}$$

as defined in Sec. 6, we first rewrite I as

$$I = R(x_1) \int_c \frac{e^{-izt}}{(z - x_1)} dz + \int_c \left[f(z) - \frac{R(x_1)}{z - x_1} \right] e^{-izt} \, dz. \quad \text{(C2)}$$

The first integral can be split into three parts as follows,

$$J(t) \equiv \int_c \frac{e^{-izt}}{z - x_1} dz = \int_{c'} \frac{e^{-izt}}{z - x_1} dz - \int_{-\infty}^{a} \frac{e^{-izt}}{x - x_1} dz - \int_b^{\infty} \frac{e^{-izt}}{x - x_1} dx, \quad \text{(C3)}$$

where the contour c' goes from $-\infty$ to $+\infty$ and above the pole at x_1. The first integral in (C3) can be evaluated exactly by closing the contour with an infinite semicircle in the upper half-plane for $t < 0$, and in the lower half-plane for $t > 0$, to give

$$\int_{c'} \frac{e^{-izt}}{z - x_1}\, dz = -2\pi i\theta(t). \tag{C4}$$

An asymptotic expression can be generated for the remaining two integrals by successive partial integrations, and it can be readily seen that they go as $1/t$, so that

$$J(t) \sim -2\pi i\theta(t), \tag{C5}$$

with corrections going as $1/t$.

The integrand of the second integral in (C2) is well behaved at $z = x_1$, but still has poles at x_2 and x_n. By performing the same procedure successively at the remaining ploes, we generate the asymptotic approximation

$$I(t) \sim -2\pi i\theta(t) \sum_{n=1}^{N} e^{-ix_n t} R(x_n) + \int_c g(z) e^{-izt}\, dz, \tag{C6}$$

$$g(x) = f(z) - \sum_{m=1}^{N} \frac{R(x_m)}{z - x_m}, \tag{C7}$$

with corrections going as $1/t$. The integrand of the remaining integral is now of bounded variation everywhere along the integration contour, so we may integrate directly along the real axis, and deduce from the Riemann–Lebesgue lemma[5] that it is $O(1/t)$.

[5] E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge University Press, Cambridge, England, 1952), 4th ed., Chap. 9, p. 172.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 8 AUGUST 1966

Concise Derivation of the Formulas for Lattice Thermal Conductivity

ROBERT J. HARDY*
Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Mexico City, Mexico
(Received 1 November 1965)

A perturbation expansion for the correlation-function formula (or Kubo formula) for thermal conductivity was presented in a previous paper. The formulas obtained there for the lowest-order contribution are used here to derive the transport equations for the thermal conductivity for a lattice with imperfections and/or anharmonic forces. The result has the same form as the familiar Boltzmann equation for phonons.

THE theory most frequently used for discussing the thermal conductivity of a crystal lattice with imperfections and anharmonic forces is based on the Boltzmann equation for phonons derived by Peierls.[1] Although plausible, this approach suffers from the usual shortcomings of kinetic theories.[2]

By utilizing the correlation function formula for thermal conductivity, it is possible to establish the theory of heat conduction in lattices on a more rigorous theoretical foundation. A perturbation expansion for the correlation function formula has been given in a previous paper.[3] The formulas obtained there for the contribution to the conductivity (which is of lowest order in the perturbation) are used here to derive the lowest-order "transport equations". These transport equations have the same form as the kinetic-theory Boltzmann equations,[4] but are obtained as a direct consequence of (a) the general assumptions made in the derivation of the correlation function formula, (b) the decision to consider the conductivity only to lowest order, and (c) the choice of the Hamiltonian. The many *ad hoc* assumptions employed in kinetic theory are avoided. Nevertheless, the derivation given here bears many similarities to a derivation of a Boltzmann equation by means of the Pauli equation.[5]

The particular advantage of the approach pre-

* Present address: Department of Physics, University of Oregon, Eugene, Oregon.

[1] R. E. Peierls, Ann. Physik 3, 1055 (1929).

[2] For an enumeration of these shortcomings, see R. J. Hardy, J. Math. Phys. 6, 1749 (1965).

[3] R. J. Hardy, R. J. Swenson, and W. C. Schieve, J. Math. Phys. 6, 1741 (1965).

[4] P. G. Klemens, in *Solid State Physics*, S. Seitz and D. Turnbull, Eds. (Academic Press, Inc., New York, 1958), Vol. 7, pp. 1–98, Eqs. (4.3) and (5.6); and G. Leibfried, in *Handbuch der Physik*, S. Flügge, Ed. (Springer-Verlag, Berlin, 1955), Vol. VII-1, pp. 293–316, Eqs. (90.1), (90.5), and (93.6).

[5] L. Van Hove, in *La theorie des gas neutres et ionises* (John Wiley & Sons, Inc., New York, 1960), pp. 151–183, Eq. (3.34).

where the contour c' goes from $-\infty$ to $+\infty$ and above the pole at x_1. The first integral in (C3) can be evaluated exactly by closing the contour with an infinite semicircle in the upper half-plane for $t < 0$, and in the lower half-plane for $t > 0$, to give

$$\int_{c'} \frac{e^{-izt}}{z - x_1}\, dz = -2\pi i\theta(t). \tag{C4}$$

An asymptotic expression can be generated for the remaining two integrals by successive partial integrations, and it can be readily seen that they go as $1/t$, so that

$$J(t) \sim -2\pi i\theta(t), \tag{C5}$$

with corrections going as $1/t$.

The integrand of the second integral in (C2) is well behaved at $z = x_1$, but still has poles at x_2 and x_n. By performing the same procedure successively at the remaining ploes, we generate the asymptotic approximation

$$I(t) \sim -2\pi i\theta(t) \sum_{n=1}^{N} e^{-ix_n t} R(x_n) + \int_c g(z) e^{-izt}\, dz, \tag{C6}$$

$$g(x) = f(z) - \sum_{m=1}^{N} \frac{R(x_m)}{z - x_m}, \tag{C7}$$

with corrections going as $1/t$. The integrand of the remaining integral is now of bounded variation everywhere along the integration contour, so we may integrate directly along the real axis, and deduce from the Riemann–Lebesgue lemma[5] that it is $O(1/t)$.

[5] E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge University Press, Cambridge, England, 1952), 4th ed., Chap. 9, p. 172.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 8 AUGUST 1966

Concise Derivation of the Formulas for Lattice Thermal Conductivity

ROBERT J. HARDY*
Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Mexico City, Mexico
(Received 1 November 1965)

A perturbation expansion for the correlation-function formula (or Kubo formula) for thermal conductivity was presented in a previous paper. The formulas obtained there for the lowest-order contribution are used here to derive the transport equations for the thermal conductivity for a lattice with imperfections and/or anharmonic forces. The result has the same form as the familiar Boltzmann equation for phonons.

THE theory most frequently used for discussing the thermal conductivity of a crystal lattice with imperfections and anharmonic forces is based on the Boltzmann equation for phonons derived by Peierls.[1] Although plausible, this approach suffers from the usual shortcomings of kinetic theories.[2]

By utilizing the correlation function formula for thermal conductivity, it is possible to establish the theory of heat conduction in lattices on a more rigorous theoretical foundation. A perturbation expansion for the correlation function formula has been given in a previous paper.[3] The formulas obtained there for the contribution to the conductivity (which is of lowest order in the perturbation) are used here to derive the lowest-order "transport equations". These transport equations have the same form as the kinetic-theory Boltzmann equations,[4] but are obtained as a direct consequence of (a) the general assumptions made in the derivation of the correlation function formula, (b) the decision to consider the conductivity only to lowest order, and (c) the choice of the Hamiltonian. The many *ad hoc* assumptions employed in kinetic theory are avoided. Nevertheless, the derivation given here bears many similarities to a derivation of a Boltzmann equation by means of the Pauli equation.[5]

The particular advantage of the approach pre-

* Present address: Department of Physics, University of Oregon, Eugene, Oregon.

[1] R. E. Peierls, Ann. Physik 3, 1055 (1929).

[2] For an enumeration of these shortcomings, see R. J. Hardy, J. Math. Phys. 6, 1749 (1965).

[3] R. J. Hardy, R. J. Swenson, and W. C. Schieve, J. Math. Phys. 6, 1741 (1965).

[4] P. G. Klemens, in *Solid State Physics*, S. Seitz and D. Turnbull, Eds. (Academic Press, Inc., New York, 1958), Vol. 7, pp. 1–98, Eqs. (4.3) and (5.6); and G. Leibfried, in *Handbuch der Physik*, S. Flügge, Ed. (Springer-Verlag, Berlin, 1955), Vol. VII-1, pp. 293–316, Eqs. (90.1), (90.5), and (93.6).

[5] L. Van Hove, in *La theorie des gas neutres et ionises* (John Wiley & Sons, Inc., New York, 1960), pp. 151–183, Eq. (3.34).

sented is that it is concise as well as rigorous. The techniques introduced are used in the following paper[6] to derive the equations which determine the next term in the perturbation expansion. The conciseness of the approach greatly facilitates that calculation. In fact, it is primarily for purposes of the extension to the next order that the lowest-order term is discussed. Finally, it should be pointed out that, in treating the effect on the conductivity of anharmonic scattering, but not of imperfect scattering, it is necessary to assume "factorization" [see Eq. (27)], which enters here as an approximation. It is possible to avoid this approximation,[2] but to do so entails a large increase in the complexity of the derivation.[2]

DERIVATION OF THE TRANSPORT EQUATIONS

It has been shown that the thermal conductivity tensor to lowest order in the perturbation $\lambda H'$ is given by[7]

$$K^{ij} = \frac{V}{kT^2} \sum_{\alpha\beta} f(\alpha) S^i(\alpha) S^j(\beta) \tilde{P}_\epsilon(\beta\alpha), \tag{1}$$

where the Hamiltonian is written as $H = H^\circ + \lambda H'$ and H° is the harmonic Hamiltonian for a perfect lattice. The diagonal parts of the unperturbed energy flux operator $\mathbf{S}(\alpha)$ and of the unperturbed equilibrium density matrix $f(\alpha)$ are given by

$$\mathbf{S}(\alpha) = V^{-1} \sum_k N_k(\alpha) \hbar\omega_k \mathbf{v}_k \tag{2}$$

and

$$f(\alpha) = Z^{-1} e^{-\mathcal{E}(\alpha)/kT}, \tag{3}$$

where $Z \equiv \sum_\alpha \exp[-\mathcal{E}(\alpha)/kT]$. $\mathcal{E}(\alpha)$ is the eigenvalue of H° associated with the eigenvector $|\alpha\rangle$ and the set of occupation numbers $N_k(\alpha)$; the subscripts label the normal modes of H°. V is the volume of the system. $\tilde{P}_\epsilon(\beta\alpha)$ is related to the quantity ${}^\circ X_{E,\epsilon}(\beta\alpha)$ used by Hardy, Swenson, and Schieve[3] and to $|\langle\beta|e^{-iHt/\hbar}|\alpha\rangle|^2$, the probability of a transition from state α to state β in time t, by

$$\tilde{P}_\epsilon(\beta\alpha) = \frac{\hbar}{2\pi} \int_{-\infty}^{\infty} dE\, {}^\circ X_{E,\epsilon}(\beta\alpha) = \int_0^\infty dt\, e^{-\epsilon t} |\langle\beta| e^{-iHt/\hbar} |\alpha\rangle|^2, \tag{4}$$

where the second equality is only valid to lowest order in λ. The parameter ϵ entered through the convergence factor $e^{-\epsilon t}$ in the correlation function formula. The equation for determining ${}^\circ X_{E,\epsilon}$ is equivalent to[7]

$$\epsilon\delta_\epsilon(\mathcal{E}(\beta) - E) Q_{E,\epsilon}(\beta\alpha) - \delta_\epsilon(\mathcal{E}(\beta) - E)\delta_{\beta\alpha} = \frac{2\pi}{\hbar} \sum_\mu |\langle\beta|\, \lambda H'\, |\mu\rangle|^2 \delta_\epsilon(\mathcal{E}(\beta) - E)\delta_\epsilon(\mathcal{E}(\mu) - E) \times [Q_{E,\epsilon}(\mu\alpha) - Q_{E,\epsilon}(\beta\alpha)], \tag{5}$$

where the quantity $Q_{E,\epsilon}$ is related to ${}^\circ X_{E,\epsilon}$ by[8]

$${}^\circ X_{E,\epsilon}(\beta\alpha) \equiv (2\pi/\hbar) Q_{E,\epsilon}(\beta\alpha)\, \delta_\epsilon(\mathcal{E}(\beta) - E), \tag{6}$$

and where $\delta_\epsilon(X) \equiv (\hbar\epsilon/2\pi)[X^2 + (\frac{1}{2}\hbar\epsilon)^2]^{-1}$.

It is convenient to define two new symbols:

$$\phi_k \equiv \lim_{\epsilon\to 0} \lim_{V\to\infty} \int_{-\infty}^{\infty} dE \sum_\beta N_k(\beta) \varphi_{E,\epsilon}(\beta)\, \delta_\epsilon(\mathcal{E}(\beta) - E); \tag{7}$$

$$\varphi_{E,\epsilon}(\beta) \equiv (V/kT^2) \sum_\alpha Q_{E,\epsilon}(\beta\alpha) f(\alpha) \mathbf{S}(\alpha). \tag{8}$$

Using (1), (2), and the above, one obtains

$$\lim_{\epsilon\to 0} \lim_{V\to\infty} K^{ij} = V^{-1} \sum_k \phi_k^i \hbar\omega_k v_k^j, \tag{9}$$

where ϕ_k^i is the ith component of the vector $\boldsymbol{\phi}_k$. Here,

$$V^{-1} \textstyle\sum_k \to (2\pi)^{-3} \sum_s \int d\mathbf{k} \quad \text{as } V \to \infty,$$

where $\mathbf{k}$ and s are the wave vector and polarization index of the normal mode indicated by subscript k.

The task now is to change (5) into a "transport equation" for determining $\boldsymbol{\phi}_k$. To find the thermal conductivity, one solves the appropriate transport equation for $\boldsymbol{\phi}_k$ and substitutes the solution obtained into (9).

To proceed, multiply (5) by

$$(-V/kT^2) N_{k'}(\beta) f(\alpha) \mathbf{S}(\alpha),$$

sum over α and β, integrate over E, use (2), and ignore the term proportional to ϵ. The result is

$$\sum_k \langle N_{k'} N_k\rangle_0 (\hbar\omega_k/kT^2) \mathbf{v}_k = -\frac{V}{kT^2} \frac{2\pi}{\hbar} \int_{-\infty}^{\infty} dE \sum_{\alpha\beta\mu} |\langle\beta|\, \lambda H'\, |\mu\rangle|^2 \times \delta_\epsilon(\mathcal{E}(\beta) - E)\delta_\epsilon(\mathcal{E}(\mu) - E) \times [Q_{E,\epsilon}(\mu\alpha) - Q_{E,\epsilon}(\beta\alpha)] N_{k'}(\beta) f(\alpha) \mathbf{S}(\alpha), \tag{10}$$

[6] R. J. Hardy and W. C. Schieve, J. Math. Phys. 7, 1439 (1966).

[7] Equations (1)–(6) correspond to Eqs. (2.5), (2.8), (2.11), (2.17a), (4.1), and (4.2) in Ref. 3; see also footnote 21 in this reference.

[8] Definition (6) of $Q_{E,\epsilon}(\beta\alpha)$ requires no restricting assumptions about the nature of $X_{E,\epsilon}(\beta\alpha)$, since for $\epsilon > 0$ one has $0 < \delta_\epsilon(X) < \infty$ for all X. However, for $Q_{E,\epsilon}$ to be a useful quantity it must nowhere increase without bound as ϵ decreases and V increases. Although a proof of this is not given, the particular form of (5) suggests that it is probably true for a wide class of perturbations $\lambda H'$.

where $\langle\ \rangle_0$ indicates an equilibrium average [e.g., $\langle N_k\rangle_0 \equiv \sum_\alpha f(\alpha)N_k(\alpha)$]. By exchanging the roles of μ and β in the parts of (10) containing $Q_{E',\epsilon}(\beta\alpha)$ [using $|\langle\beta|\ \lambda H'\ |\mu\rangle|^2 = |\langle\mu|\ \lambda H'\ |\beta\rangle|^2$, making the substitution $E' = \mathcal{E}(\mu) - E$, and taking the limits $V \to \infty$ and $\epsilon \to 0$], one arrives at

$$[d\langle N_{k'}\rangle_0/dT]\mathbf{U}_{k'}$$
$$= -(2\pi/\hbar)\lim_{\epsilon\to 0}\lim_{V\to\infty}\int_{-\infty}^{\infty} dE' \sum_\mu \{\sum_\beta |\langle\beta|\ \lambda H'\ |\mu\rangle|^2 \times \delta_\epsilon(\mathcal{E}(\beta) - \mathcal{E}(\mu) + E') \times [N_{k'}(\beta) - N_{k'}(\mu)]\}\varphi_{\mathcal{E}(\mu)-E',\epsilon}(\mu)\,\delta_\epsilon(E'), \quad (11)$$

where $[d\langle N_{k'}\rangle_0/dT]\mathbf{v}_{k'}$ enters as simply an alternate way of expressing the first member of (10). One must now evaluate the quantity within braces { } in (11).

In the limit $\epsilon \to 0$ the δ_ϵ functions become Dirac δ functions, and Dirac δ functions have the property

$$\delta(\mathcal{E}(\beta) - \mathcal{E}(\mu) + E')\,\delta(E') = \delta(\mathcal{E}(\beta) - \mathcal{E}(\mu))\,\delta(E'). \quad (12)$$

However, one cannot take the limit $\epsilon \to 0$ while the eigenvalue spectrum is discrete ($\hbar\epsilon$ must be large compared to the spacing between adjacent energy levels). The spectrum is discrete for $V < \infty$. The limit $V \to \infty$ cannot be performed until after the sums over α, β, and μ have been carried out because it is difficult to give a precise meaning to these sums at $V = \infty$. However, after these sums have been carried out, the limit $V \to \infty$ can be performed, and doing so changes the discrete set of wave vectors describing the normal modes of H° into a continuous set and changes the sum over $\mathbf{k}$ into an integral [see comment below (9)]. Since the limits $V \to \infty$ and $\epsilon \to 0$ are finally taken, one need only consider those terms in (11) for which $\mathcal{E}(\beta) \cong \mathcal{E}(\mu)$; in particular, only the "energy conserving" part of $\lambda H'$ is needed. All other contributions vanish in the limit. This is indicated symbolically by replacing $\delta_\epsilon[\mathcal{E}(\beta) - \mathcal{E}(\mu) + E']$ with $\delta[\mathcal{E}(\beta) - \mathcal{E}(\mu)]$ in the evaluation of the quantity in braces in (11).

Imperfections

The "energy conserving" part of the perturbation describing the effect of imperfections has the form

$$\lambda T' + \lambda V_2 = \sum_{jk} c_{j-k}a_j a_k^\dagger, \quad (13)$$

where the $a_k^\dagger$ and a_k are the creation and annihilation operators for phonons and where $c_{k-j} = c_{-kj}^*$. The matrix elements of this are

$$\langle\beta|\ \lambda T' + \lambda V_2\ |\mu\rangle = \sum_{jk} c_{j-k}[N_j(\mu)(N_k(\mu)+1)]^{\frac{1}{2}} \times \prod_l \delta_{N_l(\beta),N_l(\mu)-\delta_{jl}+\delta_{kl}}. \quad (14)$$

Thus,

$$\{\sum_\beta |\langle\beta|\ \lambda T' + \lambda V_2\ |\mu\rangle|^2 \times \delta(\mathcal{E}(\beta) - \mathcal{E}(\mu))[N_{k'}(\beta) - N_{k'}(\mu)]\} = \frac{1}{\hbar}\sum_j |c_{j-k'}|^2\,\delta(\omega_j - \omega_{k'})[N_j(\mu) - N_{k'}(\mu)]. \quad (15)$$

The substitution of (15) into (11) yields *the transport equation for imperfection scattering*:

$$[d\langle N_k\rangle_0/dT]\mathbf{v}_{k'} = V^{-1}\sum_l \tilde{\Lambda}_{k'l}\phi_l, \quad (16)$$

where

$$\tilde{\Lambda}_{k'l} \equiv -(2\pi/\hbar^2)\sum_j |c_{j-k'}|^2 \times \delta(\omega_j - \omega_{k'})(V\delta_{jl} - V\delta_{k'l}) \quad (17)$$

and where $V\delta_{kk'} \to (2\pi)^3\delta_{ss'}\delta(\mathbf{k} - \mathbf{k}')$ as $V \to \infty$.

Anharmonic Forces

The "energy conserving" part of the perturbation describing anharmonic forces is

$$\lambda V_3 = \tfrac{1}{2}\sum_{jkl} b_{jkl}(a_j a_{-k}^\dagger a_{-l}^\dagger - a_{-j}^\dagger a_k a_l), \quad (18)$$

where $-b_{j-k-l}^* = b_{jkl} = b_{kjl} =$ etc. Each nonzero element $\langle\beta|\ \lambda V_3\ |\mu\rangle$ contains contributions from two of the terms in the sum in (18). Keeping this in mind, one can show that

$$\{\sum_\beta |\langle\beta|\ \lambda V_3\ |\mu\rangle|^2 \times \delta(\mathcal{E}(\beta) - \mathcal{E}(\mu))[N_{k'}(\beta) - N_{k'}(\mu)]\} = \frac{1}{2\hbar}\sum_{jkl} |b_{jkl}|^2\,\delta(\omega_j - \omega_k - \omega_l) \times [-\delta_{k'j} + \delta_{k'-k} + \delta_{k'-l}] \times [N_j(\mu)N_{-k}(\mu) + N_j(\mu)N_{-l}(\mu) + N_j(\mu) - N_{-k}(\mu)N_{-l}(\mu)]. \quad (19)$$

The substitution of this into (11) yields

$$[d\langle N_{k'}\rangle_0/dT]\mathbf{v}_{k'} = \frac{\pi}{\hbar^2 V}\sum_{jkl} |b_{jkl}|^2\,\delta(\omega_j - \omega_k - \omega_l) \times [V\delta_{k'j} - V\delta_{k'-k} - V\delta_{k'-l}] \times (\phi_{j-k} + \phi_{j-l} + \phi_j - \phi_{-k-l}), \quad (20)$$

where [cf. (7)]

$$\phi_{kl} \equiv \lim_{\epsilon\to 0}\lim_{V\to\infty}\int_{-\infty}^{\infty} dE' \sum_\mu N_k(\mu)N_l(\mu) \times \varphi_{\mathcal{E}(\mu)-E',\epsilon}(\mu)\,\delta_\epsilon(E'). \quad (21)$$

In order to write (20) as a transport equation

one must first express the ϕ_{jk} in terms of the ϕ_j. To do this, consider a density matrix $\rho_l(t)$ whose diagonal elements $\langle\alpha|\ \rho_l(t)\ |\alpha\rangle$ at time $t = 0$ are $(\langle N_l\rangle_0)^{-1}f(\alpha)N_l(\alpha)$, where the factor $(\langle N_l\rangle_0)^{-1}$ is for normalization. Represent the averages formed with $\rho_l(t)$ by

$$\langle A; t\rangle_l \equiv \mathrm{Tr}\, A\rho_l(t). \tag{22}$$

The average of $N_j(\alpha)$ at $t = 0$ is

$$\langle N_j; 0\rangle_l = \langle N_j\rangle_0 + \delta_{jl}(\langle N_l\rangle_0 + 1). \tag{23}$$

Thus, the density matrix $\rho_l(t)$ describes a system in which all of the normal modes of H°, except mode l, are occupied to their equilibrium value at time $t = 0$. The diagonal elements of $\rho_l(t)$ for times $t > 0$ are

$$\langle\mu|\ \rho_l(t)\ |\mu\rangle = \sum_\alpha |\langle\mu|\ e^{-iHt/\hbar}\ |\alpha\rangle|^2\langle\alpha|\ \rho_l(0)\ |\alpha\rangle. \tag{24}$$

Now, using (2), (4), (6), (7), and (8), one can show to lowest order in λ that

$$\phi_k = \lim_{\epsilon\to 0}\lim_{V\to\infty} V^{-1}\sum_l \left[\frac{V}{kT^2}\int_0^\infty dt\, e^{-\epsilon t}\langle N_l\rangle_0\hbar\omega_l\mathbf{v}_l\right] \times \langle N_k; t\rangle_l. \tag{25}$$

A similar relation, but with $\langle N_k; t\rangle_l$ replaced by $\langle N_jN_k; t\rangle_l$, exists for ϕ_{jk}. The following identity is now needed:

$$\langle N_jN_k; t\rangle_l = \langle N_j\rangle_0\langle N_k; t\rangle_l + \langle N_k\rangle_0\langle N_j; t\rangle_l - \langle N_j\rangle_0\langle N_k\rangle_0 + \langle (N_j - \langle N_j\rangle_0)(N_k - \langle N_k\rangle_0); t\rangle_l. \tag{26}$$

Multiply this by the quantity in square brackets in (25), sum over l, and take the limits. Then, by using the fact that the equilibrium average $V^{-1}\sum_l \langle N_l\rangle_0\hbar\omega_l\mathbf{v}_l$ is zero, and by ignoring the last term on the right of (26), one obtains the "factorized" result

$$\phi_{jk} = \phi_j\langle N_k\rangle_0 + \phi_k\langle N_j\rangle_0. \tag{27}$$

The last term on the right of (26) is the average of a product of two deviations from equilibrium for a system excited so that (even at the initial time) only the deviation from equilibrium of one mode (mode l) is significant; consequently, any error introduced by ignoring the time integral of

$$\langle (N_j - \langle N_j\rangle_0)(N_k - \langle N_k\rangle_0); t\rangle_l$$

should be negligibly small.

The substitution of (27) into (20) gives *the transport equation for anharmonic scattering*:

$$[d\langle N_{k'}\rangle_0/dT]\mathbf{v}_{k'} = V^{-1}\sum_l \tilde{\Gamma}_{k'l}\phi_l, \tag{28}$$

where $\Gamma_{k'l}$ is defined by

$$V^{-1}\sum_l \tilde{\Gamma}_{k'l}\phi_l \equiv (\pi/2\hbar^2)\sum_{kl} |b_{jkl}|^2\,(\mathcal{S}_j\mathcal{S}_k\mathcal{S}_l)^{-1} \times [\delta(\omega_j - \omega_k - \omega_l)(\mathcal{S}_j^2\phi_j - \mathcal{S}_k^2\phi_{-k} - \mathcal{S}_l^2\phi_{-l}) + 2\delta(\omega_k - \omega_l - \omega_j)(\mathcal{S}_j^2\phi_j - \mathcal{S}_k^2\phi_{-k} + \mathcal{S}_l^2\phi_l)], \tag{29}$$

where $\mathcal{S}_j \equiv \sinh(\hbar\omega_j/2kT)$, and where $\langle N_j\rangle_0 = [\exp(\hbar\omega_j/kT) - 1]^{-1}$ has been used.

Since

$$|\langle\beta|\ \lambda T' + \lambda V_2 + \lambda V_3\ |\mu\rangle|^2 = |\langle\beta|\ \lambda T' + \lambda V_2\ |\mu\rangle|^2 + |\langle\beta|\ \lambda V_3\ |\mu\rangle|^2,$$

the transport equation for the combined perturbation $\lambda H' = \lambda T' + \lambda V^2 + \lambda V^3$ *is*

$$[d\langle N_{k'}\rangle_0/dT]\mathbf{v}_{k'} = V^{-1}\sum_l (\tilde{\Lambda}_{k'l} + \tilde{\Gamma}_{k'l})\phi_l. \tag{30}$$

DISCUSSION

The lattice thermal conductivity is determined by solving the appropriate transport equation [(16), (28), or (30)] for ϕ_k and substituting the solution obtained into (9). If one multiplies the transport equation by the temperature gradient ∇T and interprets $-\phi_j\cdot\nabla T$ as the derivation of the average number of phonons in mode k from the equilibrium value $\langle N_k\rangle_0$, the resulting equation is identical to the corresponding Boltzmann equation.[4] Similarly, the multiplication of (9) by ∇T yields the kinetic-theory expression for relating the average number of phonons per mode to the heat flux.

Notice the similarity of Eq. (5) to the Laplace transform of the Pauli equation,[5] which is

$$\epsilon\tilde{P}_\epsilon(\beta\alpha) - \delta_{\alpha\beta} = \frac{2\pi}{\hbar}\sum_\mu |\langle\beta|\ \lambda H'\ |\mu\rangle|^2\,\delta(\mathcal{E}(\beta) - \mathcal{E}(\mu)) \times [\tilde{P}_\epsilon(\mu\alpha) - \tilde{P}_\epsilon(\beta\alpha)]. \tag{31}$$

It is because of the similarity between Eqs. (5) and (31) that the present derivation is similar in parts to the derivation of the Boltzmann equation in kinetic theory. Of course, the results obtained here, unlike those of kinetic theory, do not depend on the assumptions and approximations necessary to derive the Pauli equation[9] and its transform (31). In particular, Eq. (5) is based on a simple interaction of an exact expression.[3]

ACKNOWLEDGMENT

The author takes pleasure in thanking Professor Arturo Rosenblueth for his kind hospitality at the Centro de Investigacion y de Estudios Avanzados.

[9] For a discussion of these assumptions and approximations, see Van Hove, Ref. 5.

Higher-Order Corrections to the Lattice Thermal Conductivity*

ROBERT J. HARDY
Department of Physics, University of Oregon, Eugene, Oregon
AND
WILLIAM C. SCHIEVE
United States Naval Radiological Defense Laboratory, San Francisco, California
(Received 1 November 1965)

In a previous paper the correlation function formula, or Kubo formula, was used to derive formulas for the lowest-order (λ^{-2}) and first-order (λ^{-1}) corrections to the lattice thermal conductivity (the Hamiltonian is written $H = H^\circ + \lambda H'$). The formulas obtained there are used here to derive the transport equations for the calculation of the first-order correction to the conductivity. These transport equations have the same homogeneous part as the familiar Boltzmann equation for phonons; however, their inhomogeneous parts are different and depend on the nature of the perturbation. Formulas for these inhomogeneous parts are given for both the scattering due to randomly distributed point imperfections and that due to anharmonic forces. At high temperature, the first-order correction for anharmonic scattering is independent of the temperature.

1. INTRODUCTION

RECENTLY, expressions have been derived for the energy flux operator for a crystal lattice with imperfections and anharmonic forces,[1] and a perturbation expansion has been presented for the correlation function formula for thermal conductivity.[2] The Hamiltonian is written $H = H^\circ + \lambda H'$, where H° is the harmonic Hamiltonian for a perfect lattice and $\lambda H'$ is the perturbation. The lowest-order term in the perturbation expansion is of order λ^{-2} and has been discussed in detail.[3,4] Here, we treat the λ^{-1} -order term.

A perturbation expansion is most appropriate for discussing either the scattering due to anharmonic forces or that due to weak imperfections, i.e., imperfections which disturb the lattice only slightly and which may occur in any concentration. Imperfections which cause strong local disturbances, but occur in low concentration are better treated by expanding in powers of the density of impurities.[5]

The formulas obtained in Refs. 1 and 2 are used in Sec. 2 to derive the transport equation for calculating the λ^{-1}-order part of the conductivity. The techniques employed are similar to those used to derive the lowest-order result in the preceding paper.[4] The homogeneous part of this transport equation—an inhomogeneous integral equation—is identical to the homogeneous part of the lowest-order transport equation, which in turn has the same homogeneous part as the familiar Boltzmann equation.[6] The expressions relating the inhomogeneous part of the λ^{-1}-order transport equation to the parameters characterizing the perturbations are given both for imperfection scattering (Sec. 3) and for anharmonic scattering (Appendix B).

The significance of the results obtained is discussed in Sec. 3. Our present limited understanding of the precise effect of imperfections and anharmonicities on the interparticle forces makes a quantitative comparison of the predicted and measured values of the conductivity virtually impossible, even in lowest order. As a result, a quantitative check of the small corrections predicted here is out of the question. However, it is possible to make qualitative statements about the general properties of these corrections, and this is done in Sec. 3. In particular, it is found that anharmonic forces give rise to a constant first-order correction to the conductivity at high temperatures, which distinguishes it from the T^{-1} dependence (T is temperature) predicted in the lowest-order approximation.

The present work is a good example of the usefulness of the correlation function formulas: The formula for the thermal conductivity makes possible both (a) the proof given in Refs. 3 and 4 that the usual lowest-order equations for determining the con-

* This work was supported in part by the U. S. Office of Naval Research.

[1] R. J. Hardy, Phys. Rev. **132**, 168 (1963); this will be referred to as Ref. 1. The energy flux operator has also been discussed by P. Choquard [Helv. Phys. Acta **36**, 415 (1963)] and L. M. Magid [Phys. Rev. **134**, A158 (1964)].

[2] R. J. Hardy, R. J. Swenson, and W. C. Schieve, J. Math. Phys. **6**, 1741 (1965).

[3] R. J. Hardy, J. Math. Phys. **6**, 1749 (1965); this will be referred to as Ref. 3. Also, see W. C. Schieve and R. L. Peterson, Phys. Rev. **126**, 1458 (1962).

[4] R. J. Hardy, J. Math. Phys. **7**, 1435 (1966).

[5] M. V. Klein, Phys. Rev. **131**, 1500 (1963); A. A. Maradudin, J. Am. Chem. Soc. **86**, 3405 (1964).

[6] By "Boltzmann equation" we mean the lowest-order collision equation obtained in kinetic theory. See G. Leibfried, in *Handbuch der Physik*, S. Flügge, Ed. (Springer-Verlag, Berlin, 1955), Vol. VII-1, Eq. (90.9); or Ref. 3, Eq. (2.22a).

ductivity do not depend on the many *ad hoc* assumptions employed in their derivation in kinetic theory and (b) the derivation given here of the equations for calculating the λ^{-1}-order correction to the lattice thermal conductivity. As discussed in Sec. 3, the λ^{-1}-order correction for anharmonic scattering contains contributions which would be impossible to anticipate by any reasonable extension of kinetic theory.

To clarify the notation and for future reference, we give the equations for determining the lowest-order contribution to the conductivity[4] K^{ij}:

$$K^{ij} = V^{-1} \sum_k \phi_k^i \hbar\omega_k \mathbf{v}_k^j, \tag{1.1}$$

ϕ_k^i is the ith component of the vector $\boldsymbol{\phi}_k$, which is the solution of the lowest-order transport equation

$$[d\langle N_k\rangle_0/dT]\mathbf{v}_k = V^{-1} \sum_l \tilde{\Omega}_{kl}\boldsymbol{\phi}_l. \tag{1.2}$$

V is the volume of the system; T denotes temperature; ω_k, $\mathbf{v}_k$, and $\langle N_k\rangle_0$ are, respectively, the frequency, the group velocity, and the equilibrium average of the occupation number. The normal modes of H° are designated by subscripts j, k, l, etc.; $k = (\mathbf{k}, s)$, where $\mathbf{k}$ is the wave vector and $s = 1, 2, 3$ is the polarization index; only a Bravais lattice is considered. We use $\omega_k = \omega_{-k}$, $\mathbf{v}_k = -\mathbf{v}_{-k}$, and $\mathbf{e}_k = -\mathbf{e}_{-k}$, where $-k = (-\mathbf{k}, s)$ and $\mathbf{e}_k$ is the polarization vector. The form of $\tilde{\Omega}_{kl}$ is determined by the perturbation. In particular, $\tilde{\Omega}_{kl}$ is represented by $\tilde{\Lambda}_{kl}$ when $\lambda H' = \lambda T' + \lambda V_2$ (imperfection scattering): $\lambda T'$ is the perturbation to the kinetic energy; λV_2 is the perturbation to the quadratic part of the potential energy. $\tilde{\Omega}_{kl}$ is represented by $\tilde{\Gamma}_{kl}$ when $\lambda H' = \lambda V_3$ (anharmonic scattering): λV_3 is the cubic term in the potential energy expression. The detailed forms of $\tilde{\Lambda}_{kl}$ and $\tilde{\Gamma}_{kl}$ are given by Ref. 4, (17) and (29), respectively [see comment following (3.9)]. We emphasize that if (1.2) is multiplied by ∇T and if $-\boldsymbol{\phi}_k\cdot\nabla T$ is interpreted as the deviation of the average number of phonons in mode k from $\langle N_k\rangle_0$, then the resulting equation is the same as the usual Boltzmann equation for phonons.[6]

2. DERIVATION OF TRANSPORT EQUATIONS

In this section the techniques introduced in the preceding paper will be used to express the formulas for the λ^{-1} contribution to the conductivity in the form of transport equations. The λ^{-1} contribution to the conductivity is given by [Ref. 2, (4.4)–(4.8), (4.10)]

$$\mathcal{K}^{ij} = \sum_a K_a^{ij}, \tag{2.1}$$

where

$$K_1^{ij} = \frac{V}{kT^2} \sum_{\alpha\beta} f(\alpha)S^i(\alpha) \times \langle\beta|\,\lambda S'^j\,|\beta\rangle \frac{\hbar}{2\pi}\int_{-\infty}^{\infty} dE\,{}^0X_{E,\epsilon}(\beta\alpha), \tag{2.2}$$

$$K_2^{ij} = \frac{V}{kT^2} \sum_{\alpha\beta} f(\alpha) \times \langle\alpha|\,\lambda S'^i\,|\alpha\rangle S^j(\beta) \frac{\hbar}{2\pi}\int_{-\infty}^{\infty} dE\,{}^0X_{E,\epsilon}(\beta\alpha), \tag{2.3}$$

$$K_3^{ij} = \frac{V}{kT^2}\,\mathrm{Re} \sum_{\alpha\beta} \langle\alpha|\,\lambda f_0' S^{\circ j}_{nd}\,|\alpha\rangle \times S^i(\beta) \frac{\hbar}{2\pi}\int_{-\infty}^{\infty} dE\,{}^0X_{E,\epsilon}(\beta\alpha), \tag{2.4}$$

$$K_4^{ij} = \frac{V}{kT^2}\,\mathrm{Re} \sum_{\beta\gamma\sigma} f(\beta)S^i(\beta) \times \langle\sigma|\,S^{\circ i}_{nd}\,|\gamma\rangle \frac{\hbar}{2\pi}\int_{-\infty}^{\infty} dE\,{}^0Y_{E,-\epsilon}(\beta\gamma\sigma), \tag{2.5}$$

$$K_5^{ij} = \frac{V}{kT^2}\,\mathrm{Re} \sum_{\beta\gamma\sigma} f(\sigma) \times \langle\sigma|\,S^{\circ i}_{nd}\,|\gamma\rangle S^j(\beta) \frac{\hbar}{2\pi}\int_{-\infty}^{\infty} dE\,{}^0Y_{E,+\epsilon}(\beta\gamma\sigma), \tag{2.6}$$

and

$$K_6^{ij} = \frac{V}{kT^2} \sum_{\alpha\beta} f(\alpha)S^i(\alpha)S^j(\beta) \times \frac{\hbar}{2\pi}\int_{-\infty}^{\infty} dE\,{}^1X_{E,\epsilon}(\beta\alpha). \tag{2.7}$$

The effect of the interactions between the normal modes of H° is included in the quantities ${}^0X_{E,\epsilon}$, ${}^0Y_{E,\pm\epsilon}$, and ${}^1X_{E,\epsilon}$. The parameter ϵ entered through the convergence factor $e^{-\epsilon t}$ in the correlation function formula; the limit $\epsilon \to 0$ may be taken only after performing the limit $V \to \infty$.[4] As indicated in Ref. 2, K_6^{ij} not only determines part of the λ^{-1} contribution to the conductivity, but it also determines the entire lowest-order contribution. In addition, it produces contributions proportional to λ^n for all value of $n \geqq 0$; these are to be neglected.[2] $\lambda\mathbf{S}'$ and $\lambda f_0'$ are the λ-order terms in the perturbation expansions of the energy flux and the equilibrium density matrix, respectively; that is, they are the second terms in the series $\mathbf{S} = \mathbf{S}^0 + \lambda\mathbf{S}' + \lambda^2\mathbf{S}'' + \cdots$ and $f_0 = f_0{}^\circ + \lambda f_0' + \cdots$. The essential parts of the expressions for $\lambda S' = \lambda S_3' + \lambda S_4'$ and for $\lambda f_0'$ are given by (B.13) and (2.47). $\langle\sigma|\,\mathbf{S}^\circ_{nd}\,|\gamma\rangle$ designates

the *nondiagonal* elements of the lowest-order part of the energy flux: $\mathbf{S}^{\circ}_{nd} = \mathbf{S}^{\circ}_{2,nd} + \mathbf{S}^{\circ}_{3}$ [see (3.3) and (B5)]. $\mathbf{S}(\alpha)$ (which is $\langle\alpha|\ \mathbf{S}^{\circ}\ |\alpha\rangle$ in Ref. 2) represents the *diagonal* elements of the lowest-order part of the flux. $f(\alpha)$ (which is $\langle\alpha|\ f_0^{\circ}\ |\alpha\rangle$ in Ref. 2) is the entire lowest-order part of the equilibrium density matrix, which is diagonal. One has [Ref. 1, (3.30)]

$$\mathbf{S}(\alpha) = V^{-1} \sum_k N_k(\alpha)\hbar\omega_k \mathbf{v}_k \tag{2.8}$$

and

$$f(\alpha) = Z^{-1} e^{-\mathcal{E}(\alpha)/kT}, \tag{2.9}$$

where

$$Z = \sum_\alpha e^{-\mathcal{E}(\alpha)/kT}; \qquad \mathcal{E}(\alpha) = \sum_k N_k(\alpha)\hbar\omega_k. \tag{2.10}$$

$\mathcal{E}(\alpha)$ is the unperturbed energy eigenvalue associated with the eigenvector $|\alpha\rangle$, and $N_k(\alpha)$ is the occupation number of mode k when the system is in state α. Re indicates that the real part is to be taken, and k is Boltzmann's constant.

Using the symmetric property of the conductivity tensor, we express the limiting value of $\mathcal{K}^{ij}$ as

$$\lim_{\epsilon\to 0}\lim_{V\to\infty} \mathcal{K}^{ij} = \tfrac{1}{2}[K_1^{ij} + V^{-1}\sum_k \sum_{a=2}^{6} \mathrm{Re}\ \phi^i_{a,k}\hbar\omega_k \mathbf{v}^j_k] + \text{transpose}, \tag{2.11}$$

where the $\phi_{a,k}$ are defined as follows[7]:

$$\phi_{2,k} \equiv \frac{V}{kT^2}\sum_{\alpha\beta} f(\alpha)\langle\alpha|\ \lambda\mathbf{S}'\ |\alpha\rangle N_k(\beta) \times \int_{-\infty}^{\infty} dE\ \delta_\epsilon(\mathcal{E}(\beta) - E)Q_{E,\epsilon}(\beta\alpha\alpha); \tag{2.12}$$

$$\phi_{3,k} \equiv \frac{V}{kT^2}\sum_{\alpha\beta} \langle\alpha|\ \lambda f_0'\mathbf{S}^{\circ}_{nd}\ |\alpha\rangle N_k(\beta) \times \int_{-\infty}^{\infty} dE\ \delta_\epsilon(\mathcal{E}(\beta) - E)Q_{E,\epsilon}(\beta\alpha\alpha); \tag{2.13}$$

$$\phi_{4,k} \equiv \frac{V}{kT^2}\sum_{\beta\gamma\sigma} f(\beta)N_k(\beta)\langle\sigma|\ \mathbf{S}^{\circ}_{nd}\ |\gamma\rangle \times \int_{-\infty}^{\infty} dE\ \delta_\epsilon(\mathcal{E}(\beta) - E)Q_{E,-\epsilon}(\beta\gamma\sigma); \tag{2.14}$$

$$\phi_{5,k} \equiv \frac{V}{kT^2}\sum_{\beta\gamma\sigma} f(\sigma)\langle\sigma|\ \mathbf{S}^{\circ}_{nd}\ |\gamma\rangle N_k(\beta) \times \int_{-\infty}^{\infty} dE\ \delta_\epsilon(\mathcal{E}(\beta) - E)Q_{E,+\epsilon}(\beta\gamma\sigma); \tag{2.15}$$

$$\phi_{6,k} \equiv \frac{V}{kT^2}\sum_{\alpha\beta} f(\alpha)\mathbf{S}(\alpha)N_k(\beta) \times \int_{-\infty}^{\infty} dE\ \delta_\epsilon(\mathcal{E}(\beta) - E)R_{E,\epsilon}(\beta\alpha). \tag{2.16}$$

$Q_{E,\pm\epsilon}$, $R_{E,\epsilon}$, and δ_ϵ are defined by[8]

$$(\hbar/2\pi)^0 X_{E,\epsilon}(\beta\alpha) \equiv \delta_\epsilon(\mathcal{E}(\beta) - E)Q_{E,+\epsilon}(\beta\alpha), \tag{2.17}$$

$$(\hbar/2\pi)^0 Y_{E,\pm\epsilon}(\beta\gamma\sigma) \equiv \delta_\epsilon(\mathcal{E}(\beta) - E)Q_{E,\pm\epsilon}(\beta\gamma\sigma), \tag{2.18}$$

$$(\hbar/2\pi)^1 X_{E,\epsilon}(\beta\alpha) \equiv \delta_\epsilon(\mathcal{E}(\beta) - E)R_{E,\epsilon}(\beta\alpha), \tag{2.19}$$

and

$$\delta_\epsilon(X) \equiv (\tfrac{1}{2}\hbar\epsilon/\pi)[X^2 + (\tfrac{1}{2}\hbar\epsilon)^2]^{-1}. \tag{2.20}$$

Although the limits $V \to \infty$ and $\epsilon \to 0$ are only explicitly indicated in the first member of (2.11), it is to be understood here (and in the following) that the $\phi_{a,k}$'s (and K_1^{ij}) designate the limiting value of the quantity defined. The limits $V \to \infty$ and $\epsilon \to 0$ are to be taken only after all sums over states α, β, γ, etc. have been performed[4]; $\delta_\epsilon(X)$ becomes a Dirac δ function in this limit.

The equation for determining $Q_{E,\pm\epsilon}$ is[8]

$$\epsilon\delta_\epsilon(\mathcal{E}(\beta) - E)Q_{E,\pm\epsilon}(\beta\gamma\sigma) - \delta_{\beta\gamma}\delta_{\gamma\sigma}\,\delta_\epsilon(\mathcal{E}(\beta) - E) - \delta_\epsilon(\mathcal{E}(\beta) - E)\,\lambda V_{E,\pm\epsilon}(\beta\gamma\sigma) = \frac{2\pi}{\hbar}\sum_\mu |\langle\beta|\ \lambda H'\ |\mu\rangle|^2 \times \delta_\epsilon(\mathcal{E}(\beta) - E)\delta_\epsilon(\mathcal{E}(\mu) - E) \times [Q_{E,\pm\epsilon}(\mu\gamma\sigma) - Q_{E,\pm\epsilon}(\beta\gamma\sigma)], \tag{2.21}$$

where $\lambda V_{E,\pm\epsilon}$ is given by (2.48); the equations for determining ${}^1X_{E,\epsilon}$ are (2.19) ,(2.52), and (A2). Note that $\lambda V_{E,\pm\epsilon}$ does not enter into the calculation of $Q_{E,\pm\epsilon}(\beta\alpha\alpha)$ [remember that $\langle\alpha|\ \lambda H'\ |\alpha\rangle = 0$, see (2.48)] and that the term $\delta_{\beta\gamma}\delta_{\gamma\sigma}\delta_\epsilon(\mathcal{E}(\beta) - E)$ does not enter into the calculation of $Q_{E,\pm\epsilon}(\beta\gamma\sigma)$ when $\gamma \neq \sigma$.

Equations for K_1^{ij}

The diagonal part of the contribution to the energy flux due to λV_3 is [see (B12), (B16), and (B17)]

$$\langle\beta|\ \lambda\mathbf{S}'\ |\beta\rangle = V^{-2}\sum_{kl} \mathbf{D}_{kl}N_l(\beta) \times [N_k(\beta) + N_{-k}(\beta) + 1]. \tag{2.22}$$

As explained in Sec. 3, the contribution to $\lambda\mathbf{S}'$

[7] Note that $\phi_{4,k}{}^i$ is related to $K_4{}^{ii}$ in the same way that the $\phi_{a,k}{}^i$ for $a = 2$, 3, 5, and 6 are related to $K_a{}^{ij}$.

[8] Equations (2.17), (2.18), and (2.21) are equivalent to Eqs. (4.2) and (4.9) of Ref. 2, while (2.19) and (2.52) are equivalent to Eq. (4.11) of Ref. 2. Also, see footnote 21 in Ref. 2 and footnote 8 in Ref. 4.

from $\lambda T' + \lambda V_2$ has no diagonal part. Using (2.2) and (2.17), one can show that

$$K_1^{ij} = V^{-2} \sum_{kl} D_{kl}^{i}(\phi_{lk}^{i} + \phi_{l-k}^{i} + \phi_l^{i}), \qquad (2.23)$$

where ϕ_k and ϕ_{kl} are defined as in (7) and (21) of Ref. 4:

$$\phi_k \equiv \frac{V}{kT^2} \sum_{\alpha\beta} f(\alpha)\mathbf{S}(\alpha)N_k(\beta) \times \int_{-\infty}^{\infty} dE\ \delta_\epsilon(\mathcal{E}(\beta) - E)Q_{E,\epsilon}(\beta\alpha\alpha); \qquad (2.24)$$

$$\phi_{kl} \equiv \frac{V}{kT^2} \sum_{\alpha\beta} f(\alpha)\mathbf{S}(\alpha)N_k(\beta)N_l(\beta) \times \int_{-\infty}^{\infty} dE\ \delta_\epsilon(\mathcal{E}(\beta) - E)Q_{E,\epsilon}(\beta\alpha\alpha). \qquad (2.25)$$

We now need the "factorized" relation [Ref. 4, (27)]

$$\phi_{kl} = \phi_k\langle N_l\rangle_0 + \phi_l\langle N_k\rangle_0. \qquad (2.26)$$

In Ref. 4 this relation was derived by neglecting the average of the product of deviations

$$(N_k(\mu) - \langle N_k\rangle_0)(N_l(\mu) - \langle N_l\rangle_0)$$

on the basis that for most times this product is small compared to the averages of $N_k(\mu)N_l(\mu)$ and $N_k(\mu)\langle N_l\rangle_0$; the average was formed with a density matrix which describes the evolution in time of a system in which all normal modes but one are initially in equilibrium. An alternate justification of relations such as (2.26) has been given by DeVault and McLennan.[9]

The substitution of (2.26) into (2.23) gives

$$K_1^{ij} = V^{-1} \sum_k \phi_k^i \hbar\omega_k V_k^i, \qquad (2.27)$$

where

$$\mathbf{V}_k \equiv (\hbar\omega_k V)^{-1} \times \sum_l [(2\mathbf{D}_{lk} + \mathbf{D}_{kl} + \mathbf{D}_{-kl})\langle N_l\rangle_0 + \mathbf{D}_{lk}]. \qquad (2.28)$$

ϕ_k is determined by the transport equation (1.2).

Transport Equation for $\phi_{2,k}$

The derivation of the transport equation for $\phi_{2,k}$ follows through the same steps as the derivation of the lowest-order transport equation in Ref. 4. One simply uses $\langle\alpha|\ \lambda\mathbf{S}'\ |\alpha\rangle$ here instead of $\mathbf{S}(\alpha)$. For a more detailed discussion of the derivation see Ref. 4.

[9] G. P. DeVault and J. A. McLennan, Phys. Rev. **138**, A856 (1965). See their discussion of "factorization."

One proceeds by multiplying (2.21) with γ and σ set equal to α by

$$f(\alpha)\langle\sigma|\ \lambda\mathbf{S}'\ |\alpha\rangle N_k(\beta)$$

and by summing over α and β. Then, exchanging the summation indices μ and β in the terms containing $Q_{E,\epsilon}(\beta\sigma\sigma)$, while leaving the indices unchanged in the terms containing $Q_{E,\epsilon}(\mu\alpha\alpha)$, using $\langle\beta|\ \lambda H'\ |\mu\rangle|^2 = |\langle\mu|\ \lambda H'\ |\beta\rangle|^2$, introducing $E' = \mathcal{E}(\mu) - E$, multiplying by $(-V/kT^2)$, taking the limits $V \to \infty$ and $\epsilon \to 0$, and integrating over E', one arrives at [cf. Ref. 4, (11)]

$$\mathbf{I}_{2,k} = -(2\pi/\hbar) \int_{-\infty}^{\infty} dE'\ \delta_\epsilon(E') \times \sum_\mu \{\sum_\beta |\langle\beta|\ \lambda H'\ |\mu\rangle|^2 \times \delta_\epsilon(\mathcal{E}(\beta) - \mathcal{E}(\mu) + E')(N_k(\beta) - N_k(\mu))\} \times (V/kT^2) \sum_\alpha Q_{\mathcal{E}(\mu)-E',\epsilon}(\mu\alpha\alpha)f(\alpha)\langle\alpha|\ \lambda\mathbf{S}'\ |\alpha\rangle, \qquad (2.29)$$

where $\mathbf{I}_{2,k}$ is defined by

$$\mathbf{I}_{2,k} \equiv (V/kT^2) \sum_\beta f(\beta)\langle\beta|\ \lambda\mathbf{S}'\ |\beta\rangle N_k(\beta). \qquad (2.30)$$

The second member of (2.29) is the same as that of Ref. 4, (11) except that $\langle\alpha|\ \lambda\mathbf{S}'\ |\alpha\rangle$ occurs in (2.29) instead of $\mathbf{S}(\alpha)$. The same difference distinguishes $\phi_{2,k}$ and ϕ_k. Consequently, the arguments used in Ref. 4 to derive the lowest-order transport equation (1.2) also apply here. Using these arguments, we obtain the following *transport equation for* $\phi_{2,k}$:

$$\mathbf{I}_{2,k} = V^{-1} \sum_l \tilde{\Omega}_{kl}\phi_{2,l}. \qquad (2.31)$$

The form of $\tilde{\Omega}_{kl}$ for imperfection scattering ($\tilde{\Lambda}_{kl}$) and for anharmonic scattering ($\tilde{\Gamma}_{kl}$) are given by Ref. 4, (17) and (29), respectively.

Just as (2.26) was needed in Ref. 4 to obtain the lowest-order transport equation for anharmonic scattering, the following, analogous relation is needed here[10]:

$$\phi_{a,jk} = \phi_{a,j}\langle N_k\rangle_0 + \phi_{a,k}\langle N_j\rangle_0, \qquad (2.32)$$

where $a = 2$. We emphasize that (2.32) is not needed

[10] The arguments made in Ref. 4 to justify (2.26) are easily generalized to apply to (2.32) with $a = 2, 3$, or 6 but are more difficult to generalize with $a = 4$ or 5. This is because the quantities which would replace the density matrix $\rho_l(t)$ in Ref. 4 have zero trace for $a = 4$ or 5 and, hence, cannot be interpreted as density matrices. However, even for $a = 4$ or 5 the validity of the above equation still depends on the time integral of a type of "average" of the product of derivations $(N_k(\mu) - \langle N_k\rangle_0)(N_l(\mu) - \langle N_l\rangle_0)$ being small compared to the same "average" of $N_k(\mu)\ N_l(\mu)$ and $N_k(\mu)\ \langle N_l\rangle_0$. Consequently, we assume (2.32) with $a = 4$ and 5 as well as with 2, 3, and 6. See Appendix A for more comments on the case $a = 6$.

for imperfection scattering acting alone (i.e., for $\lambda H' = \lambda T' + \lambda V_2$). Note also that in Ref. 3 it was possible to obtain the lowest-order transport equation, even for anharmonic scattering, without using (2.26). However, this increase in rigor was obtained at the expense of a large increase in the length of the derivation. Presumably, a procedure similar to that of Ref. 3 could be used to avoid using (2.32) here.

Simplification of $\mathbf{I}_{2,k}$

The substitution of (2.22) into (2.30) gives

$$\mathbf{I}_{2,k'} = (kT^2V)^{-1} \sum_{kl} \mathbf{D}_{kl}\langle N_l(N_k + N_{-k} + 1)N_{k'}\rangle_0, \tag{2.33}$$

where

$$\langle N_i \cdots N_k\rangle_0 \equiv \sum_{\alpha} f(\alpha)N_i(\alpha) \cdots N_k(\alpha). \tag{2.34}$$

A simple lemma is now needed: Consider quantities of the form

$$\mathbf{J}_{k'} \equiv V^{-n+1} \sum_{j_1 \cdots j_n} \mathbf{A}_{j_1 \cdots j_n}\langle N_{j_1} \cdots N_{j_n}N_{k'}\rangle_0. \tag{2.35}$$

Assume that $A_{j_1 \cdots j_n}$ is an odd function of the wave vectors, i.e., that

$$\mathbf{A}_{j_1 \cdots j_n} = -\mathbf{A}_{-j_1 \cdots -j_n}. \tag{2.36}$$

It follows from (2.9), (2.10), (2.34), and $\omega_k = \omega_{-k}$ that the averages $\langle N_i \cdots N_k\rangle_0$ are independent of the signs before their subscripts, i.e., that

$$\langle \cdots N_iN_k \cdots\rangle_0 = \langle \cdots N_{-i}N_k \cdots\rangle_0 = \text{etc.}$$

Using this fact, condition (2.36), and $\langle N_k\rangle_0 = [\exp(\hbar\omega_k/kT) - 1]^{-1}$, one obtains

$$\begin{aligned}\mathbf{J}_{k'} &= V^{-n+1} \sum_{j_1 \cdots j_n} \mathbf{A}_{j_1 \cdots j_n}(\langle N_{j_1} \cdots N_{j_n}N_{k'}\rangle_0 \\ &\quad - \langle N_{j_1} \cdots N_{j_n}\rangle_0\langle N_{k'}\rangle_0) \\ &= V^{-n} \sum_{j_1 \cdots j_n} \{V\delta_{k'j_1}\langle N_{j_2} \cdots N_{j_n}\rangle_0 \\ &\quad + \cdots + V\delta_{k'j_n}\langle N_{j_1} \cdots N_{j_{n-1}}\rangle_0\} \\ &\quad \times [d\langle N_{k'}\rangle_0/dT](kT^2/\hbar\omega_{k'}),\end{aligned} \tag{2.37}$$

where the limit $V \to \infty$ is understood. The terms in $\mathbf{J}_{k'}$, from the special cases when k' equals two or more of the summed-over subscripts, vanish in the limit $V \to \infty$ [see (B1)].

According to (B18) one has $\mathbf{D}_{kl} = -\mathbf{D}_{-k-l}$; thus, the above lemma applies to (2.33). Using the lemma, we write the inhomogeneous term $\mathbf{I}_{2,k'}$ as

$$\mathbf{I}_{2,k'} = [d\langle N_{k'}\rangle_0/dT]\mathbf{V}_{k'}, \tag{2.38}$$

where $\mathbf{V}_{k'}$ is defined by (2.28). The temperature derivative has been introduced solely to bring out the analogy between $\mathbf{I}_{2,k'}$ and the inhomogeneous term in the lowest-order transport equation.

Transport Equation for $\phi_{3,k}$

In analogy to the treatment of $\phi_{2,k}$, multiply (2.21) with γ and σ set it equal to α by

$$\langle\alpha|\ \lambda f_0'\mathbf{S}^\circ_{nd}\ |\alpha\rangle N_k(\beta),$$

and sum over α and β. After exchanging summation indices, etc., one obtains an equation that is similar to (2.29) but with $f(\alpha)\langle\alpha|\ \lambda\mathbf{S}'\ |\alpha\rangle$ replaced by $\langle\alpha|\ \lambda f_0'\mathbf{S}^\circ_{nd}\ |\alpha\rangle$. The same difference distinguishes $\phi_{2,k}$ and $\phi_{3,k}$. Thus, just as in the derivation of (2.31), the arguments used in Ref. 4 to derive (1.2) apply here. Their application gives the *transport equation for* $\phi_{3,k}$:

$$\mathbf{I}_{3,k} = V^{-1} \sum_{l} \tilde{\Omega}_{kl}\phi_{3,l}, \tag{2.39}$$

where the inhomogeneous term is

$$\begin{aligned}\mathbf{I}_{3,k} \equiv (V/kT^2) &\int_{-\infty}^{\infty} dE\ \delta_\epsilon(\mathcal{E}(\beta) - E) \\ &\times \langle\beta|\ \lambda f_0'\mathbf{S}^\circ_{nd}\ |\beta\rangle N_k(\beta).\end{aligned} \tag{2.40}$$

Transport Equation for $\phi_{4,k}$

Again, following the analogy of the treatment of $\phi_{2,k}$, multiply (2.21) with $\gamma \neq \sigma$ by

$$(-V/kT^2)f(\beta)N_k(\beta)\langle\sigma|\ \mathbf{S}^\circ_{nd}\ |\gamma\rangle,$$

sum over β, σ, and γ, introduce $E' = \mathcal{E}(\mu) - E$, take the limits $V \to \infty$ and $\epsilon \to 0$, and integrate over E'; one obtains

$$\begin{aligned}\mathbf{I}_{4,k} &= -(2\pi/\hbar) \int_{-\infty}^{\infty} dE'\ \delta_\epsilon(E') \\ &\times \sum_{\mu} \sum_{\beta} |\langle\beta|\ \lambda H'\ |\mu\rangle|^2\ \delta_\epsilon(\mathcal{E}(\beta) - \mathcal{E}(\mu) + E') \\ &\times [N_k(\beta)f(\beta) - N_k(\mu)f(\mu)] \\ &\times (V/kT^2) \sum_{\sigma\gamma} Q_{\mathcal{E}(\mu)-E',-\epsilon}(\mu\gamma\sigma)\ \langle\sigma|\ \mathbf{S}^\circ_{nd}\ |\gamma\rangle,\end{aligned} \tag{2.41}$$

where

$$\begin{aligned}\mathbf{I}_{4,k} \equiv (V/kT^2) &\sum_{\beta\gamma\sigma} \int_{-\infty}^{\infty} dE\ \delta_\epsilon(\mathcal{E}(\beta) - E) \\ &\times \lambda V_{E,-\epsilon}(\beta\gamma\sigma)f(\beta)N_k(\beta)\langle\sigma|\ \mathbf{S}^\circ_{nd}\ |\gamma\rangle.\end{aligned} \tag{2.42}$$

In the limits $V \to \infty$ and $\epsilon \to 0$ the δ_ϵ functions become Dirac δ functions, which are such that $\delta(E')\delta(\mathcal{E}(\beta) - \mathcal{E}(\mu) + E') = \delta(E')\delta(\mathcal{E}(\beta) - \mathcal{E}(\mu))$. Thus, only states for which $\mathcal{E}(\beta) = \mathcal{E}(\mu)$ will contribute in the end to the second member of (2.41). Since $f(\alpha)$ depends on α only through $\mathcal{E}(\alpha)$ [see (2.9)], the difference $[N_k(\beta)f(\beta) - N_k(\mu)f(\mu)]$ can

be replaced by $[N_k(\beta) - N_k(\mu)]f(\mu)$. After making this replacement, the second members of (2.41) and (2.29) are the same except that, instead of $Q_{E,\epsilon}(\mu\alpha\alpha)\, f(\alpha)\langle\alpha|\, \lambda S' \,|\alpha\rangle$, one has

$$f(\mu)Q_{E,-\epsilon}(\mu\gamma\sigma)\langle\sigma|\, \mathbf{S}^{\circ}{}_{nd} \,|\gamma\rangle;$$

the same difference distinguishes $\phi_{2,k}$ and $\phi_{4,k}$. Consequently, by repeating the arguments used to deduce (2.31) from (2.29), one obtains *the transport equation for* $\phi_{4,k}$:

$$\mathbf{I}_{4,k} = V^{-1} \sum_l \tilde{\Omega}_{kl}\phi_{4,l}. \tag{2.43}$$

Transport Equations for $\phi_{5,k}$

Multiply (2.21) by $f(\sigma)\langle\sigma|\, \mathbf{S}^{\circ}{}_{nd} \,|\gamma\rangle N_k(\beta)$, sum over, σ, γ, and β, exchange summation indices, etc. The result obtained is similar to (2.29) but with $Q_{E,\epsilon}(\mu\alpha\alpha)f(\alpha)\langle\alpha|\, \lambda\mathbf{S}' \,|\alpha\rangle$ replaced by

$$Q_{E,+\epsilon}(\mu\gamma\sigma)f(\sigma)\langle\sigma|\, S^{\circ}{}_{nd} \,|\gamma\rangle.$$

By continuing through the steps used to derive (2.31), one obtains the *transport equation for* $\phi_{5,k}$:

$$\mathbf{I}_{5,k} = V^{-1} \sum_l \tilde{\Omega}_{kl}\phi_{5,l}, \tag{2.44}$$

where the inhomogeneous term is

$$\mathbf{I}_{5,k} \equiv (V/kT^2) \int_{-\infty}^{\infty} dE\; \delta_{\epsilon}(\mathcal{E}(\beta) - E)f(\sigma)\langle\sigma|\, \mathbf{S}^{\circ}{}_{nd} \,|\gamma\rangle \times N_k(\beta)\, \lambda V_{E,+\epsilon}(\beta\gamma\sigma). \tag{2.45}$$

Simplification of $\mathbf{I}_{3,k} + \mathbf{I}_{4,k} + \mathbf{I}_{5,k}$

Since only the sum $\phi_{3,k} + \phi_{4,k} + \phi_{5,k}$ is needed in (2.11) for the evaluation of $\mathfrak{K}^{ij}$. The transport equations (2.39), (2.43), and (2.44) can be combined into a single transport equation:

$$\mathbf{I}_{3,k} + \mathbf{I}_{4,k} + \mathbf{I}_{5,k} = V^{-1} \sum_l \tilde{\Omega}_{kl}(\phi_{3,k} + \phi_{4,k} + \phi_{5,k}). \tag{2.46}$$

The combined inhomogeneous term $\mathbf{I}_{3,k'} + \mathbf{I}_{4,k'} + \mathbf{I}_{5,k'}$ has a simpler form than any of its three parts. To see this, the following formulas are needed [Ref. 2, (2.9) and (B7)]:

$$\langle\beta|\, \lambda f_0' \,|\sigma\rangle = \langle\beta|\, \lambda H' \,|\sigma\rangle[f(\beta) - f(\sigma)]/[\mathcal{E}(\beta) - \mathcal{E}(\sigma)]; \tag{2.47}$$

$$\lambda V_{E,\pm\epsilon}(\beta\gamma\sigma) = \langle\gamma|\, \lambda H' \,|\sigma\rangle \times \left[\frac{\delta_{\beta\gamma}}{\mathcal{E}(\sigma) - E \mp \frac{1}{2}i\hbar\epsilon} + \frac{\delta_{\beta\sigma}}{\mathcal{E}(\gamma) - E \pm \frac{1}{2}i\hbar\epsilon}\right]. \tag{2.48}$$

Using these and adding together the equations defininig $\mathbf{I}_{3,k}$, $\mathbf{I}_{4,k}$, and $\mathbf{I}_{5,k}$, one obtains

$$\begin{aligned}\mathbf{I}_{3,k} + \mathbf{I}_{4,k} + \mathbf{I}_{5,k} &= (V/kT^2) \int_{-\infty}^{\infty} dE \sum_{\beta\gamma\sigma} \delta(\mathcal{E}(\beta) - E)N_k(\beta) \\ &\times \langle\gamma|\, \lambda H' \,|\sigma\rangle\langle\sigma|\, \mathbf{S}^{\circ}{}_{nd} \,|\gamma\rangle \Bigg\{\frac{f(\beta) - f(\sigma)}{\mathcal{E}(\beta) - \mathcal{E}(\sigma)}\, \delta_{\beta\gamma} \\ &- \left[\frac{\delta_{\beta\gamma}}{\mathcal{E}(\sigma) - E + \frac{1}{2}i\hbar\epsilon} + \frac{\delta_{\beta\sigma}}{\mathcal{E}(\gamma) - E - \frac{1}{2}i\hbar\epsilon}\right] \\ &- \left[\frac{\delta_{\beta\gamma}}{\mathcal{E}(\sigma) - E - \frac{1}{2}i\hbar\epsilon} + \frac{\delta_{\beta\sigma}}{\mathcal{E}(\gamma) - E + \frac{1}{2}i\hbar\epsilon}\right]\Bigg\}.\end{aligned} \tag{2.49}$$

To further reduce this the following are needed: (1)

$$\frac{1}{X \pm \frac{1}{2}i\hbar\epsilon} = \frac{\mathcal{P}}{X} \mp \pi i\delta(X), \tag{2.50}$$

where $\mathcal{P}$ indicates that the principal part is to be taken; (2) the Hermitian character of $\mathbf{S}^{\circ}{}_{nd}$ and $\lambda H'$; (3) $f(\beta)\delta(\mathcal{E}(\sigma) - \mathcal{E}(\beta)) = f(\sigma)\delta(\mathcal{E}(\sigma) - \mathcal{E}(\beta))$. By integrating over E', one can now show that

$$\begin{aligned}\mathbf{I}_{3,k} + \mathbf{I}_{4,k} + \mathbf{I}_{5,k} &= \frac{4V}{kT^2} \sum_{\gamma\sigma} \mathrm{Re}\,(\langle\gamma|\, \lambda H' \,|\sigma\rangle\langle\sigma|\, \mathbf{S}^{\circ}{}_{nd} \,|\gamma\rangle) \\ &\times \frac{\mathcal{P}}{\mathcal{E}(\gamma) - \mathcal{E}(\sigma)} f(\gamma)N_k(\gamma).\end{aligned} \tag{2.51}$$

Transport Equation for $\phi_{6,k}$

The equation for determining ${}^1X_{E,\epsilon}$ [Ref. 2, (4.11)] is equivalent to (2.19) and[8]

$$\begin{aligned}\epsilon\delta_{\epsilon}(\mathcal{E}(\beta) - E)R_{E,\epsilon}(\beta\alpha) - \delta_{\epsilon}(\mathcal{E}(\beta) - E)\delta_{\beta\alpha} &= \frac{2\pi}{\hbar} \sum_{\mu} [|\langle\beta|\, \lambda H' \,|\mu\rangle|^2 - \lambda^3 F_{E,\epsilon}(\beta\mu)] \\ &\times \delta_{\epsilon}(\mathcal{E}(\beta) - E)\delta_{\epsilon}(\mathcal{E}(\mu) - E)R_{E,\epsilon}(\mu\alpha) \\ &- \frac{2\pi}{\hbar} \sum_{\mu} [|\langle\mu|\, \lambda H' \,|\beta\rangle|^2 - \lambda^3 F_{E,\epsilon}(\mu\beta)] \\ &\times \delta_{\epsilon}(\mathcal{E}(\mu) - E)\delta_{\epsilon}(\mathcal{E}(\beta) - E)R_{E,\epsilon}(\beta\alpha),\end{aligned} \tag{2.52}$$

where $\lambda^3 F_{E,\epsilon}(\beta\mu)$ is given by (A2).

To obtain the transport equation for $\phi_{6,k}$, multiply (2.52) by $(-V/kT^2)f(\alpha)\mathbf{S}(\alpha)N_k(\beta)$, sum over α and β, and exchange the summation indices μ and β in the terms arising from the part of (2.52) containing $R_{E,\epsilon}(\beta\alpha)$. Then, by introducing $E' = \mathcal{E}(\mu) - E$, integrating over E', and taking the limits $V \to \infty$ and $\epsilon \to 0$, one obtains

$$\mathbf{I}_{6,k} = \mp(2\pi/\hbar)\int_{-\infty}^{\infty} dE'\ \delta_\epsilon(E')$$
$$\times \sum_\mu \{\sum_\beta |\langle\beta|\ \lambda H'\ |\mu\rangle|^2\ \delta_\epsilon(\mathcal{E}(\beta) - \mathcal{E}(\mu) + E')$$
$$\times [N_k(\beta) - N_k(\mu)]\}$$
$$\times (V/kT^2)R_{\mathcal{E}(\gamma)-E',\epsilon}(\mu\alpha)f(\alpha)\mathbf{S}(\alpha) + \mathbf{F}_k, \qquad (2.53)$$

where

$$\mathbf{F}_k \equiv (V/kT^2)(2\pi/\hbar)\int_{-\infty}^{\infty} dE'\ \delta_\epsilon(E')$$
$$\times \sum_\mu \{\sum_\beta \lambda^3 F_{\mathcal{E}(\gamma)-E',\epsilon}(\beta\mu)\delta_\epsilon(\mathcal{E}(\beta) - \mathcal{E}(\mu) + E')$$
$$\times [N_k(\beta) - N_k(\mu)]\}$$
$$\times \sum_\alpha R_{\mathcal{E}(\gamma)-E',\epsilon}(\mu\alpha)f(\alpha)\mathbf{S}(\alpha) \qquad (2.54)$$

and

$$\mathbf{I}_{6,k} \equiv (V/kT^2)\sum_\alpha f(\alpha)\mathbf{S}(\alpha)N_k(\alpha)$$
$$= [d\langle N_k\rangle_0/dT]\mathbf{v}_k. \qquad (2.55)$$

Note that $\mathbf{I}_{6,k}$ is identical to the inhomogeneous term of (1.2) [to verify (2.55) use (2.8) and (2.35)–(2.37)].

The first part of the second member of (2.53) is the same as the second member of (2.29) except that instead of $Q_{E',\epsilon}(\mu\alpha\alpha)\langle\alpha|\ \lambda\mathbf{S}'\ |\alpha\rangle$ one has $R_{E',\epsilon}(\mu\alpha)\mathbf{S}(\alpha)$; the same difference distinguishes the definitions of $\phi_{2,k}$ and $\phi_{6,k}$. Consequently, the arguments from Ref. 4, which were used to obtain (2.31) from (2.29), apply here. By using those arguments, the first part of the second member of (2.53) can be rewritten as $V^{-1}\sum \tilde{\Omega}_{kl}\phi_{6,l}$.

Before $\mathbf{F}_k$ can be expressed in terms of the $\phi_{6,l}$, a specific perturbation must be specified. Then, the techniques already employed to derive the transport equations for ϕ_k, $\phi_{2,k}$, etc. can be used to express $\mathbf{F}_k$ as

$$\mathbf{F}_k = V^{-1}\sum_l \tilde{F}_{kl}\phi_{6,l}. \qquad (2.56)$$

The coefficients $\tilde{F}_{kl}$ for imperfection scattering are discussed in Sec. 3 and Appendix A.

It is now apparent that *the transport equation for* $\phi_{6,k}$ *is*

$$[d\langle N_k\rangle_0/dT]\mathbf{v}_k = V^{-1}\sum_l (\tilde{\Omega}_{kl} + \tilde{F}_{kl})\phi_{6,l}. \qquad (2.57)$$

Note that

$$\tilde{\Omega}_{kl} \propto \lambda^2 \quad \text{and} \quad \tilde{F}_{kl} \propto \lambda^3. \qquad (2.58)$$

From (2.57) and (2.58) it follows that $\phi_{6,k}$, and thus also K_6^{ij}, contains parts proportional to all powers of λ^n with $n \geq -2$, of which the parts with $n \geq 0$ are to be neglected. This suggests that $\phi_{6,k}$ be expanded as follows:

$$\phi_{6,k} = \phi_k^\circ + \phi_k' + \cdots, \qquad (2.59)$$

where $\phi_k^\circ \propto \lambda^{-2}$, $\phi_k' \propto \lambda^{-1}$, etc. The substitution of this into (2.57) and the equating of terms proportional to the same power of λ gives

$$[d\langle N_k\rangle_0/dT]\mathbf{v}_k = V^{-1}\sum_l \tilde{\Omega}_{kl}\phi_l^\circ, \qquad (2.60)$$

$$-V\sum_l \tilde{F}_{kl}\phi_l^\circ = V^{-1}\sum_l \tilde{\Omega}_{kl}\phi_l'. \qquad (2.61)$$

As expected, (2.60) is identical to (1.2), the transport equation for the λ^{-2} part of the conductivity. *Equation* (2.61) *is the transport equation for* ϕ_l'. The use of (2.61) to solve for ϕ_l' requires that the solution of (2.60) be known. This is in general not known, so that for practical purposes it may be easier to get information about ϕ_l' by using (2.57) and (2.59) rather than (2.61).

3. DISCUSSION

The equations which determine the λ^{-1} contribution to the conductivity in the limits $V \to \infty$ and $\epsilon \to 0$ can be summarized as follows: Equation (2.11) as modified by (2.27) and (2.59) becomes

$$\mathcal{K}^{ij} = [\tfrac{1}{2}V^{-1}\sum_k (\psi_k^i v_k^j + \phi_k^i V_k^j)\hbar\omega_k] + \text{transpose}, \qquad (3.1)$$

where

$$\psi_k = \sum_{a=2}^{5}\phi_{a,k} + \phi_k',$$

where ϕ_k is the solution of (1.2) and where the transport equations (2.31), (2.46), and (2.61) are replaced by a single *transport equation for* ψ_k:

$$[d\langle N_k\rangle_0/dT]\mathbf{V}_k + (\mathbf{I}_{3,k} + \mathbf{I}_{4,k} + \mathbf{I}_{5,k})$$
$$- V^{-1}\sum_l \tilde{F}_{kl}\phi_l = V^{-1}\sum_l \tilde{\Omega}_{kl}\psi_l; \qquad (3.2)$$

$\mathbf{V}_k$ is defined by (2.28). The inhomogeneous terms in (3.2) and the coefficients $\tilde{\Omega}_{kl}$ are real; consequently, the taking of the real part as prescribed in (2.11) is not necessary. Note that the homogeneous parts of (3.2) and of the lowest-order transport equations (1.2) are identical. However, the inhomogeneous part of (3.2) is proportional to λ, while that of (1.2) is independent of λ.

Considering its origin in the correlation function formula, it is very difficult to see how the term $\mathbf{I}_{3,k} + \mathbf{I}_{4,k} + \mathbf{I}_{5,k}$ could have been predicted by any extension of the arguments of kinetic theory. The terms involving $\mathbf{V}_k$ in (3.1) and (3.2), however, could perhaps have been anticipated had the form

of $\lambda\mathbf{S}'$ been known, while the term $-V^{-1}\sum \bar{F}_{kl}\phi_l$ is obtainable from kinetic theory in quite a natural way. $-V^{-1}\sum \bar{F}_{kl}\phi_l$ comes from the analysis of K_6^{ij}, which differs from the lowest-order contribution to the conductivity only in the replacement of ${}^0X_{E,\epsilon}$ in the lowest-order contribution by ${}^1X_{E,\epsilon}$ in K_6^{ij}. The difference between ${}^0X_{E,\epsilon}$ and ${}^1X_{E,\epsilon}$ is in the replacement of the $|\langle\beta|\ \lambda H'\ |\mu\rangle|^2$ in (2.21) by $[|\langle\beta|\ \lambda H'\ |\mu\rangle|^2 - \lambda^3 F_{E,\epsilon}(\beta\mu)]$ in (2.52) [also see (2.17) and (2.19)]. The additional term $-\lambda^3 F_{E,\epsilon}(\beta\mu)$, which is defined by (A2), has the same form as the correction to $|\langle\beta|\ \lambda H'\ |\mu\rangle|^2$ predicted by taking the transition rate from ordinary perturbation theory[11] to one order higher in λ, provided that E is set equal to $\mathcal{E}(\mu)$ as it eventually is. Consequently, just as the equations for the lowest-order contribution to the conductivity determined in Ref. 4 are the same as the equations predicted by kinetic theory, the contribution to the conductivity labeled K_6^{ij} is the same as the contribution determined by taking the transition rate in kinetic theory to one order higher in λ.

With lattice imperfections alone ($\lambda H' = \lambda T' + \lambda V_2$), *the only contribution to the conductivity in the* λ^{-1} *order arises from the scattering mechanism,* i.e., from $-V^{-1}\sum \bar{F}_{kl}\phi_l$. For this case $\mathbf{V}_k$ and $\mathbf{I}_{3,k} + \mathbf{I}_{4,k} + \mathbf{I}_{5,k}$ are zero.

$\mathbf{V}_k$ arose from the diagonal part of $\lambda\mathbf{S}'$. Now, it has been shown (Ref. 1, p. 174) that the separation of H into H° and $\lambda T' + \lambda V_2$, which fulfills the requirement [Ref. 2, (2.3)] that $\langle\alpha|\ \lambda T' + \lambda V_2\ |\alpha\rangle$ be zero, causes the contribution of $\lambda T' + \lambda V_2$ to the diagonal part of $\lambda\mathbf{S}'$ to also be zero. Thus, $\mathbf{V}_k$ vanishes.

$\mathbf{I}_{3,k} + \mathbf{I}_{4,k} + \mathbf{I}_{5,k}$ involves a sum over the product $\langle\gamma|\ \lambda H'\ |\sigma\rangle\langle\sigma|\ \mathbf{S}^\circ{}_{nd}\ |\gamma\rangle$ [see (2.49)]. The perturbation $\lambda H' = \lambda T' + \lambda V_2$ is a linear combination of products of the type $a_{\mathbf{k}s}a_{\mathbf{k}'s'}$, $a_{\mathbf{k}s}a^\dagger_{-\mathbf{k}'s'}$, $a^\dagger_{-\mathbf{k}s}a_{\mathbf{k}'s'}$, and $a^\dagger_{-\mathbf{k}s}a^\dagger_{-\mathbf{k}'s'}$, where the coefficients vanish for the terms with $\mathbf{k} = -\mathbf{k}'$ [see (3.5) and Ref. 1, p. 174]. Obviously, only that part of $\mathbf{S}^\circ{}_{nd}$ which contains products of two of the $a^\dagger_{\mathbf{k}s}$ and $a_{\mathbf{k}s}$ can possibly give nonzero elements $\langle\sigma|\ \mathbf{S}^\circ{}_{nd}\ |\gamma\rangle$ for states σ and γ for which $\langle\gamma|\ \lambda T' + \lambda V_2\ |\sigma\rangle$ is also nonzero. The necessary part of $\mathbf{S}^\circ{}_{nd}$ is [Ref. 1, (3.31)]

$$\mathbf{S}^\circ{}_{2,nd} = -\frac{1}{2V}\sum_{\mathbf{k}ss',s\neq s'}(a_{\mathbf{k}s} + a^\dagger_{-\mathbf{k}s})(a_{-\mathbf{k}s'} - a^\dagger_{\mathbf{k}s'}) \times \hbar\omega_{\mathbf{k}s}\mathbf{U}_{\mathbf{k}ss'}. \quad (3.3)$$

It can now be seen that a σ and γ for which $\langle\gamma|\ S^\circ{}_{2,nd}\ |\sigma\rangle$ is nonzero picks out a combination of creation and annihilation operators from $\lambda T' + \lambda V_2$ which has a vanishing coefficient. Hence, there are no two states γ and σ for which both $\langle\gamma|\ \lambda T' + \lambda V_2\ |\sigma\rangle$ and $\langle\sigma|\ \mathbf{S}^\circ{}_{2,nd}\ |\gamma\rangle$ are nonzero, so that *for lattice imperfections*

$$\mathbf{I}_{3,k} + \mathbf{I}_{4,k} + \mathbf{I}_{5,k} = 0. \quad (3.4)$$

With anhormonic forces alone ($\lambda H' = \lambda V_3$), *the inhomogeneous term* $V^{-1}\sum \bar{F}_{kl}\phi_l$ *is zero, but the other terms, which arise from corrections to the density matrix and energy flux, are in general nonzero.* The appropriate formulas in Sec. 2 and Ref. 1 are used in Appendix B to derive Eqs. (B10) and (B22), which with (2.28) give $\mathbf{I}_{3,k} + \mathbf{I}_{4,k} + \mathbf{I}_{5,k}$ and $\mathbf{V}_k$ as functions of the anharmonic force constants, the frequencies, the polarization vectors, etc. It is also shown there that these first two inhomogeneous terms have the proper dependence on volume to neither diverge nor vanish in the limit $V \to \infty$ and that they are odd functions of their subscripts [see (B11) and (B18)]. Because of this latter property, the inhomogeneous terms are orthogonal to the known solution $\hat{\phi}^H_l$ of the transposed homogeneous equation $V^{-1}\sum_k \bar{\Gamma}_{kl}\hat{\phi}^H_k = 0$, as is required for a solution to (3.2) to exist.[12]

We now demonstrate that $-V^{-1}\sum \bar{F}_{kl}\phi_l$ vanishes when $\lambda H' = \lambda V_3$ [(B3) defines λV_3]. This term in (3.2) arose from the simplification of $\mathbf{F}_k$ [see (2.54)]. From (2.54) and (A2) it is readily seen that $\mathbf{F}_k$ is zero if

$$\langle\mu|\ \lambda V_3\ |\beta\rangle\langle\beta|\ \lambda V_3\ |\nu\rangle\langle\nu|\ \lambda V_3\ |\mu\rangle \times \delta_\epsilon(\mathcal{E}(\beta) - \mathcal{E}(\mu) + E')\delta_\epsilon(E')$$

is zero. Since the δ_ϵ functions become Dirac δ's, only states for which $\mathcal{E}(\mu) = \mathcal{E}(\beta)$ contribute. States μ and β for which $\mathcal{E}(\mu) = \mathcal{E}(\beta)$ and for which $\langle\mu|\ \lambda V_3\ |\beta\rangle$ is nonzero differ by an increase (or decrease) in the occupation numbers of two modes and a decrease (or increase) in the occupation number of one other mode. For such states μ and β there are no states ν such that $\langle\beta|\ \lambda V_3\ |\nu\rangle\langle\nu|\ \lambda V_3\ |\mu\rangle$ is also nonzero. This is because $\langle\beta|\ \lambda V_3\ |\nu\rangle$ and $\langle\nu|\ \lambda V_3\ |\mu\rangle$ are only nonzero for states β, ν and states ν, μ which differ: by an increase (or decrease) in the occupation numbers of three modes; by an increase (or decrease) of the occupation numbers for two modes and the decrease (or increase) for one mode. It is readily verified

[11] P. A. M. Dirac, *Principles of Quantum Mechanics* (Oxford University Press, London, 1958), pp. 178–181. The transition rate to lowest order in λ is

$$(2\pi/\hbar)\,|<\beta|\ \lambda\ H'|\ \mu\ >|^2\ \delta(\mathcal{E}(\beta) - \mathcal{E}(\mu)).$$

[12] For a statement of the theorem involved, see R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience Publishers, Inc., New York, 1953), p. 6. It follows from the form of $\bar{\Gamma}_{kl}$ given by Ref. 4, (29), that $\hat{\phi}^H \propto \omega_k$ is a solution of the transposed homogeneous equation.

that there are no states ν such that both β, ν and ν, μ are related in one of these four ways and such that $\langle\mu|\ \lambda V_3\ |\beta\rangle$ is also nonzero. Consequently, $\mathbf{F}_k$ is zero.

It is consistent in discussing anharmonic scattering to order λ^{-1} to ignore the quartic term in the potential energy expansion provided that one makes the usual assumption that the quartic term is of one order higher in the perturbation than the cubic term (i.e., that $H = H^0 + \lambda T' + \lambda V_2 + \lambda V_3 + \lambda^2 V_4 + \cdots$). In this case, the first contribution to the conductivity of the quartic term is of order λ^0. In other words, if one is interested in corrections to the lowest-order contribution to conductivity, one should consider the corrections to the energy flux and the density matrix due to λV_3 (i.e., $\mathbf{I}_{3,k} + \mathbf{I}_{4,k} + \mathbf{I}_{5,k}$ and $\mathbf{V}_k$) before considering the various corrections due to $\lambda^2 V_4$.

Finally, note that *with* $\lambda H' = \lambda T' + \lambda V_2 + \lambda V_3$ *there is a mixing of the effects of the anharmonic and the imperfection scattering in* $-V^{-1}\sum \tilde{F}_{kl}\phi_l$. In particular, if the two types of perturbation are characterized by different parameters, e.g., $\lambda_2 T' + \lambda_2 V_2$ for imperfections and $\lambda_3 V_3$ for anharmonicities, there will be mixed terms in $\tilde{F}_{kl}$ proportional to $\lambda_2(\lambda_3)^2$.

$\tilde{F}_{kl}$ for Point Imperfections

The perturbations to the kinetic energy and to the quadratic part of the potential energy due to imperfections are, respectively,

$$\lambda T' = \tfrac{1}{2}\sum_{jk} c^T_{jk}(a_j + a^\dagger_{-j})(a_k + a^\dagger_{-k}); \tag{3.5a}$$

$$\lambda V_2 = \tfrac{1}{2}\sum_{jk} c^V_{jk}(a_j - a^\dagger_{-j})(a_k - a^\dagger_{-k}). \tag{3.5b}$$

The coefficients in both of these satisfy the relations $c_{jk}{}^* = c_{-j-k}$; $c_{jk} = c_{kj}$; $c_{\mathbf{k}s,-\mathbf{k}s'} = 0$ (see Ref. 1, p. 174).

It is shown in Appendix A that, for $\lambda H' = \lambda T' + \lambda V_2$,

$$\tilde{F}_{kl} = \tilde{L}_{kl} + \tilde{M}_{kl}; \tag{3.6}$$

$$\tilde{L}_{k'l} = \frac{4\pi}{\hbar^3}\sum_{jk}\left[\mathrm{Re}\,(c^-_{k'-j}c^-_{j-k}c^-_{k-k'})\frac{\mathcal{P}}{\omega_k - \omega_{k'}} + \mathrm{Re}\,(c^-_{k'-j}c^+_{j-k}c^+_{k-k'})\frac{\mathcal{P}}{\omega_k + \omega_{k'}}\right] \times \delta(\omega_{k'} - \omega_j)[V\delta_{jl} - V\delta_{k'l}]; \tag{3.7}$$

$$\tilde{M}_{k'l} = -\frac{4\pi^2}{\hbar^3}\sum_{jk}\mathrm{Im}\,(c^-_{k'-j}c^-_{j-k}c^-_{k-k'}) \times \delta(\omega_{k'} - \omega_j)\,\delta(\omega_{k'} - \omega_k)[4\langle N_{k'}\rangle_0(\langle N_{k'}\rangle_0 + 1) \times V(\delta_{jl} + \delta_{k'l} + \delta_{kl}) + V(\delta_{jl} + \delta_{k'l})]; \tag{3.8}$$

where

$$c^\pm_{j-k} \equiv c^T_{j-k} \pm c^V_{j-k} \tag{3.9}$$

(c^-_{jk} is written as simply c_{jk} in Ref. 4).

As an example of the constants $c^-_{k'-j}c^\pm_{j-k}c^\pm_{k-k'}$, consider the case of isotopic scattering ($\lambda H' = \lambda T'$); in particular, consider a lattice containing randomly distributed particles of two masses, m_1 and m_2. It is shown in Appendix A that in this case

$$c^-_{k-j}c^\pm_{j-l}c^\pm_{l-k} = N^{-2}(1 - 3f + 2f^2)f\, f_{k-j}f_{j-l}f_{l-k}, \tag{3.10}$$

where

$$f_{jk} = \tfrac{1}{2}\hbar(\omega_j\omega_k)^{\frac{1}{2}}(\mathbf{e}_j\cdot\mathbf{e}_k)m\left(\frac{1}{m_1} - \frac{1}{m_2}\right). \tag{3.11}$$

N is the number of unit cells in the system ($N \propto V$), and f is the fraction of particles which are of mass m_2. For comparison, we give the expression for the product of coefficients needed in $\tilde{\Lambda}_{kl}$ [see Ref. 4, (17)]:

$$|c^-_{j-k}|^2 = N^{-1}(1 - f)f|f_{j-k}|^2. \tag{3.12}$$

The concentration f may take on all values between zero and one. This is particularly useful when studying the thermal resistance due to a mixture of two isotopes or of two chemically similar atoms. In such systems the perturbation is small because the mass difference is small, not because the concentration is small. Note that the coefficients (3.10) and (3.12) vanish for both $f=0$ and $f=1$, which is a consequence of the requirement that $\langle\alpha|\ \lambda H'\ |\alpha\rangle = 0$ [Ref. 2, (2.3)]. To fulfill this requirement the mass m must be changed from m_1 to m_2 as f goes from $f = 0$ to $f = 1$ [see (A12)].

Now consider whether or not the existence of solutions of the homogeneous equation associated with (1.2) and (3.2) causes any arbitrariness in $\mathcal{K}^{ij}$. For imperfection scattering, (3.2) becomes

$$-V^{-1}\sum_l(\tilde{L}_{kl} + \tilde{M}_{kl})\phi_l = V^{-1}\sum_l\tilde{\Lambda}_{kl}\psi_l. \tag{3.13}$$

For a solution to exist, the first member of this must be orthogonal to any solution φ^H_l of the homogeneous equation

$$V^{-1}\sum_l\tilde{\Lambda}_{kl}\varphi^H_l = 0. \tag{3.14}$$

(Since $\tilde{\Lambda}_{kl} = \tilde{\Lambda}_{lk}$, the homogeneous and the transposed homogeneous equations are identical.) Equation (3.14) has the solution $\varphi^H_l = \mathbf{f}(\omega_l)$, where $\mathbf{f}(\omega_l)$ is an arbitrary vector function of ω_l [to verify, use Ref. 4, (17)]. Since $\omega_l = \omega_{-l}$, φ^H_l is an even function of its subscript, i.e., $\varphi^H_l = \varphi^H_{-l}$. Since (3.1) is only sensitive to the odd part of ψ_l, i.e., to $\frac{1}{2}(\psi_l - \psi_{-l})$, the solution $\mathbf{f}(\omega_l)$ does not affect $\mathcal{K}^{ij}$ *provided that* a solution to (3.13) exists.

A study of the inhomogeneous part of (3.13) requires a knowledge of the properties of ϕ_l, the solution of (1.2). It follows from Ref. 4, (17) and $c_{jk} = c^*_{j-k}$ that $\tilde{\Lambda}_{kl} = \tilde{\Lambda}_{-k-l}$. From this, from $\langle N_k\rangle_0 = \langle N_{-k}\rangle_0$, and from $\mathbf{v}_k = -\mathbf{v}_{-k}$ it follows that $\phi_l^{\text{odd}} = \frac{1}{2}(\phi_l - \phi_{-l})$—the odd part of any solution of (1.2)—is also a solution. Then, noting that (3.14) is also the homogeneous part of (1.2), one can represent any solution of (1.2) as

$$\phi_l = \phi_l^{\text{odd}} + \varphi_l^H, \tag{3.15}$$

and, assuming that $\varphi_l^H = \mathbf{f}(\omega_l)$ is the only solution of (3.14), ϕ_l^{odd} is unique. Now, since $c_{jk} = c^*_{-j-k}$, one has

$$\text{Re}\,(c_{k-j}c_{j-l}c_{l-k}) = \text{Re}\,(c_{-kj}c_{-jl}c_{-lk}) \tag{3.16a}$$

and

$$\text{Im}\,(c_{k-j}c_{j-l}c_{l-k}) = -\text{Im}\,(c_{-kj}c_{-jl}c_{-lk}). \tag{3.16b}$$

Using these, (3.7) and (3.8), one can show that

$$\tilde{M}_{kl} = -\tilde{M}_{-k-l}; \qquad \tilde{L}_{kl} = \tilde{L}_{-k-l};$$
$$V^{-1}\sum_l \tilde{L}_{kl}\varphi_l^H = 0. \tag{3.17}$$

It is now easily demonstrated that

$$-V^{-1}\sum_l \tilde{L}_{-kl}\phi_l = V^{-1}\sum_l \tilde{L}_{kl}\phi_l \tag{3.18}$$

and that

$$V^{-1}\sum_l \tilde{M}_{-kl}(\phi_l^{\text{odd}} + \varphi_l^H) = V^{-1}\sum_l \tilde{M}_{kl}(\phi_l^{\text{odd}} - \varphi_l^H). \tag{3.19}$$

Since $-V^{-1}\sum \tilde{L}_{kl}\phi_l$ and $-V^{-1}\sum \tilde{M}_{kl}\varphi_l^H$ are odd functions of k, they are orthogonal to $\varphi_k^H = \mathbf{f}(\omega_k)$ and, when considered alone, yield a soluble equation (3.13). Since $-V^{-1}\sum \tilde{L}_{kl}\phi_l$ depends only on the uniquely determined ϕ_l^{odd}, it leads to a correction to the conductivity which is completely determined by the perturbation, as it should be. However, $-V^{-1}\sum \tilde{M}_{kl}\varphi_l^H$ gives a correction which depends on the arbitrary choice of $\mathbf{f}(\omega_l)$, and $-V^{-1}\sum \tilde{M}_{kl}\phi_l^{\text{odd}}$ leads to an insoluble equation for ψ_l, since it is not in general orthogonal to φ_l^H. Thus, we are only capable of treating systems for which $\tilde{M}_{kl} = 0$, so that we are restricted to perturbations for which

$$\tilde{F}_{kl} = \tilde{L}_{kl}. \tag{3.20}$$

What is the significance of restriction (3.20)? It is apparent from (3.8) that restriction (3.20) is satisfied if the $c_{k-j}c_{j-l}c_{l-k}$ are all real. It follows from (3.10) and (3.11) that, for the scattering due to randomly distributed particles of two different masses, the $c_{k-j}c_{j-l}c_{l-k}$ are indeed real. Actually, this is valid for particles with any number of different masses. Thus, (3.20) is satisfied for isotopic scattering. However, for randomly distributed point imperfections *with an associated strain field*, a more surprising result is obtained: (3.20) is then satisfied *only* when the static strain field associated with each imperfection possesses inversion symmetry about the lattice site at the center of the imperfection. Such a symmetry is usually assumed, but is not a necessary property. [The calculation of the coefficients $c_{k-j}c_{j-l}c_{l-k}$ for static strain fields is similar to the calculation of (3.10) and (3.11), and is not given here.]

A Relaxation Time Solution; Isotopic Scattering

An exact, explicit expression for the λ^{-1}-order contribution to the conductivity exists for isotopic scattering ($\lambda H' = \lambda T'$).[13] It follows from (3.11) that $f_{jk} = -f_{-jk} = -f_{j-k}$; from this and (3.10), and (3.12) it follows that $c_{k-j}c_{j-l}c_{l-k} = c_{-k-j}c_{j-l}c_{lk}$ and that $|c^-_{j-k}|^2 = |c^-_{jk}|^2$. Then, using (3.7), and Ref. 4, (17) we find that $\tilde{\Lambda}_{kl} = \tilde{\Lambda}_{-kl}$ and $\tilde{L}_{kl} = \tilde{L}_{-kl}$ for $k \neq \pm l$. Note also that the form of $\tilde{L}_{kl}$ is such that

$$V^{-1}\sum_l \tilde{L}_{k'l}\phi_l = V^{-1}\sum_{l(\neq k')} \tilde{L}_{k'l}(\phi_l - \phi_{k'}), \tag{3.21}$$

where the omission of the term $l = k'$ in the sum signifies that the part of $\tilde{L}_{kl}$ containing the δ function $\delta_{k'l}$ is to be omitted; it follows from Ref. 4, (17) that a similar result holds for $\tilde{\Lambda}_{kl}$. By using these properties of $\tilde{L}_{kl}$ and $\tilde{\Lambda}_{kl}$, one can rewrite (1.2) and (3.2) in the form of explicit expressions for ϕ_l and ψ_l. The substitution of these expressions into (1.1) and (3.1) gives

$$K^{ij} + \mathcal{K}^{ij} = V^{-1}\sum_k \tau_k[1 - (\tau_k/\tau_k')] \times [d\langle N_k\rangle_0/dT]\hbar\omega_k v_k^i v_k^j, \tag{3.22}$$

where

$$(\tau_k)^{-1} \equiv -V^{-1}\sum_{l(\neq k)} \tilde{\Lambda}_{kl} = (2\pi/\hbar^2)\sum_j |c^-_{j-k}|^2\delta(\omega_j - \omega_k) \propto \lambda^2 \tag{3.23}$$

[Ref. 4, (17) has been used here] and

$$(\tau_k')^{-1} \equiv -V^{-1}\sum_{l(\neq k)} \tilde{L}_{kl} \propto \lambda^3. \tag{3.24}$$

The addition of the correction $-\tau_k/\tau_k'$ does not eliminate the familiar divergence in the lowest-order

[13] The fact that an explicit solution of this type can be exact for lowest order was pointed out by Schieve and Peterson (Ref. 3).

contribution to the conductivity for isotopic scattering.[14]

Temperature Dependence of $\mathcal{K}^{ij}$

$\tilde{L}_{kl}$ and $\tilde{\Lambda}_{kl}$ contain no constants which depend on the temperature; thus, for imperfection scattering the temperature variation of $\mathcal{K}^{ij}$ is entirely due to the variation of ϕ_l [see (3.1), (3.13), and (3.20)]. Since ϕ_l also determines the temperature dependence of the lowest-order contribution, no significant alteration of the conductivity–temperature curve is likely to result from the adding in of $\mathcal{K}^{ij}$. [Note that for isotopic scattering the entire temperature dependence comes from the factor $d\langle N_k\rangle_0/dT$ in (3.22).]

At high temperatures the temperature dependence of the conductivity is more easily discussed because of the classical behavior of the system; in particular, $kT \gg \hbar\,\omega_k$, $\langle N_k\rangle_0 \gg 1$, and $\langle N_k\rangle_0 \propto T$. For anharmonic forces [see (2.28) and (B10)] the inhomogeneous terms $[d\langle N_k\rangle_0/dT]\mathbf{V}_k$ and $\mathbf{I}_{3,k} + \mathbf{I}_{4,k} + \mathbf{I}_{5,k}$ are proportional to T. From Ref. 4, (29) it follows that $\tilde{\Gamma}_{kl} \propto T$. Since $\tilde{F}_{kl} = 0$ for $\lambda H' = \lambda V_3$, we find from (3.1), (3.2), and the above that $\mathcal{K}^{ij}$ becomes *temperature independent* at high temperatures with anharmonic forces alone. This is to be compared with a T^{-1} dependence for the lowest-order contribution to the conductivity.

Finally, note that the correction $\mathcal{K}^{ij}$ may be either positive or negative depending on the character of the imperfections and anharmonic forces.

ACKNOWLEDGMENT

One of us (R.J.H.) wishes to thank Professor G. H. Wannier for his kind hospitality at the University of Oregon.

APPENDIX A. CALCULATIONS FOR $\lambda H' = \lambda T' + \lambda V_2$

$\mathbf{F}_k$ will now be expressed in the form $V^{-1}\sum \tilde{F}_{kl}\phi_{6,l}$, where $\mathbf{F}_k$ is defined by (2.54). We begin by considering the quantity

$$\sum_\beta \lambda^3 F_{E,\epsilon}(\beta\mu)\,\delta(\mathcal{E}(\beta) - E)[N_k(\beta) - N_k(\mu)] \qquad \text{(A1)}$$

evaluated at $E = \mathcal{E}(\mu)$. $\lambda^3 F_{E,\epsilon}$ is defined by Ref. 2, (B6):

$$\lambda^3 F_{E,\epsilon}(\beta\mu) \equiv \Big\{\sum_\nu \langle\mu|\,\lambda H'\,|\beta\rangle\langle\beta|\,\lambda H'\,|\nu\rangle\langle\nu|\,\lambda H'\,|\mu\rangle \times \Big(\frac{\mathcal{P}}{\mathcal{E}(\nu) - E} + \pi i\,\delta(\mathcal{E}(\nu) - E)\Big)\Big\} + \text{c.c.}, \qquad \text{(A2)}$$

where (2.50) has been used, and where +c.c. indicates that the complex conjugate is to be added. Because of the $\delta_\epsilon(E')$ in (2.54), where $E' = \mathcal{E}(\mu) - E$, we are interested only in (A1) with $E = \mathcal{E}(\mu)$. For $\lambda H' = \lambda T' + \lambda V_2$ one can show by using (3.5) and (3.9) that

$$\langle\mu|\,\lambda H'\,|\beta\rangle\,\delta(\mathcal{E}(\beta) - \mathcal{E}(\mu)) = \frac{1}{\hbar}\sum_{ij}(c_{i-j}^-)^*[N_i(\mu)(N_j(\mu)+1)]^{\frac{1}{2}} \times \delta(\omega_i - \omega_j)\prod_l \delta_{N_l(\beta),N_l(\mu)-\delta_{il}+\delta_{jl}}. \qquad \text{(A3)}$$

Furthermore, for β related to μ by $N_l(\beta) = N_l(\mu) - \delta_{il} + \delta_{jl}$, one has

$$\begin{aligned}\langle\beta|\,\lambda H'\,|\nu\rangle\langle\nu|\,\lambda H'\,|\mu\rangle &= \sum_{kl}\sum_{mn}\{c_{k-l}^- c_{m-n}^-\langle\beta|\,a_k a_l^\dagger\,|\nu\rangle\langle\nu|\,a_m a_n^\dagger\,|\mu\rangle \\ &+ \tfrac{1}{2}c_{kl}^+ c_{mn}^+(\langle\beta|\,a_k a_l\,|\nu\rangle\langle\nu|\,a_{-m}^\dagger a_{-n}^\dagger\,|\mu\rangle \\ &+ \langle\beta|\,a_{-k}^\dagger a_{-l}^\dagger\,|\nu\rangle\langle\nu|\,a_m a_n\,|\mu\rangle)\}. \qquad \text{(A4)}\end{aligned}$$

Note, for example, that, for any given m, n, and $|\mu\rangle$, there is only one member of the complete set $\langle\nu|$ for which $\langle\nu|\,a_m a_n\,|\mu\rangle$ is nonzero. Similarly, there is only one member of the set $|\beta\rangle$ for which the product of Kronecker δ functions in (A3) is nonzero. By combining (A3) and (A4) and using $\delta(\omega_i + \omega_k) = 0$ and $(c_{j-i}c_{i-l}c_{l-j})^* = c_{i-j}c_{j-l}c_{l-i}$, one can show that

$$\begin{aligned}\lambda^3 F_{\mathcal{E}(\beta),\epsilon}(\beta\mu)\,\delta(\mathcal{E}(\beta) - \mathcal{E}(\mu))[N_{k'}(\beta) - N_{k'}(\mu)] &= -\frac{2\pi}{\hbar^2}\sum_{jl}\operatorname{Im}(c_{k'-j}^- c_{j-l}^- c_{l-k'}^-)\,\delta(\omega_{k'} - \omega_j)\,\delta(\omega_{k'} - \omega_l) \\ &\times [4N_j(\mu)N_{k'}(\mu)N_l(\mu) \\ &+ 2N_j(\mu)N_{k'}(\mu) + 2N_{k'}(\mu)N_l(\mu) \\ &+ 2N_l(\mu)N_j(\mu) + N_{k'}(\mu) + N_j(\mu)] \\ &+ \frac{2}{\hbar^2}\sum_{jl}\Big[\operatorname{Re}(c_{k'-j}^- c_{j-l}^- c_{l-k'}^-)\frac{\mathcal{P}}{\omega_l - \omega_{k'}} \\ &+ \operatorname{Re}(c_{k'-j}^- c_{j-l}^+ c_{l-k'}^+)\frac{\mathcal{P}}{\omega_l + \omega_{k'}}\Big] \\ &\times \delta(\omega_{k'} - \omega_j)[N_j(\mu) - N_{k'}(\mu)]. \qquad \text{(A5)}\end{aligned}$$

When (A5) is multiplied by

$$(2\pi/\hbar)(V/kT^2)\sum_\alpha \int dE'\,R_{\mathcal{E}(\mu)-E'}(\mu\alpha)f(\alpha)\mathrm{S}(\alpha) \qquad \text{(A6)}$$

and is summed over μ, one obtains an expression for $\mathbf{F}_k$ which has the same form as the second member of (A5) except that: the factors $N_l(\mu)$, etc. are replaced by $\phi_{6,l}$, etc. [$\phi_{6,k'}$ is defined by (2.16)]; the

[14] See, e.g., P. Carruthers, Rev. Mod. Phys. 33, 120 (1961).

factors $N_k(\mu)N_l(\mu)$, etc., are replaced by $\phi_{6,kl}$, etc., where

$$\phi_{6,kl} \equiv \frac{V}{kT^2} \sum_{\mu} N_k(\mu)N_l(\mu)$$

$$\times \sum_{\alpha} \int_{-\infty}^{\infty} dE' \; \delta_\epsilon(E')R_{\varepsilon(\mu)-E',\epsilon}(\mu\alpha)f(\alpha)\mathrm{S}(\alpha); \qquad \text{(A7)}$$

the factors $N_k(\mu)N_{k'}(\mu)N_l(\mu)$ are replaced by $\phi_{6,kk'l}$, where the definition of $\phi_{6,kk'l}$ is the obvious generalization of (A7). The arguments made to derive (2.26) are equally valid here (the time dependence is simply taken to one order higher in λ); thus,

$$\phi_{6,kl} = \phi_{6,k}\langle N_l\rangle_0 + \phi_{6,l}\langle N_k\rangle_0. \qquad \text{(A8a)}$$

The appropriate generalization of (A8a) for $\phi_{6,jkl}$ follows from an application of the same arguments used to derive (A8a) and from the use of (A8a) itself; the result is

$$\phi_{6,jkl} = \phi_{6,j}\langle N_k\rangle_0\langle N_l\rangle_0 + \phi_{6,k}\langle N_l\rangle_0\langle N_j\rangle_0 + \phi_{6,l}\langle N_j\rangle_0\langle N_k\rangle_0. \qquad \text{(A8b)}$$

Multiplying (A5) by (A6), summing over μ, using (A8) and the properties of the Dirac δ functions, one obtains

$$\tilde{\mathbf{F}}_{k'} = V^{-1} \sum_{l} (\tilde{L}_{k'l} + \tilde{M}_{k'l})\phi_{6,l}, \qquad \text{(A9)}$$

where $\tilde{L}_{k'l}$ and $\tilde{M}_{k'l}$ are given by (3.7) and (3.8).

Randomly Distributed Point Imperfections

A comparison of (3.5a) with Ref. 1, (4.19) shows that

$$c_{jk}^{T} = \frac{\hbar}{2}(\omega_j\omega_k)^{\frac{1}{2}}(\mathbf{e}_j\cdot\mathbf{e}_k)N^{-1} \sum_{i} \frac{\delta m_i}{m + \delta m_i} e^{i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{x}_i}, \qquad \text{(A10)}$$

where m_i is the mass of the ith particle, $\delta m_i = m_i - m$, $m^{-1} = N^{-1}\sum_i (m_i)^{-1}$, and where N without a subscript is the number of unit cells in the lattice ($N \propto V$). For a system made up of particles of two masses, m_1 and m_2, which are randomly distributed and where f is the fraction of the particles that are of mass m_2, one has

$$\frac{\delta m_i}{m + \delta m_i} = [f(\mathbf{x}_i) - f]m\left(\frac{1}{m_1} - \frac{1}{m_2}\right), \qquad \text{(A11)}$$

where $f(\mathbf{x}_i)$ equals one when particle i (associated with lattice vector $\mathbf{x}_i$) is of mass m_2 and is zero otherwise. Obviously,

$$N^{-1}\sum_i f(\mathbf{x}_i) = f, \qquad \frac{1}{m} = \frac{f}{m_2} + \frac{(1-f)}{m_1}, \qquad \text{(A12)}$$

and

$$c_{jk}^{T} = N^{-1}\sum_i [f(\mathbf{x}_i) - f]e^{i(\mathbf{j}+\mathbf{k})\cdot\mathbf{x}_i} f_{jk}, \qquad \text{(A13)}$$

where f_{jk} is defined by (3.11).

For the evaluation of $|c_{j-k}^{-}|^2$ and $c_{j-k}^{-}c_{k-l}^{\pm}c_{l-j}^{\pm}$ we need

$$N^{-2}\sum_{mn}[f(\mathbf{x}_m) - f][f(\mathbf{x}_n) - f]e^{i(\mathbf{k}-\mathbf{k}')\cdot(\mathbf{x}_m-\mathbf{x}_n)} = N^{-1}(1-f)f \qquad \text{(A14a)}$$

and

$$N^{-3}\sum_{lmn}[f(\mathbf{x}_l) - f][f(\mathbf{x}_m) - f][f(\mathbf{x}_n) - f] \times e^{i[(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}_l + (\mathbf{k}'-\mathbf{k}'')\cdot\mathbf{x}_m + (\mathbf{k}''-\mathbf{k})\cdot\mathbf{x}_n]} = N^{-2}(1 - 3f + 2f^2)f, \qquad \text{(A14b)}$$

which are valid for random distributions of imperfections. To check (A14b), for example, make the substitutions $\mathbf{x}_m = \mathbf{x}_{m'} + \mathbf{x}_l$ and $\mathbf{x}_n = \mathbf{x}_{n'} + \mathbf{x}_l$ and sum over l. The quantity to be summed is $[f(\mathbf{x}_l) - f][f(\mathbf{x}_{m'} + \mathbf{x}_l) - f][f(\mathbf{x}_{n'} + \mathbf{x}_l) - f]$. Multiplying this out, one obtains eight terms. (A14b) is obtained by evaluating each of these terms using: (a) $\sum f(\mathbf{x}_l) = Nf$; (b) $\sum f(\mathbf{x}_l)f(\mathbf{x}_l + \mathbf{x}_k)$ equals fN when $\mathbf{x}_k = 0$ and equals f^2N otherwise; (c) the result analogous to the above for $\sum f(\mathbf{x}_l)f(\mathbf{x}_l + \mathbf{x}_k) \times f(\mathbf{x}_l + \mathbf{x}_j)$. Results (3.10) and (3.12) follow immediately from (A13) and (A14).

APPENDIX B: CALCULATIONS FOR $\lambda H' = \lambda V_3$

The detailed expressions for $\mathbf{V}_k$ and $\mathbf{I}_{3,k} + \mathbf{I}_{4,k} + \mathbf{I}_{5,k}$ are derived here, and it is shown that they: (1) are finite in the limit $V \to \infty$; (2) are real; (3) are odd functions of k.

In the limit $V \to \infty$, one has

$$V^{-1}\sum_k \to (2\pi)^{-3}\sum_s \int d\mathbf{k}, \qquad \text{(B1a)}$$

$$V\delta_{jk} \to (2\pi)^3\delta_{ss'}\delta(\mathbf{j} - \mathbf{k}), \qquad \text{(B1b)}$$

and

$$V\Delta_{\mathbf{j}+\mathbf{k}+\mathbf{l}} \to (2\pi)^3\delta(\mathbf{j} + \mathbf{k} + \mathbf{l} - \mathbf{K}), \qquad \text{(B1c)}$$

where $k = (\mathbf{k}, s)$; $\delta(\mathbf{k})$ is a Dirac δ function; and $\mathbf{K}$ is any reciprocal lattice vector including zero. $\Delta_{\mathbf{j}+\mathbf{k}+\mathbf{l}}$ was introduced in Ref. 1 and equals one when $\mathbf{j} + \mathbf{k} + \mathbf{l} = \mathbf{K}$ and equals zero otherwise. For quantities to converge to a nonzero value in the limit $V \to \infty$ all factors of V (or N, which is proportional to V), summations in $\mathbf{k}$ space, Kronecker δ functions, and Δ functions must occur in the combinations indicated on the left of (B1).

The cubic term in the potential energy expansion is [Ref. 1, (4.1)]

$$\lambda V_3 = \frac{1}{3!}\sum_{mnr}\sum_{abc} B^{abc}(\mathbf{x}_m, \mathbf{x}_n, \mathbf{x}_r)Q^a(\mathbf{x}_m)Q^b(\mathbf{x}_n)Q^c(\mathbf{x}_r), \tag{B2}$$

where $\mathbf{Q}(\mathbf{x}_i)$ is the displacement of particle i from lattice position $\mathbf{x}_i$. By introducing creation and annihilation operators for phonons, $a_k^\dagger$ and a_k, respectively [see Ref. 1, (3.6) and (3.12)], and neglecting the commutation relations $[a_i, a_k^\dagger] = \delta_{ik}$, Eq. (B2) becomes

$$\lambda V_3 = \frac{1}{2N^{\frac{1}{2}}}\sum_{jkl}\hat{b}_{jkl}[\tfrac{1}{3}(a_ja_ka_l - a_{-j}^\dagger a_{-k}^\dagger a_{-l}^\dagger) + (a_ja_{-k}^\dagger a_{-l}^\dagger - a_{-j}^\dagger a_ka_l)], \tag{B3}$$

where

$$\hat{b}_{jkl} = \left(\frac{\hbar}{2m}\right)^{\frac{3}{2}}\Delta_{j+k+l}(\omega_j\omega_k\omega_l)^{-\frac{1}{2}} \times \sum_{abc} e_j^a e_k^b e_l^c \sum_{mn} B^{abc}(0, \mathbf{x}_m, \mathbf{x}_n)e^{i(\mathbf{k}\cdot\mathbf{x}_m + l\cdot\mathbf{x}_n)} \tag{B4}$$

(b_{jkl} is used in Ref. 4; $b_{jkl} = N^{-\frac{1}{2}}\hat{b}_{jkl}$). The additional terms which the commutation relations introduce into the complete expression for λV_3 are unimportant as they contain an additional Kronecker δ, so that they vanish when the limit $V \to \infty$ is taken.

The Term $\mathbf{I}_{3,k} + \mathbf{I}_{4,k} + \mathbf{I}_{5,k}$

The evaluation of $\mathbf{I}_{3,k} + \mathbf{I}_{4,k} + \mathbf{I}_{5,k}$ for $\lambda H' = \lambda V_3$ requires a knowledge of the value of

$$\mathrm{Re}\,(\langle\sigma|\,\lambda V_3\,|\beta\rangle^*\langle\sigma|\,\mathbf{S}^\circ{}_{nd}\,|\beta\rangle)$$

[see (2.51)]. Since λV_3 is a cubic function of creation and annihilation operators, only the part of $\mathbf{S}^\circ{}_{nd}$ that is also cubic contributes to $\mathbf{I}_{3k,} + \mathbf{I}_{4,k} + \mathbf{I}_{5,k}$. The cubic part of $\mathbf{S}^\circ{}_{nd}$ is [Ref. 1, (3.12) and (3.33)]

$$\mathbf{S}_3^\circ = \frac{1}{2VN^{\frac{1}{2}}}\sum_{jkl}[\mathbf{s}_{jkl}^{++}(a_ja_ka_l + a_{-j}^\dagger a_{-k}^\dagger a_{-l}^\dagger) + \mathbf{s}_{jkl}^{+-}(a_ja_{-k}^\dagger a_{-l}^\dagger + a_{-j}^\dagger a_ka_l) + \mathbf{s}_{jkl}^{-+}(a_{-j}^\dagger a_{-k}^\dagger a_l + a_ja_ka_{-l}^\dagger) + \mathbf{s}_{jkl}^{--}(a_{-j}^\dagger a_ka_{-l}^\dagger + a_ja_{-k}^\dagger a_l)] + \mathrm{H.c.}, \tag{B5}$$

where

$$\mathbf{s}_{jkl}^{(a)(b)} \equiv \frac{i}{2}\left(\frac{\hbar}{2}\right)^{\frac{3}{2}}\frac{1}{m^{\frac{1}{2}}}\Delta_{j+k+l}\mathbf{e}_j(\mathbf{e}_k\cdot\mathbf{e}_l) \times [(\omega_j\omega_k\omega_l)^{\frac{1}{2}} - (a)(\omega_j\omega_l^3/\omega_k)^{\frac{1}{2}} - (b)(\omega_l/\omega_j\omega_k)^{\frac{1}{2}}(\omega_l^2 - \omega_k^2)], \tag{B6}$$

where a and b equal $+1$ or -1. It follows from (B6) that

$$-\mathbf{s}_{jkl}^{(a)(b)*} = \mathbf{s}_{jkl}^{(a)(b)} = -\mathbf{s}_{-j-k-l}^{(a)(b)}. \tag{B7}$$

It is apparent from this that the quantity in brackets in (B5) is Hermitian as given. Ignoring the commutation relations, one finds that

$$\begin{aligned}&\mathrm{Re}\,(\langle\sigma|\,\lambda V_3\,|\beta\rangle^*\langle\sigma|\,\mathbf{S}_{nd}^\circ\,|\beta\rangle)\\ &= \frac{1}{2VN}\sum_{jkl}\{\mathrm{Re}\,[\hat{b}_{jkl}^*\,\mathbf{s}_{jkl}^{++}]\\ &\quad\times 3!\,(\tfrac{1}{3}\,|\langle\sigma|\,a_ja_ka_l\,|\beta\rangle|^2 - \tfrac{1}{3}\,|\langle\sigma|\,a_{-j}^\dagger a_{-k}^\dagger a_{-l}^\dagger\,|\beta\rangle|^2)\\ &\quad+\mathrm{Re}\,[\hat{b}_{jkl}^*\,(\mathbf{s}_{jkl}^{+-} + \mathbf{s}_{kjl}^{--} + \mathbf{s}_{ljk}^{-+})]\\ &\quad\times 2!\,(|\langle\sigma|\,a_ja_{-k}^\dagger a_{-l}^\dagger\,|\beta\rangle|^2 - |\langle\sigma|\,a_{-j}^\dagger a_ka_l\,|\beta\rangle|^2)\},\end{aligned} \tag{B8}$$

where the 3! and 2! give the number of ways the subscripts in λV_3 can be matched with the subscripts in $\lambda\mathbf{S}^\circ{}_3$ so that the matrix elements $\langle\sigma|\,\lambda V_3\,|\beta\rangle^*$ and $\langle\sigma|\,\mathbf{S}^\circ{}_{nd}\,|\beta\rangle$ are both nonzero. One can show, using (B4) and (B7), that

$$\mathrm{Re}\,[\hat{b}_{jkl}^*\,\mathbf{s}_{jkl}^{(a)(b)}] = -\mathrm{Re}\,[\hat{b}_{-j-k-l}^*\,\mathbf{s}_{-j-k-l}^{(a)(b)}]. \tag{B9}$$

The substitution of (B8) into (2.51) with the aid of (B9) and lemmas (2.35)–(2.37) gives

$$\begin{aligned}&\mathbf{I}_{3,k'} + \mathbf{I}_{4,k'} + \mathbf{I}_{5,k'} = -[d\langle N_{k'}\rangle_0/dT](\hbar\omega_{k'})^{-1}\\ &\times\frac{4}{N}\sum_{jkl}\{\mathrm{Re}\,[\hat{b}_{jkl}^*\,\mathbf{s}_{jkl}^{++}]\frac{\mathcal{P}}{\omega_j+\omega_k+\omega_l}\\ &\times(\delta_{k'j}\langle N_k + N_l + 1\rangle_0 + \delta_{k'k}\langle N_j + N_l + 1\rangle_0\\ &+\delta_{k'l}\langle N_j + N_k + 1\rangle_0)\\ &-\mathrm{Re}\,[\hat{b}_{jkl}^*\,(\mathbf{s}_{jkl}^{+-} + \mathbf{s}_{klj}^{--} + \mathbf{s}_{ljk}^{-+})]\\ &\times\frac{\mathcal{P}}{\omega_j-\omega_k-\omega_l}(\delta_{k'j}\langle N_k + N_l + 1\rangle_0\\ &+\delta_{k'k}\langle N_l - N_j\rangle_0 + \delta_{k'l}\langle N_k - N_j\rangle_0)\},\end{aligned} \tag{B10}$$

where the limit $V \to \infty$ is understood.

Both $\hat{b}_{jkl}$ and $\mathbf{s}_{jkl}^{(a)(b)}$ contain Δ_{j+k+l} as a factor. Since $(\Delta_{j+k+l})^2 = \Delta_{j+k+l}$, there is only one Δ-function factor in (B10). Using this, one can easily verify that $\mathbf{I}_{3,k} + \mathbf{I}_{4,k} + \mathbf{I}_{5,k}$ has the correct number of factors of V, etc., for convergence to a finite value in the limit $V \to \infty$. That $\mathbf{I}_{3,k} + \mathbf{I}_{4,k} + \mathbf{I}_{5,k}$ is real is obvious. Finally, by using (B9) one can prove that

$$\mathbf{I}_{3,k} + \mathbf{I}_{4,k} + \mathbf{I}_{5,k} = -\mathbf{I}_{3,-k} - \mathbf{I}_{4,-k} - \mathbf{I}_{5,-k}. \tag{B11}$$

$\mathbf{D}_{kl}$ and $\mathbf{V}_k$

λV_3 gives rise to a contribution to the energy flux operator of the form $\lambda\mathbf{S}' = \lambda\mathbf{S}_3' + \lambda\mathbf{S}_4'$ [see Ref.

1, (4.8)], where $\lambda S_3'$ is a cubic function of the $a_k^\dagger$ and a_k and $\lambda S_4'$ is a quartic function. Obviously $\langle\alpha|\ \lambda S_3'\ |\alpha\rangle = 0$, so that

$$\langle\alpha|\ \lambda S'\ |\alpha\rangle = \langle\alpha|\ \lambda S_4'\ |\alpha\rangle. \tag{B12}$$

According to Ref. 1, (4.14), one has

$$\lambda S_4' = \frac{1}{2VN}\{\sum_{ijkl} \mathbf{b}_{ijkl}(a_i - a_{-i}^\dagger)(a_j - a_{-j}^\dagger) \times (a_k - a_{-k}^\dagger)(a_l + a_{-l}^\dagger)\} + \text{H.c.}, \tag{B13}$$

where

$$\mathbf{b}_{ijkl} = \frac{(-i\hbar^2)}{24m^2}\Delta_{i+j+k+l}\left(\frac{\omega_l}{\omega_i\omega_j\omega_k}\right)^{\frac{1}{2}} \times \sum_{abc} e_i^b e_k^c \sum_{mn} B^{abc}(0, \mathbf{x}_m, \mathbf{x}_n)e^{i(\mathbf{i}\cdot\mathbf{x}_m + \mathbf{k}\cdot\mathbf{x}_n)} \times [e_i^a\mathbf{e}_l + 2\mathbf{e}_i e_l^a(e^{i(\mathbf{i}\cdot\mathbf{x}_m)} - 1)]. \tag{B14}$$

Note that

$$-\mathbf{b}_{-i-j-k-l}^* = \mathbf{b}_{ijkl}. \tag{B15}$$

A straightforward but somewhat lengthy calculation starting with (B13) leads to

$$\langle\alpha|\ \lambda S_4'\ |\alpha\rangle = \frac{2}{VN}\sum_{jk} \text{Re}\,(\mathbf{b}_{-j-kjk} + \mathbf{b}_{-k-jjk} + \mathbf{b}_{-jj-kk}) \times N_k(\alpha)[N_j(\alpha) + N_{-j}(\alpha) + 1], \tag{B16}$$

where (B15) has been used and where the addition of the Hermitian conjugate (+H.c.) results in the real part of $\mathbf{b}_{jk-j-k}$, etc. being taken.

The comparison of (B16) with (2.22) shows that

$$\mathbf{D}_{jk} = 2(V/N)\ \text{Re}\,(\mathbf{b}_{-j-kjk} + \mathbf{b}_{-k-jjk} + \mathbf{b}_{-jj-kk}). \tag{B17}$$

The particular combination of subscripts on the $\mathbf{b}_{ijkl}$'s in (B16) are such that the subscripts on the Δ function in (B14) always add up to zero, in which case $\Delta_{j+k+l} = 1$; thus, no factor of V is needed to compensate for the Δ functions; so that $\mathbf{b}_{-j-kjk}$, $\mathbf{b}_{kj-j-k}$, and $\mathbf{b}_{j-jk-k}$ converge to a finite value in the limit $V \to \infty$. Consequently, $\mathbf{D}_{jk}$ converges to a finite value. $\mathbf{V}_k$ is related to the $\mathbf{D}_{jk}$ by (2.28) and is thus also finite. That $\mathbf{D}_{jk}$ and $\mathbf{V}_k$ are real is obvious. Finally, it follows from (B15), (B17), and (2.28) that

$$\mathbf{D}_{jk} = -\mathbf{D}_{-j-k} \quad \text{and} \quad \mathbf{V}_k = -\mathbf{V}_{-k}. \tag{B18}$$

General Spherical Harmonic Tensors in the Boltzmann Equation

TUDOR WYATT JOHNSTON
RCA Victor Company, Ltd., Research Laboratories, Montreal, Canada
(Received 6 January 1966)

The irreducible velocity space direction cosine tensors associated with velocity magnitude spherical harmonic expansion of the distribution function are manipulated in the Boltzmann–Vlasov flow terms to yield a linked chain of equations whose general (lth) equation is given explicitly. This generalizes earlier results for $l = 0, 1, 2, 3$.

INTRODUCTION

THE object of this note is the presentation of a simple derivation from the Boltzmann equation of the general set of equations for the irreducible base tensors associated with the velocity space spherical harmonic expansion of the one-particle distribution function for charged particles.

To the author's knowledge, Wallace[1] was the first (in connection with neutron transport) to give a general explicit direction cosine tensor generalization of the spherical harmonics themselves. Ikenberry[2] also evolved an equivalent form for statistical mechanics problems. Unfortunately, this work, unknown to the author in 1960, was not mentioned before.[3] Delcroix[4] has hinted at the tensor application in the Boltzmann equation, based on the spherical harmonic work of Jancel and Kakan.[5]

The next step is to obtain the equations resulting from the substitution into the Boltzmann equation. Allis[6] had given the zero-order (scalar) and first-order (vector) equations and the general one-dimensional form in which the spherical harmonics and the tensors reduce to Legendre polynomials. The author[3] then derived the tensor equations up to the third order and Shkarofsky[7] included the intrinsic velocity effects up to the second order, but each case was calculated separately. The (successful) object of this work was to obtain the general form for the equations to all orders.

Using a bit of hindsight, together with the extremely useful approach developed by Wallace,[1] the general tensor equation including intrinsic velocity for any order is derived here in a manner much simpler than the brute force methods[3,7] previously used for the second-order and third-order results. The tensor equations are far more compact, symmetric and understandable than the clumsy spherical harmonic result.

SPHERICAL HARMONIC TENSORS

Owing to the habit of using powers of v with coefficients of one in the velocity moment equations, the tensor form used here differs by a numerical constant C_l from that of Wallace[1] but agrees with Ikenberry[2] in having the first term coefficient equal to 1. The fully symmetric lth-order tensor $\mathbf{T}_l$ is therefore defined as follows:

$$\mathbf{T}_l(\mathbf{u}) = T_l\left(\frac{\mathbf{v}}{v}\right) = \frac{(-1)^l}{l!}\, C_l \nabla_{\mathbf{v}}^l \left(\frac{1}{v}\right)_{v=1}, \qquad (1)$$

where

$$C_l \equiv \frac{2^l l!\, l!}{(2l)!} = \frac{l!}{1\cdot 3\cdot 5 \cdots (2l-1)},$$

$\mathbf{v}$ is the velocity vector with magnitude v, $\nabla_{\mathbf{v}}$ is the gradient operator in velocity space, and $\mathbf{u} = \mathbf{v}/v$ is the velocity direction cosine vector of unit magnitude.

As Wallace[1] points out, $1/v$ is a solution of the Laplace equation in velocity space, i.e., $\nabla_{\mathbf{v}}^2(1/v)$ is zero, therefore $\mathbf{T}_l$ is an irreducible or base tensor, one for which any contraction gives zero ($\sum_i \mathbf{T}_l \cdots i \cdots i \cdots = 0$). Each of the $\frac{1}{2}(l+1)(l+2)$ elements of $\mathbf{T}_l$ is a linear combination of the $2l + 1$ spherical harmonics of order l, but the $\frac{1}{2}l(l-1)$ conditions from the irreducibility feature leave just $2l + 1$ independent elements.[1,3] An equivalent situation exists in considering multipoles and spherical harmonics in electrostatic problems.[8]

The z^l element $T_{l(z)}$ is particularly simple, being

[1] P. R. Wallace, Can. J. Res. **A26**, 99 (1948).
[2] E. Ikenberry, Ann. Math. Monthly **62**, 719 (1955); E. Ikenberry and C. Truesdell, J. Ratl. Mech. Anal. **5**, 1 (1956); J. Math Anal. Appl. **3**, 355 (1961).
[3] T. W. Johnston, Phys. Rev. **120**, 1103, 2277 (1960).
[4] J. L. Delcroix, *Introduction à la théorie des gaz ionisés* (Dunod Cle., Paris, 1959), p. 69 [English transl.: *Introduction to the Theory of Ionized Gases* (Interscience Publishers, Inc., New York, 1960), p. 59].
[5] R. Jancel and T. Kahan, J. Phys. Radium **20**, 35, 804 (1959); later work [by C. A. Carpenter and F. W. Metzger, J. Math. Phys. **2**, 694 (1961)] appears to be very similar.
[6] W. P. Allis, in *Handbuch der Physik*, S. Flügge, Ed. (Springer-Verlag, Berlin, 1956), Vol. 21, pp. 404–408.
[7] I. P. Shkarofsky, Can. J. Phys. **41**, 1776 (1963).
[8] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), pp. 1276–1283; M. H. Cohen, Phys. Rev. **95**, 674 (1954).

equal to the product of C_l and the lth-order Legendre polynomial in μ_z,

$$T_{l(z)} = C_l P_l(\mu_z) = \mu_z^l - \frac{l(l-1)}{2}\frac{\mu^{l-2}}{(2l-1)} + \frac{l(l-1)(l-2)(l-3)}{8(2l-1)(2l-3)}\mu^{l-4} \mp \cdots .$$

By unique extension the general element is then

$$\mathbf{T}_l(\mathfrak{y}) = \mathfrak{y}^l - \frac{l(l-1)}{2}[\mathbf{I}\mathfrak{y}^{l-2}]_l + \frac{l(l-1)(l-2)(l-3)}{8(2l-1)(2l-3)}[\mathbf{II}\mathfrak{y}^{l-4}]_l \mp \cdots . \quad (2)$$

Here $\mathfrak{y}^l$ is the symmetric lth-order vector product tensor of $\mathfrak{y}$ and $[\ ]_l$ denotes an lth-order symmetrization operation, adding all the $l!$ permutations and dividing by $l!$, $\mathbf{I}$ is the diagonal limit or identity tensor.

Wallace showed that the complete contraction or scalar product of two lth-order spherical tensors of different argument is simply given by the formula

$$\mathbf{T}_l(\mathfrak{y})_l \cdot \mathbf{T}_l(\mathfrak{y}') = C_l P_l(\mathfrak{y}\cdot\mathfrak{y}').$$

This means that $f(v)$ can be expanded as follows:

$$f(\mathbf{v}) = \frac{1}{4\pi}\sum_l (2l+1)\int f(v')P_l(\mathfrak{y}\cdot\mathfrak{y}')\,d^2\Omega' = \frac{1}{4\pi}\sum_l \frac{(2l+1)}{C_l}\left\{\int f(v')\mathbf{T}_l(\mathfrak{y}')\,d^2\Omega\right\}_l \cdot \mathbf{T}_l(\mathfrak{y}) \quad (3)$$
$$= \sum_l \mathbf{f}_l(v)_l \cdot \mathbf{T}_l(\mathfrak{y}),$$

where[9]

$$\mathbf{f}_l \equiv \frac{2l+1}{4\pi C_l}\int f(v)\mathbf{T}_l(\mathfrak{y})\,d^2\Omega. \quad (4)$$

Note that the $f_{l(z)}$ element is just the coefficient of the lth-order $m = 0$ Legendre polynomial in the spherical harmonic expansion,[3] for from Eq. (4) we have

$$f_{l(z)} = \frac{2l+1}{4\pi C_l}\int f(v)T_{l(z)}\,d^2\Omega = \frac{2l+1}{4\pi}\int f(v)P_l(\mu_z)\,d^2\Omega \equiv f_{l00}.$$

Thus, the $f_{l(z)}$ tensor equation can be checked immediately with the polar spherical harmonic ($m = 0$) result given by Allis.[6]

Because $\mathbf{T}_l$ is irreducible, any contraction on $\mathbf{f}_l$ which gives a nonzero result cannot appear in the result and should be eliminated, and indeed this is the result of the definition of Eq. (4) for $\mathbf{f}_l$. Note that once we have made $\mathbf{f}_l$ irreducible by definition[10] then other $\mathfrak{y}^l$ polynomials can be used and, in particular, using Eq. (2) and the fact that $\mathbf{f}_l \cdot [\mathbf{I}_2\mathfrak{y}^{l-2}] = 0$, the following combinations are equivalent:

$$\mathbf{f}_l \vdots\, \mathbf{T}_l = \mathbf{f}_l \vdots\, \mathfrak{y}^l. \quad (5)$$

It is this equivalence that enables the simple derivation of the chain of tensor equations. Note that the only property required for $\mathbf{f}_l$ is irreducibility. The same result will hold if the tensor is not fully symmetric. We have (since $\mathfrak{y}^l$ and $\mathbf{T}_l$ are fully symmetric) the following result {[ ]$_l$ defined after Eq. (2)}:

$$\mathbf{g}_l \vdots\, \mathfrak{y}^l = \mathbf{g}_l \vdots\, \mathbf{T}_l = [\mathbf{g}_l] \vdots\, \mathbf{T}_l = [\mathbf{g}_l] \vdots\, \mathfrak{y}^l.$$

In that case, however, the tensor obtained by tensor multiplication by $\mathbf{T}_l$ and integration over angle is $[\mathbf{g}_l]_l$ the symmetrized version of the arbitrary irreducible tensor $\mathbf{g}_l$.

BOLTZMANN–VLASOV EQUATION

The collision terms are not discussed here. Shkarofsky[11] has treated the Fokker–Planck equation and spherical harmonic tensors in considerable detail and electron–neutron collision effects are well known.[6,12] Only the flow terms, those common to the Boltzmann and Vlasov equations, are discussed. To tackle the problem in two stages, the straightforward extrinsic flow terms referred to a rest frame are treated first, providing the generalization for the particular equations given previously.[3,6] The intrinsic velocity (velocity referred to some velocity $\mathbf{C}$) generalization of the particular results of Shkarofsky[12] (which introduce additional terms) are then derived.

EXTRINSIC VELOCITY EQUATION

The application to the Boltzmann–Vlasov flow terms is at first just like the previous work.[3] The flow terms $D(f)$ are as follows:

$$D(f) \equiv \frac{\partial f}{\partial t} + \mathbf{v}\cdot\nabla f + (\mathbf{a} + \mathbf{v}\times\omega_b)\cdot\nabla_v f, \quad (6)$$

where ∇ is the configuration space gradient operator, ∇_v is the velocity space gradient operator, $\mathbf{a}$ is the velocity-independent acceleration; $\mathbf{a}$ =

[9] One should not leap to the incorrect conclusion that the $\mathbf{T}_l$ are orthogonal in angle integration. Contributions to an $\mathbf{f}_l$ element in (4) come from other elements as well.

[10] An indication of this is given in the book by A. Sommerfeld, *Lectures on Theoretical Physics, Vol. 5, Thermodynamics and Statistical Mechanics* (Academic Press, Inc., New York, 1964), p. 338. (This section was actually completed after the author's death by F. Bopp and J. Meixner.)

[11] T. W. Johnston, Can. J. Phys. 41, 1208 (1962).

[12] I. P. Shkarofsky, Can. J. Phys. 41, 1753 (1963).

$(q/m)\mathbf{E} - \nabla\psi$ with q, m the particle charge and mass, $\mathbf{E}$ the electric field, ψ the gravitational potential, and $\boldsymbol{\omega}_b$ is the magnetic cyclotron angular frequency vector $q\mathbf{B}/m$ so that $\mathbf{v}\times\boldsymbol{\omega}_b$ is the acceleration due to the magnetic field. Substituting the irreducible direction cosine expansion of Eq. (3) gives, as before,[3]

$$D(f) = \sum_l \frac{\partial \mathbf{f}_l}{\partial t}\cdot\underset{l}{\boldsymbol{\mu}}^l + v\nabla\mathbf{f}_l \cdot \underset{l+1}{\boldsymbol{\mu}}^{l+1} + av^l \frac{\partial}{\partial v}\left(\frac{\mathbf{f}_l}{v^l}\right)\cdot \underset{l+1}{\boldsymbol{\mu}}^{l+1} + \frac{l\mathbf{a}\cdot\mathbf{f}_l}{v}\cdot \underset{l-1}{\boldsymbol{\mu}}^{l-1} + l\boldsymbol{\omega}_b\times\mathbf{f}_l\cdot\underset{l}{\boldsymbol{\mu}}^l = 0.$$

As before we wish to group the terms by $\boldsymbol{\mu}^l$ rather than by $\mathbf{f}_l$. Now, however, we recognize the special value of arranging to have irreducible tensor forms multiplying $\boldsymbol{\mu}^l$, $\boldsymbol{\mu}^{l+1}\boldsymbol{\mu}^{l-1}$. We can symmetrize immediately but irreducibility must be contrived. It is evident that only the $\boldsymbol{\mu}^{l+1}$ terms require special treatment; the coefficients of $\boldsymbol{\mu}^l$ and $\boldsymbol{\mu}^{l-1}$ are already irreducible. Both $\boldsymbol{\mu}^{l+1}$ terms are of the form $\mathbf{Af}_l$. In order to form the irreducible tensors, we add and subtract the nonzero results of the contractions of $\mathbf{Af}_l$. The only nonzero contraction is $A\cdot f_l$, and the form we require is

$$[\mathbf{Af}_l]_{l+1} = [\mathbf{Af}_l - \alpha(\mathbf{A}\cdot\mathbf{f}_l)\mathbf{I}]_{l+1} + \alpha[(\mathbf{A}\cdot\mathbf{f}_l)\mathbf{I}]_{l+1}.$$

The required coefficient of α is that which will make the first tensor on the right zero on any contraction. There are l indices in $\mathbf{f}_l$ to choose in contracting $[\mathbf{Af}_l]_{l+1}$ and $2(l-1)+3 = 2l+1$ in $[\mathbf{A}\cdot\mathbf{f}_l\mathbf{I}]_{l+1}$, since contraction on $\mathbf{I}$ gives 3 and either index in $\mathbf{I}$ may be equated with the $l-1$ free indices in $\mathbf{A}\cdot\mathbf{f}_l$. Thus, α is $l/(2l+1)$ and the required form is

$$\underset{l}{\mathbf{Af}_l}\cdot\boldsymbol{\mu}^{l+1} = \underset{l}{[\mathbf{Af}_l]}\cdot\boldsymbol{\mu}^{l+1} = \left[\mathbf{Af}_l - \frac{l}{2l+1}\mathbf{A}\cdot\mathbf{f}_l\mathbf{I}\right]_{l+1}\cdot\boldsymbol{\mu}^{l+1} + \frac{l}{2l+1}\underset{l-1}{[\mathbf{A}\cdot\mathbf{f}_l]}\cdot\boldsymbol{\mu}^{l-1}.$$

(We have used $\mathbf{I}:\boldsymbol{\mu}\boldsymbol{\mu} = \boldsymbol{\mu}\cdot\boldsymbol{\mu} = 1$.) Note that $\mathbf{A}\cdot\mathbf{f}_l$ is irreducible as well.

We can therefore write Eq. (7) as follows:

$$D(f) = \sum_l \frac{\partial \mathbf{f}_l}{\partial t}\cdot\underset{l}{\boldsymbol{\mu}}^l + \left\{v\left(\nabla\mathbf{f}_l - \frac{l}{2l+1}\mathbf{I}\nabla\cdot\mathbf{f}_l\right) + v^l\frac{\partial}{\partial v}\left[\frac{1}{v^l}\left(\mathbf{af}_l - \frac{l}{2l+1}\mathbf{I}\mathbf{a}\cdot\mathbf{f}_l\right)\right]\right\}_{l+1}\cdot\boldsymbol{\mu}^{l+1} + l[\boldsymbol{\omega}_b\times\mathbf{f}_l]\cdot\underset{l}{\boldsymbol{\mu}}^l + \frac{l}{2l+1}\left[v\nabla\cdot\mathbf{f}_l + v^l\frac{\partial}{\partial v}\left(\frac{\mathbf{a}\cdot\mathbf{f}_l}{v^l}\right) + \frac{l\mathbf{a}\cdot\mathbf{f}_l}{v}\right]_{l-1}\cdot\boldsymbol{\mu}^{l-1}.$$

Grouping now by $\boldsymbol{\mu}^l$ and using the following identity on the $\mathbf{a}\cdot\mathbf{f}_{l-1}\underset{l-1}{\boldsymbol{\mu}}^{l-1}$ terms,

$$\frac{1}{v^p}\frac{\partial v^p f}{\partial v} = \frac{pf}{v} + \frac{\partial f}{\partial v}, \tag{7}$$

we obtain

$$D(f) = \sum_l \underset{l}{\boldsymbol{\mu}}^l\cdot\mathbf{D}_l \equiv \sum_l \underset{l}{\boldsymbol{\mu}}^l\cdot\left\{\frac{\partial\mathbf{f}_l}{\partial t} + v\left(\nabla_r\mathbf{f}_{l-1} - \frac{l-1}{2l-1}\mathbf{I}\nabla_r\cdot\mathbf{f}_{l-1}\right) + v^{l-1}\frac{\partial}{\partial v}\left(\frac{\mathbf{af}_{l-1}}{v^{l-1}} - \frac{l-1}{2l-1}\mathbf{I}\frac{\mathbf{a}\cdot\mathbf{f}_{l-1}}{v^{l-1}}\right) + l\boldsymbol{\omega}_b\times\mathbf{f}_l + \frac{l+1}{2l+3}\left[v\nabla\cdot\mathbf{f}_{l+1} + \frac{1}{v^{l+2}}\frac{\partial}{\partial v}(v^{l+2}\mathbf{a}\cdot\mathbf{f}_{l+1})\right]\right\}_l. \tag{8}$$

Now, because $\mathbf{D}_l$ is irreducible we can immediately apply Eq. (4), that is multiply by $\mathbf{T}_l$ and integrate over angle to obtain the $\mathbf{D}_l$ elements in a chain of equations which contain $\mathbf{f}_{l-1}$, $\mathbf{f}_l$, and $\mathbf{f}_{l+1}$. Provided one can do this for the collisions as well ($C = \sum_l \mathbf{C}_l\cdot\boldsymbol{\mu}^l$, with irreducible $\mathbf{C}_l$), then the result is the chain of equations

$$\mathbf{C}_l = \mathbf{D}_l = \left[\frac{\partial\mathbf{f}_l}{\partial t} + v\left(\nabla\mathbf{f}_{l-1} - \frac{l-1}{2l-1}\mathbf{I}\nabla\cdot\mathbf{f}_{l-1}\right) + v^{l-1}\frac{\partial}{\partial v}\left(\frac{\mathbf{af}_{l-1}}{v^{l-1}} - \frac{l-1}{2l-1}\mathbf{I}\frac{\mathbf{a}\cdot\mathbf{f}_{l-1}}{v^{l-1}}\right) + l\boldsymbol{\omega}_b\times\mathbf{f}_l + \frac{l+1}{2l+3}\left(v\nabla\cdot\mathbf{f}_{l+1} + \frac{1}{v^{l+2}}\frac{\partial}{\partial v}(v^{l+2}\mathbf{a}\cdot\mathbf{f}_{l+1})\right)\right]_l \tag{9}$$

instead of the original equation

$$C = D.$$

The $\mathbf{D}_l$ elements for $\mathbf{f}_l$ with l less than 2 were obtained by Allis[6] but irreducibility is not necessary. The equations for $\mathbf{D}_l$ and $\mathbf{f}_l$ with $l = 2, 3$ were given by Johnston[3] by direct calculation and agree with Eq. (9). The calculation shown here gives the result for all l with less labor than that required for $l = 2$ or $l = 3$ by the direct approach. The tensor form of $\mathbf{D}_l$ in Eq. (9) is far more symmetric and compact than the direct spherical harmonic form.[5]

Another check is the one-dimensional result, for which the magnetic field is along the z axis, say, and Eq. (8) becomes simply

$$D_l = \frac{\partial f_{l(z)}}{\partial t} + \frac{1}{2l-1}\times\left\{\frac{\partial}{\partial z}f_{l-1(z)} + v^{l-1}a_z\frac{\partial}{\partial v}\left[\frac{f_{l-1(z)}}{v^{l-1}}\right]\right\} + \frac{l+1}{2l+3}\left(\frac{v\,\partial f_{l+1(z)}}{\partial z} + \frac{a_z}{v^{l+2}}\frac{\partial v^{l+2}f_{l+1(z)}}{\partial z}\right).$$

Since, as pointed out above $f_{l(s)} = f_{l00}$, this can be compared directly with Allis'[6] one-dimensional result obtained from Legendre polynomial recursion relations; the two results are identical.

INTRINSIC VELOCITY EQUATION

The intrinsic velocity flow term $D^w(f)$ involves the intrinsic velocity $\mathbf{w} = \mathbf{v} - \mathbf{C}$, where $\mathbf{C}$ is the reference velocity. The intrinsic flow terms, as given by Bernstein and Trehan,[13] are

$$D^w(f) = \frac{df}{dt} + \mathbf{w}\cdot\nabla f + (\mathbf{h} + \mathbf{w}\times\boldsymbol{\omega}_b)\cdot\nabla_w f - \mathbf{w}\cdot\nabla\mathbf{C}\cdot\nabla_w f, \qquad (10)$$

where

$$d/dt = \partial/\partial t + \mathbf{C}\cdot\nabla,$$

$$\mathbf{h} \equiv \mathbf{a} + \mathbf{C}\times\boldsymbol{\omega}_b - d\mathbf{C}/dt.$$

Define

$$D^{wa}(f) \equiv \frac{df}{dt} + \mathbf{w}\cdot\nabla f + (\mathbf{h} + \mathbf{w}\times\boldsymbol{\omega}_b)\cdot\nabla_w f,$$

$$D^{wb}(f) \equiv -\mathbf{w}\cdot\nabla\mathbf{C}\cdot\nabla_w f.$$

Thus

$$D^{wb}(f) = D^{wa}(f) + D^{wb}(f).$$

Evidently the $D^{wa}(f)$ term is just like $D(f)$ with d/dt replacing $\partial/\partial t$, $\mathbf{w}$ replacing $\mathbf{v}$ and $\mathbf{h}$ replacing $\mathbf{a}$, and the final result for $D_l^{wa}(f)$ can be obtained with these transformations in Eq. (8).

The term D^{wb} needs more treatment. Substitution of $\mathbf{f}_l(w)\mathbf{w}^l$ in D^{wb} gives the result quoted by Shkarofsky[7] [his Eq. (9)]:

$$D^{wb}(f) = -\mathbf{w}\cdot\nabla\mathbf{C}\cdot\nabla_w f = \sum_l - l\nabla\mathbf{C}\cdot\mathbf{f}_{ll}\mathfrak{y}^l - \nabla\mathbf{C}w^{l+1}\frac{\partial}{\partial w}\left(\frac{\mathbf{f}_l}{w^l}\right)_{l+2}\cdot\mathfrak{y}^{l+2},$$

where $\mathfrak{y}$ now is given by $\mathfrak{y} = \mathbf{w}/w$.

The $\nabla\mathbf{C}\cdot\mathbf{f}_l$ term is easily dealt with by the same type of reasoning as above. We see that it can be written as

$$-l\nabla\mathbf{C}\cdot\mathbf{f}_{ll}\mathfrak{y}^l = -\left[\nabla\mathbf{C}\cdot\mathbf{f}_l - \frac{l-1}{2l-1}\mathsf{I}\nabla\mathbf{C}:\mathbf{f}_l\right]_{ll}\mathfrak{y}^l + \frac{l-1}{2l-1}\nabla\mathbf{C}:\mathbf{f}_l\cdot\mathfrak{y}^{l-2}.$$

The last term in D^{wb} is of the form $\nabla\mathbf{C}\mathbf{f}_{ll}i_{+2}\mathfrak{y}^{l+2}$ and requires the subtraction and addition of two terms with I and II to reach the desired irreducible form.

If we contract and symmetrize $\nabla\mathbf{C}\mathbf{f}_l$, the result is $[\nabla\cdot\mathbf{C}\mathbf{f}_l + 2l\mathbf{f}_l\cdot[\nabla\mathbf{C}]_2]_l$; hence the terms to be subtracted and added are in the form

$$\pm\,[-\beta\mathsf{I}\nabla\cdot\mathbf{C}\mathbf{f}_l + 2\gamma\mathsf{I}\mathbf{f}_l\cdot[\nabla\mathbf{C}]_2 + \delta\mathsf{II}\mathbf{f}_l:\nabla\mathbf{C}]_{l+2}.$$

Contracting, we have

$$\mathsf{I}:[\nabla\mathbf{C}\mathbf{f}_l - \beta\mathsf{I}\nabla\cdot\mathbf{C}\mathbf{f}_l - 2\gamma\mathsf{I}\mathbf{f}_l\cdot[\nabla\mathbf{C}]_2 - \delta\mathsf{II}\mathbf{f}_l:\nabla\mathbf{C}]_{l+2}$$
$$= [\nabla\cdot\mathbf{C}\mathbf{f}_l + 2l[\nabla\mathbf{C}]_2\cdot\mathbf{f}_l - \beta(2l+3)\nabla\cdot\mathbf{C}\mathbf{f}_l - 2\gamma(2l+3)[\nabla\mathbf{C}]_2\cdot\mathbf{f}_l - 2\gamma(l-1)\mathsf{I}\nabla\mathbf{C}:\mathbf{f}_l - (2\times 3 + 4 + 4(l-2))\delta\nabla\mathbf{C}:\mathbf{f}_l\mathsf{I}]_l.$$

Setting the coefficients of $\nabla\cdot\mathbf{C}\mathbf{f}_l$, $[\nabla\mathbf{C}]_2\cdot\mathbf{f}_l$, and $\nabla\mathbf{C}:\mathbf{f}_l$ equal to zero gives the following results:

$$\beta = \frac{1}{2l+3}, \qquad \gamma = \frac{l}{2l+3},$$

$$\delta = -\frac{(l-1)\gamma}{2l+1} = -\frac{l(l-1)}{(2l+1)(2l+3)}.$$

Thus, the expression

$$\left[\nabla\mathbf{C}\mathbf{f}_l - \frac{\mathsf{I}}{2l+3}(2l[\nabla\mathbf{C}]_2\cdot\mathbf{f}_l + \nabla\cdot\mathbf{C}\mathbf{f}_l) + \frac{l(l-1)}{(2l+1)(2l+3)}\mathsf{II}\nabla\mathbf{C}:\mathbf{f}_l\right]_{l+2}$$

is irreducible. We also require that the coefficients of the I and II terms are each irreducible. This is automatically true for $\mathbf{f}_l:\nabla\mathbf{C}$, the II coefficient and for $[\nabla\cdot\mathbf{C}\mathbf{f}_l]_l$ but not for the $[\nabla\mathbf{C}]_2\cdot\mathbf{f}_l$ combinations, to which a term of the form $\epsilon[\mathsf{I}\nabla\mathbf{C}:\mathbf{f}_l]_l$ must be added and subtracted.

We have already done this for the $\nabla\mathbf{C}\cdot\mathbf{f}_l$ coefficient earlier, the coefficient being $(l-1)/(2l-1)$ for each $\nabla\mathbf{C}\cdot\mathbf{f}_l$ term, so ϵ is then given by

$$-\epsilon = \frac{2l(l-1)}{(2l+3)(2l-1)}.$$

We have a term of the identical form already, so the final coefficient for the last II term is given below

$$\epsilon - \delta = \frac{l(l-1)}{2l+3}\left(\frac{1}{2l+1} - \frac{2}{2l-1}\right) = -\frac{l(l-1)}{(2l+1)(2l-1)}.$$

The final result is that we can write $\nabla\mathbf{C}\mathbf{f}_{ll}i_{+2}\mathfrak{y}^{l+2}$ as follows:

[13] I. B. Bernstein and S. K. Trehan, Nucl. Fusion 1, 3 (1960), Chap. 1, Eq. (42).

$$\nabla\mathbf{C}\mathbf{f}_l \cdot \underset{l+2}{\mathfrak{y}}^{l+2}$$
$$= \left(\nabla\mathbf{C}\mathbf{f}_l - \frac{\mathbf{I}}{2l+3}(2l[\nabla\mathbf{C}]_2\cdot\mathbf{f}_l + \nabla\cdot\mathbf{C}\mathbf{f}_l) + \frac{l(l-1)}{(2l+1)(2l+3)}\mathbf{II}\nabla\mathbf{C}:\mathbf{f}_l\right)_{l+2}\cdot\mathfrak{y}^{l+2} + \frac{\mathbf{I}}{2l+3}$$
$$\times\left(2l[\nabla\mathbf{C}]_2\cdot\mathbf{f}_l + \nabla\cdot\mathbf{C}\mathbf{f}_l - \frac{2(l-1)l}{2l-1}\mathbf{I}\nabla\mathbf{C}:\mathbf{f}_l\right)_l\cdot\mathfrak{y}^l + \frac{l(l-1)}{(2l+1)(2l-1)}(\nabla\mathbf{C}:\mathbf{f}_l)_{l+2}\cdot\mathfrak{y}^{l-2}.$$

All the $\mathfrak{y}^l$, $\mathfrak{y}^{l+2}$, $\mathfrak{y}^{l-2}$ coefficients are now irreducible. As before, we group by $\mathfrak{y}^l$ rather than by $\mathbf{f}_l$ and then isolate by multiplication by $\mathbf{T}_l(\mathfrak{y})$ and integration over w angle, the result being

$$\mathbf{D}_l^{wb}(f) = -w^{l-1}\frac{\partial}{\partial w}\left[\frac{1}{w^{l-2}}\left(\nabla\mathbf{C}\mathbf{f}_{l-2} - \frac{\mathbf{I}}{2l-1}\right.\right.$$
$$\times(2(l-2)[\nabla\mathbf{C}]_2\cdot\mathbf{f}_{l-2} + \nabla\cdot\mathbf{C}\mathbf{f}_{l-2})$$
$$\left.\left.+\frac{(l-2)(l-3)}{(2l-3)(2l-1)}\mathbf{II}\mathbf{C}:\mathbf{f}_{l-2}\right)\right]_l$$
$$-l\left[\nabla\mathbf{C}\cdot\mathbf{f}_l - \frac{l-1}{2l-1}\mathbf{I}\nabla\mathbf{C}:\mathbf{f}_l\right]_l$$
$$-\frac{w^{l+1}}{2l+3}\frac{\partial}{\partial w}\left[\frac{1}{w^l}\left(2l[\nabla\mathbf{C}]_2\cdot\mathbf{f}_l\right.\right.$$
$$\left.\left.+\nabla\cdot\mathbf{C}\mathbf{f}_l - \frac{2(l-1)l}{2l-1}\mathbf{I}\nabla\mathbf{C}\cdot\mathbf{f}_l\right)\right]_l$$
$$-\frac{(l+2)(l+1)}{(2l+3)(2l+5)}\frac{1}{w^{l+2}}\frac{\partial}{\partial w}(w^{l+3}\nabla\mathbf{C}:\mathbf{f}_{l+2}). \quad (11)$$

The identity of Eq. (7) has been used on the $\nabla\mathbf{C}:\mathbf{f}_{l+2}$ terms to collapse them into one term.

The final form, with both D^{wa} and D^{wb} included, is

$$\mathbf{D}_l^w(f) = D_l^{wa} + D_l^{wb}$$
$$= \left[\frac{d\mathbf{f}_l}{dt} + l\omega_b\times\mathbf{f}_l + w\left(\nabla\mathbf{f}_{l-1} - \frac{l-1}{2l-1}\mathbf{I}\nabla\cdot\mathbf{f}_{l-1}\right)\right.$$
$$+w^{l-1}\frac{\partial}{\partial w}\left(\frac{\mathbf{a}\mathbf{f}_{l-1}}{w^{l-1}} - \frac{l-1}{2l-1}\mathbf{I}\mathbf{a}\cdot\mathbf{f}_{l-1}\right)$$
$$\left.+\frac{l+1}{2l+3}\left[w\nabla\cdot\mathbf{f}_{l+1} + \frac{1}{w^{l+2}}\frac{\partial(w^{l+2}\mathbf{a}\cdot\mathbf{f}_{l+1})}{\partial w}\right]\right.$$
$$-w^{l-1}\frac{\partial}{\partial w}\left(\frac{1}{w^{l-2}}\right)\left\{\nabla\mathbf{C}\mathbf{f}_{l-2} - \frac{\mathbf{I}}{2l-1}\right.$$
$$\times\left[2(l-2)[\nabla\mathbf{C}]_2\cdot\mathbf{f}_{l-2} + \nabla\cdot\mathbf{C}\mathbf{f}_{l-2}\right]$$
$$\left.+\frac{(l-2)(l-3)}{(2l-3)(2l-1)}\mathbf{II}\nabla\mathbf{C}:\mathbf{f}_{l-2}\right\}$$
$$-l\left(\nabla\mathbf{C}\cdot\mathbf{f}_l - \frac{l-1}{2l-1}\mathbf{I}\nabla\mathbf{C}:\mathbf{f}_l\right)$$
$$-\frac{w^{l+1}}{2l+3}\frac{\partial}{\partial w}\left\{\frac{1}{w^l}\left[2l[\nabla\mathbf{C}]_2\cdot\mathbf{f}_l\right.\right.$$
$$\left.\left.+\nabla\cdot\mathbf{C}\mathbf{f}_l - \frac{2(l-1)l}{2l-1}\mathbf{I}\nabla\mathbf{C}:\mathbf{f}_l\right]\right\}$$
$$\left.-\frac{(l+1)(l+2)}{(2l+3)(2l+5)}\frac{1}{w^{l+2}}\frac{\partial}{\partial w}(w^{l+3}\nabla\mathbf{C}:\mathbf{f}_{l+2})\right]_l. \quad (12)$$

The result can be checked for the $D_{l(z)}$ element in the same way as for the extrinsic velocity case, by comparison with the Legendre polynomial result. Only the D_l^{wb} term needs checking. In the same way as before, we have in one dimension

$$-\mathbf{w}\cdot\nabla\mathbf{C}\cdot\nabla_w f = -\sum_l w\cos\theta \times\frac{\partial C_z}{\partial z}\cdot\left(\cos\theta\frac{\partial f_l}{\partial w} - \frac{\sin\theta}{w}f_l\frac{\partial P_l}{\partial\theta}\right)$$

with $f = \sum_l f_l P_l$. Using Legendre polynomial recursion relations twice and the identity of Eq. (7), we have

$$-\mathbf{w}\cdot\nabla\mathbf{C}\cdot\nabla_w f = -\frac{\partial C_z}{\partial z}\sum_l\left(P_{l-2}\frac{l(l-1)}{(2l+1)(2l-1)}\right.$$
$$\times\frac{1}{w^{l+1}}\frac{\partial(f_l w^{l+1})}{\partial w} + P_l\left\{\frac{l^2}{2l-1}\frac{f_l}{w}\right.$$
$$\left.\left.+\left[\frac{l^2}{2l-1}+\frac{(l+1)^2}{2l+3}\right]\frac{w^{l+1}}{2l+1}\frac{\partial}{\partial w}\left(\frac{f_l}{w^l}\right)\right\}\right.$$
$$\left.+P_{l+2}\frac{(l+1)(l+2)}{(2l+1)(2l+3)}w^l\frac{\partial}{\partial w}\left(\frac{f_l}{w^l}\right)\right).$$

Regrouping by P_l results in the following:

$$-\mathbf{w}\cdot\nabla\mathbf{C}\cdot\nabla_w f = +\sum P_l D_{l(z)}^{wb}$$

with

$$-D_{l(z)}^{wb} = \frac{\partial C_z}{\partial z}\left[\frac{(l-1)l}{(2l-3)(2l-1)}w^{l-1}\frac{\partial}{\partial w}\left(\frac{f_{l-2}}{w^{l-2}}\right)\right.$$
$$+\frac{l^2}{2l-1}f_l + \frac{2l(l+1)-1}{(2l-1)(2l+3)}w^{l+1}\frac{\partial}{\partial w}\left(\frac{f_l}{w^l}\right)$$
$$\left.+\frac{(l+1)(l+2)}{(2l+3)(2l+5)}\frac{1}{w^{l+2}}\frac{\partial}{\partial w}(f_{l+2}w^{l+3})\right]. \quad (13)$$

This agrees with the z^l element of $D_{l(z)}^{wb}$ in Eq. (11).

SUMMARY

The general intrinsic velocity (w) spherical harmonic equation has been derived [Eq. (12)] with the extrinsic case as a particular case [Eq. (8)] obtained when the reference velocity $\mathbf{C}$ and its derivatives are zero so that $\mathbf{v}$ and $\mathbf{w}$ are equal. Legendre polynomial recursion relations have been

used to check the result. The worth of this general result can best be appreciated by those who have laboured through the piecemeal derivation of several particular cases.

With this general expression for $\mathbf{D}_l$ available, it may now be worthwhile to examine more general expressions for collisions in order to extend Shkarofsky's[12] work on the Fokker–Planck terms on effects on irreducible anisotropic tensor pressure to other effects of higher order.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the helpfulness of several authors (P. R. Wallace, E. Ikenberry, T. Kahan) in sending me copies of their work and of Dr. Suchy and Dr. Ziering in drawing my attention to the works of A. Sommerfeld and E. Ikenberry, respectively. My thanks are due to Dr. I. P. Shkarofsky, who urged the extension of the work to obtain the intrinsic results and who was always available for invaluable and stimulating discussion.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 8 AUGUST 1966

Exact Solutions for a Semi-Infinite Square Lattice Gas

N. KARAYIANIS, C. A. MORRISON, AND D. E. WORTMAN
Harry Diamond Laboratories, Washington, D. C.
(Received 22 December 1965)

The problem of a two-dimensional lattice gas with nearest neighbor infinite repulsion is considered by obtaining exact solutions for a sequence of semi-infinite spaces. The exact solutions are obtained for $M \times \infty$ spaces with $2 \leq M \leq 14$ in even steps, and although there are no phase transitions in these spaces, a criterion for the point of "closest approach" to a phase transition is established. The values of the thermodynamic variables evaluated at this point for each M are extrapolated to obtain the properties of the two-dimensionally infinite space. The data indicate that a phase transition occurs with possibly infinite compressibility at an activity $z = 3.799$, a density $\rho/\rho_{\max} = 0.7356$, and a pressure given by $P/k_BT = 0.7914$. The density is obtained by a rigorous differentiation of the secular determinant that determines the value of the pressure for a given z, thus securing the accuracy of the calculations and enabling the extrapolated values of the thermodynamic variables to be estimated with good precision.

I. INTRODUCTION

IN this paper we are concerned with the thermodynamic properties of a lattice gas of hard squares with infinite nearest neighbor repulsion. This model is of interest in its relationship to certain physical systems, which are observed to undergo phase transitions, and in which the particle interaction is predominantly repulsive.[1] For the details concerning the history of this problem, the authors refer the reader to a recent paper by Gaunt and Fisher.[2] These authors have analyzed the two-dimensional, infinite system by various approximation techniques in order to investigate the possible existence and the nature of a phase transition. They predict a "continuous" phase transition at an activity of $z = 3.80 \pm 2$, a density of $\rho/\rho_{\max} = 0.740 \pm 8$, and a pressure given by $P/k_BT = 0.792 \pm 5$. Further, they suggest that the compressibility of the infinite system is finite at the point of phase transition.

In this paper, we report on an analysis of this problem using a different approach. Exact solutions for a sequence of semi-infinite two-dimensional systems are obtained, and values for the above thermodynamic quantities for the two-dimensional, infinite system are determined by extrapolation. The results indicate a third-order phase transition at an activity of $z = 3.799$, a density of $\rho/\rho_{\max} = 0.7356$ and a pressure given by $P/k_BT = 0.7914$. Further, it is shown that an infinite compressibility at the point of phase transition is at least consistent with the present calculations and cannot be excluded.

In Sec. II of this paper, the method for obtaining the exact solutions for the semi-infinite systems is discussed. The relationship of these solutions to the Yang and Lee[3] criterion for a phase transition is

[1] G. E. Uhlenbeck, *Statistical Physics* (W. A. Benjamin, Inc., New York, 1963), p. 47.
[2] D. S. Gaunt and M. E. Fisher, J. Chem. Phys. 43, 2840 (1965).
[3] C. N. Yang and T. D. Lee, Phys. Rev. 87, 404 (1952).

used to check the result. The worth of this general result can best be appreciated by those who have laboured through the piecemeal derivation of several particular cases.

With this general expression for $\mathbf{D}_l$ available, it may now be worthwhile to examine more general expressions for collisions in order to extend Shkarofsky's[12] work on the Fokker–Planck terms on effects on irreducible anisotropic tensor pressure to other effects of higher order.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the helpfulness of several authors (P. R. Wallace, E. Ikenberry, T. Kahan) in sending me copies of their work and of Dr. Suchy and Dr. Ziering in drawing my attention to the works of A. Sommerfeld and E. Ikenberry, respectively. My thanks are due to Dr. I. P. Shkarofsky, who urged the extension of the work to obtain the intrinsic results and who was always available for invaluable and stimulating discussion.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 8 AUGUST 1966

Exact Solutions for a Semi-Infinite Square Lattice Gas

N. KARAYIANIS, C. A. MORRISON, AND D. E. WORTMAN
Harry Diamond Laboratories, Washington, D. C.
(Received 22 December 1965)

The problem of a two-dimensional lattice gas with nearest neighbor infinite repulsion is considered by obtaining exact solutions for a sequence of semi-infinite spaces. The exact solutions are obtained for $M \times \infty$ spaces with $2 \leq M \leq 14$ in even steps, and although there are no phase transitions in these spaces, a criterion for the point of "closest approach" to a phase transition is established. The values of the thermodynamic variables evaluated at this point for each M are extrapolated to obtain the properties of the two-dimensionally infinite space. The data indicate that a phase transition occurs with possibly infinite compressibility at an activity $z = 3.799$, a density $\rho/\rho_{\max} = 0.7356$, and a pressure given by $P/k_BT = 0.7914$. The density is obtained by a rigorous differentiation of the secular determinant that determines the value of the pressure for a given z, thus securing the accuracy of the calculations and enabling the extrapolated values of the thermodynamic variables to be estimated with good precision.

I. INTRODUCTION

IN this paper we are concerned with the thermodynamic properties of a lattice gas of hard squares with infinite nearest neighbor repulsion. This model is of interest in its relationship to certain physical systems, which are observed to undergo phase transitions, and in which the particle interaction is predominantly repulsive.[1] For the details concerning the history of this problem, the authors refer the reader to a recent paper by Gaunt and Fisher.[2] These authors have analyzed the two-dimensional, infinite system by various approximation techniques in order to investigate the possible existence and the nature of a phase transition. They predict a "continuous" phase transition at an activity of $z = 3.80 \pm 2$, a density of $\rho/\rho_{\max} = 0.740 \pm 8$, and a pressure given by $P/k_BT = 0.792 \pm 5$. Further, they suggest that the compressibility of the infinite system is finite at the point of phase transition.

In this paper, we report on an analysis of this problem using a different approach. Exact solutions for a sequence of semi-infinite two-dimensional systems are obtained, and values for the above thermodynamic quantities for the two-dimensional, infinite system are determined by extrapolation. The results indicate a third-order phase transition at an activity of $z = 3.799$, a density of $\rho/\rho_{\max} = 0.7356$ and a pressure given by $P/k_BT = 0.7914$. Further, it is shown that an infinite compressibility at the point of phase transition is at least consistent with the present calculations and cannot be excluded.

In Sec. II of this paper, the method for obtaining the exact solutions for the semi-infinite systems is discussed. The relationship of these solutions to the Yang and Lee[3] criterion for a phase transition is

[1] G. E. Uhlenbeck, *Statistical Physics* (W. A. Benjamin, Inc., New York, 1963), p. 47.
[2] D. S. Gaunt and M. E. Fisher, J. Chem. Phys. 43, 2840 (1965).
[3] C. N. Yang and T. D. Lee, Phys. Rev. 87, 404 (1952).

pointed out in Sec. III. Further, it is proved that no phase transitions occur in an $M \times N$ space in the limit $N \to \infty$ for any finite M; however, it is possible to establish a criterion for determining the point of "closest approach" to a phase transition. The asymptotic extrapolations to infinite M of the thermodynamic variables determined at these points provide the limit values given above.

The results of the computer calculations are given in Sec. IV. Programs were written for a 7094 computer to solve the difference equations involved and obtain the matrices that embody the equations of state for the various systems. The calculations were carried out for $M \times \infty$ systems with periodic boundary conditions for $2 \leq M \leq 14$ in even steps, and for $M \times \infty$ systems with free boundary conditions for $M = 3, 4, 6, 8$. Fortunately, the periodic boundary systems were found to have properties that converge rapidly with increasing M; hence it was not considered imperative, in view of the rapidly increasing computer time required, to go to larger systems.

II. THE $M \times N$ SPACE

Consider an $M \times N$ square-lattice space occupied by indistinguishable particles that experience an infinite repulsion when in nearest neighbor sites and zero force otherwise. Let M be even and fixed, and let us seek the grand partition function for this system in the limit $N \to \infty$. We proceed by classifying the various states of the system that contain n particles by the configuration of the first column of M sites, i.e., by the number of particles k in the first column and their particular arrangement ν_k. If n_k is the number of possible configurations for k particles, then $1 \leq \nu_k \leq n_k$, and the partition function $P_{MN}(n)$ is given by the summation over the various possibilities,

$$P_{MN}(n) = \sum_{k=0}^{\frac{1}{2}M} \sum_{\nu_k=1}^{n_k} p_{MN}(\nu_k k n). \tag{1}$$

Here $p_{MN}(\nu_k k n)$ is the number of states with n particles where, to repeat, k of these particles are in the first column of M sites and are arranged in the particular way characterized by ν_k. In terms of the "partial" partitions p_{MN}, it is clear that the partition function $P_{MN}(n)$ is also given by

$$P_{MN}(n) = p_{MN+1}(1\ 0\ n), \tag{2}$$

i.e., the number of states in the $M \times (N+1)$ space containing n particles, none of which are in the first column of M sites, is simply the partition function

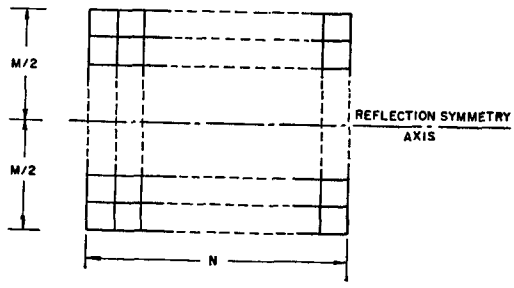


FIG. 1. The $M \times N$ space showing the reflection symmetry axis.

for the $M \times N$ space. The grand partition function Q_{MN} is therefore

$$Q_{MN} \equiv Q_{MN}(1\ 0) = \sum_n p_{MN+1}(1\ 0\ n) z^n. \tag{3}$$

Depending on the type of boundary conditions, free or periodic, there will be certain p_{MN} for a given k that will be equal, thus enabling the summation in (1) to be collapsed somewhat. To illustrate this further, we restrict the following discussion to the free-boundary case which is simpler to discuss but is otherwise similar to the periodic-boundary case. The free-boundary system is invariant under reflection through an axis parallel to the "N" direction as shown in Fig. 1. Therefore, pairs of configurations that go into each other under reflection are equivalent and can be represented by a single p_{MN}. Expression (1) thus becomes

$$p_{MN+1}(1\ 0\ n) = \sum_{k=0}^{\frac{1}{2}M} \sum_{\nu_k=1}^{n_k} g_M(\nu_k k; 1\ 0) p_{MN}(\nu_k k n), \tag{4}$$

where use has been made of (2). The coefficients g_M in (4) are equal to one or two, according to whether the configuration ν_k goes into itself under reflection or not. We assume that the p_{MN} have been renumbered in (4) so that n_k is less than it was originally assumed to be in (1). Specifically, it can be shown that n_k here is given by

$$n_k = \frac{1}{2}\left[\binom{M-k+1}{k} + \binom{\frac{1}{2}M - \frac{1}{2}k}{\frac{1}{2}k}\right], \quad k \text{ even}, \tag{5}$$

$$= \frac{1}{2}\binom{M-k+1}{k}, \quad k \text{ odd}, \tag{6}$$

where $\binom{a}{b}$ is a binomial coefficient.

In a manner similar to the determination of (4), all of the partial partitions for the $M \times N + 1$ space can be expressed in terms of the partial

partitions for the $M \times N$ space giving relationships of the form

$$p_{MN+1}(\nu_k' k' n) = \sum_{k\nu_k} g_M(\nu_k k; \nu_k' k') p_{MN}(\nu_k k n - k'). \quad (7)$$

These equations are obtained by considering the number of ways various configurations in the $M \times N$ space can exist contiguous to the specific configuration ν_k' in the first column of the $M \times (N+1)$ space, where no two particles are allowed in nearest neighbor sites. The elements of the matrix g_M are zero, one, or two, and there is a symmetry between its elements of the form

$$g_M(\nu_k' k'; \nu_k k) = 2^{\mu(\nu_k') - \mu(\nu_k)} g_M(\nu_k k; \nu_k' k'), \quad (8)$$

where $\mu(\nu_k)$ is defined to be zero if the configuration ν_k is symmetric under reflection, and unity otherwise.

Let a set of "partial" grand partition functions be defined by

$$Q_{MN}(\nu_k k) = 2^{\frac{1}{2}\mu(\nu_k)} z^{-\frac{1}{2}k} \sum_n p_{MN+1}(\nu_k k n) z^n. \quad (9)$$

This definition is consistent with (3) owing to the fact that the configuration with $k = 0$ is symmetric under reflection so that $\mu(k = 0) = 0$. In terms of these new variables, (7) becomes

$$Q_{MN}(\nu_k' k') = \sum_{k\nu_k} z^{\frac{1}{2}(k+k')} h_M(\nu_k k; \nu_k' k') Q_{MN-1}(\nu_k k), \quad (10)$$

where

$$h_M(\nu_k k; \nu_k' k') = 2^{\frac{1}{2}[\mu(\nu_k') - \mu(\nu_k)]} g_M(\nu_k k; \nu_k' k') \quad (11)$$

$$= h_M(\nu_k' k'; \nu_k k). \quad (12)$$

The set of equations represented by (10) is solved by assuming the N dependence of the $Q_{MN}(\nu_k k)$ to be of the form

$$Q_{MN}(\nu_k k) \sim \lambda^N Q_M(\nu_k k), \quad (13)$$

in which case (expressing the Q_M as a column vector $\mathbf{Q}_M$),

$$\lambda(z)\mathbf{Q}_M = \mathbf{H}_M(z)\mathbf{Q}_M \quad (14)$$

with solutions for $\lambda(z)$ given by

$$\det [\mathbf{H}_M(z) - \lambda(z)] = 0. \quad (15)$$

For positive z, it is significant to note that the matrix $\mathbf{H}_M(z)$ with elements $z^{\frac{1}{2}(k+k')} h_M(\nu_k k; \nu_k' k')$ is real and symmetric; hence, it is Hermitian and has real eigenvalues $\lambda_i(z)$.

The grand partition function is given in terms of the $\lambda_i(z)$ by

$$Q_{MN} = \sum_i a_{M_i}(z) \lambda_i^N(z), \quad (16)$$

where the $a_{M_i}(z)$ are independent of N and, in practice, are determined by the boundary conditions for Q_{MN}. The dimensionality of $\mathbf{H}_M(z)$, hence the number of $\lambda_i(z)$ in the above summation, is given by

$$d_M = \sum_{k=0}^{\frac{1}{2}M} n_k \quad (17)$$

with n_k given by (5) and (6).

With the form of the solution for Q_{MN} given by (16), we have succeeded in isolating the dependence on N, thus enabling the thermodynamic functions for the system to be determined in the limit $N \to \infty$. For this two-dimensional, semi-infinite system, we define

$$P/k_B T \equiv \Gamma(M) = \lim_{N\to\infty} (MN)^{-1} \ln Q_{MN}(z), \quad (18)$$

$$\rho(M) = z\, d\Gamma(M)/dz, \quad (19)$$

and

$$K(M) = [\rho(M)]^{-1} d\rho(M)/d\Gamma(M), \quad (20)$$

where P, k_B, and T are the pressure, Boltzmann constant, and temperature, respectively, and $\rho(M)$ and $K(M)$ the specific density and compressibility for the $M \times \infty$ system.

III. THE LIMIT $N \to \infty$

In order to examine the grand partition function in the limit $N \to \infty$, it is necessary to consider first the relative magnitude of the eigenvalues $\lambda_i(z)$ as functions of z. For real, positive z, the matrix $\mathbf{H}_M(z)$ is nonnegative; i.e., each of its elements is greater than or equal to zero, but no zeros occur in the first row or column [see the discussion after (4)]. This implies that each element of $\mathbf{H}_M^2(z)$ is greater than zero, hence $\mathbf{H}_M(z)$ is primitive.[4] The largest eigenvalue of a primitive matrix is positive, simple and greater in magnitude than each of the other eigenvalues. Letting this largest eigenvalue be $\lambda_1(z)$, we then conclude

$$\lambda_1(z) > |\lambda_i(z)|, \qquad i \neq 1, \quad (21)$$

for all real, positive z.

For arbitrary complex z, the theorem obviously does not apply. It is useful, however, to extend the concept of a largest eigenvalue to complex z.

[4] F. R. Gantmacher, *Applications of the Theory of Matrices* (Interscience Publishers, Inc., New York, 1959), p. 96.

Accordingly, let us define $\mathfrak{R}(z)$ as the set of complex z for which $|\lambda_1(z)| > |\lambda_i(z)|$ for all $i \neq 1$, i.e.,

$$\mathfrak{R}(z) = \{z \mid |\lambda_1(z)| > |\lambda_i(z)|, \text{ for all } i \neq 1\}. \quad (22)$$

Then, it follows simply that

$$\lim_{N\to\infty} (Q_{MN})^{1/N} \to \lambda_1(z), \qquad z \in \mathfrak{R}(z). \quad (23)$$

Furthermore, it follows that the grand partition function cannot vanish for $z \in \mathfrak{R}(z)$, for $\lambda_1(z) = 0$ is incompatible with the definition of $\mathfrak{R}(z)$ given by (22). Since $\mathfrak{R}(z)$ includes the real, positive z axis, we conclude on the basis of the Yang and Lee criterion that no phase transitions occur for the $M \times \infty$ two-dimensional systems for finite M.

Zeros for the grand partition function in the limit $N \to \infty$ are therefore necessarily found for z in the set complementary to $\mathfrak{R}(z)$. This set may be defined by

$$\mathfrak{N}(z) = \{z \mid |\lambda_1(z)| = |\lambda_2(z)| \geq |\lambda_i(z)|\}. \quad (24)$$

To complete the reasoning, it is clear that no real, positive z is included in $\mathfrak{N}(z)$, for this would indicate a nonsimple largest eigenvalue, in contradiction to the proved primitivity of $\mathbf{H}_M(z)$. For physical z values, therefore, (18) and (19) become

$$\Gamma(M) = (M)^{-1} \ln \lambda_1, \quad (25)$$

$$\rho(M) = z(M\lambda_1)^{-1}\, d\lambda_1/dz. \quad (26)$$

While there are no phase transitions for finite M, expression (24) suggests a method for seeking the values of real, positive z for which the $M \times \infty$ system "most closely" approaches a phase transition. That is, one may examine the function $\lambda_2(z)/\lambda_1(z)$ and seek to maximize it as a function of z on the real axis. A maximum, if it exists, is achieved for $z = z_t(M)$, where $z_t(M)$ is a solution

$$\lambda_1\lambda_2' = \lambda_2\lambda_1', \quad (27)$$

the prime representing differentiation with respect to z. In the computer calculations, the value of $z_t(M)$ was found for each system by homing in on the zero of the function $\eta = 1 - (\lambda_1\lambda_2')/(\lambda_2\lambda_1')$. The thermodynamic properties of the systems at these values of z plotted versus $1/M$ provided the curves from which the properties of the infinite system were inferred.

IV. RESULTS

In this section we present the results of the computer calculations only for the systems with periodic boundary conditions. There are two reasons for this. First, the properties of the periodic systems converge much more rapidly as M increases, and second, the matrices are much smaller than that of the corresponding free boundary system owing to the greater symmetry of the periodic systems.

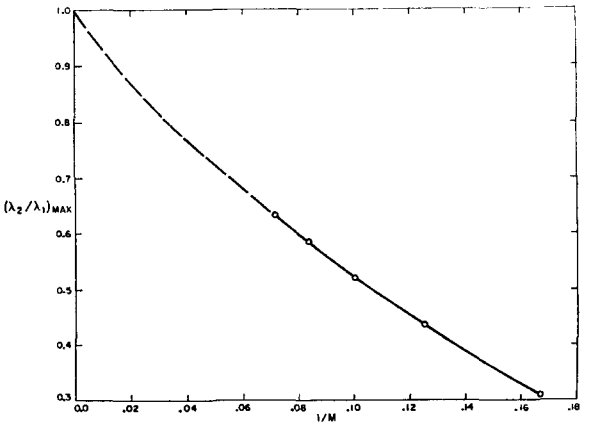


FIG. 2. The maximum ratio of the two largest roots for the spaces $6 \leq M \leq 14$.

To facilitate the following discussion, let the value of the thermodynamic variables at the point determined by the maximum of λ_2/λ_1 for each M be represented by $z_t(M)$, $\rho_t(M)$, $\Gamma_t(M)$, and $K_t(M)$. In Fig. 2 there is a plot of $(\lambda_2/\lambda_1)_{max}$ vs $1/M$ for $6 \leq M \leq 14$ in even steps. As M becomes infinite, the ratio appears to approach unity. According to the discussion in the previous section, it is clear that the degree of certainty of a phase transition occurring in the infinite system hinges primarily on the degree of certainty that the limiting value of this ratio is indeed unity.

In Fig. 3, the variables $z_t(M)$, $\rho_t(M)$, and $1/\Gamma_t(M)$ are similarly plotted vs $1/M$. A smooth extrapolation to the infinite system gives the following limiting values: $z_t(\infty) = 3.799$, $\rho_t(\infty) = 0.3678$, and $\Gamma_t(\infty) = 0.7914$. These values compare very favorably with the values of Gaunt and Fisher, i.e., $z = 3.80$, $\rho = 0.370$, and $\Gamma = 0.792$, but do not agree quite as well with the values of Runnels,[5] i.e., $z = 3.86$, $\rho = 0.369$, and $\Gamma = 0.796$.

A cross check on the value for $\rho_t(\infty)$ can be obtained by plotting $\rho_t(M)$ vs $1 - (\lambda_2/\lambda_1)^{1/M}_{max}$ as shown in Fig. 4. Assuming that $(\lambda_2/\lambda_1)_{max}$ does approach unity in the limit $M \to \infty$, a consistency check for $\rho_t(\infty)$ is provided. The value of $\rho_t(\infty)$ from this plot is again similar to 0.3678 which supports the value obtained above. Consistency between the extrapolated values of $\rho_t(M)$ and $\Gamma_t(M)$ as compared with the extrapolated value of $z_t(M)$ can be examined by plotting $\rho(M)$ and $\Gamma(M)$ cal-

[5] L. K. Runnels, Phys. Rev. Letters 15, 581 (1965).

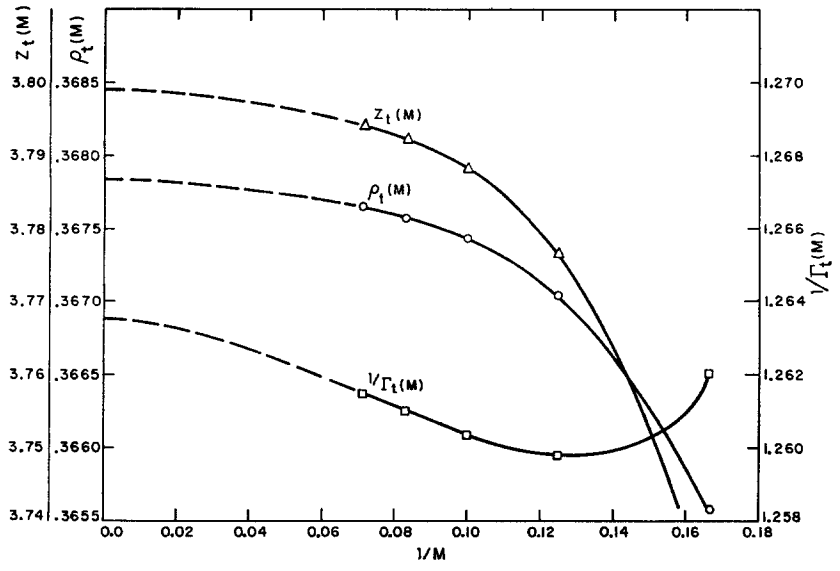


FIG. 3. The calculated values of z, ρ, and Γ evaluated at $(\lambda_2/\lambda_1)_{max}$ for the spaces $6 \leq M \leq 14$.

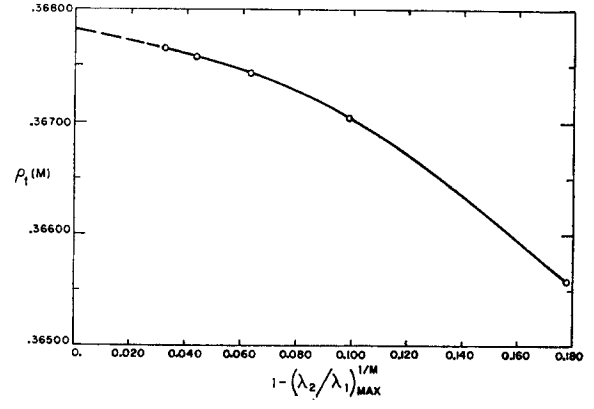


FIG. 4. The calculated values of $\rho_t(M)$ displayed as functions of $1 - (\lambda_2/\lambda_1)^{1/M}_{max}$ for the spaces $6 \leq M \leq 14$.

culated at fixed z vs $1/M$, as shown in Fig. 5. The fixed values of z are chosen to straddle $z_t(\infty)$ as follows: $z = 3.798, 3.799$, and 3.800. The extrapolated values for ρ and Γ evaluated at $z = z_t(\infty) = 3.799$ are $\rho = 0.3679$ and $\Gamma = 0.7913$, which give further support to the previous extrapolations.

The compressibility for the $M \times \infty$ systems is plotted vs the density in Fig. 6 to show the tendency towards a peak (possibly infinite) at the transition density. The vertical dash marks on the curves indicate the values $\rho_t(M)$, i.e., the values of ρ where $(\lambda_2/\lambda_1)_{max}$ obtains for each system. The rapid convergence of these points to $\rho_t(\infty)$ can be compared with the very slow relative convergence of the peaks (in the larger systems where they occur) or of the

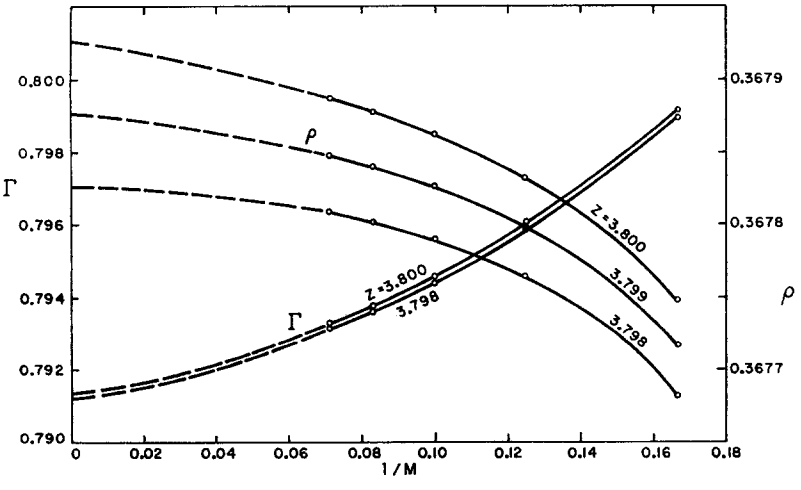


FIG. 5. The calculated values of ρ and Γ for constant z for the spaces $6 \leq M \leq 14$.

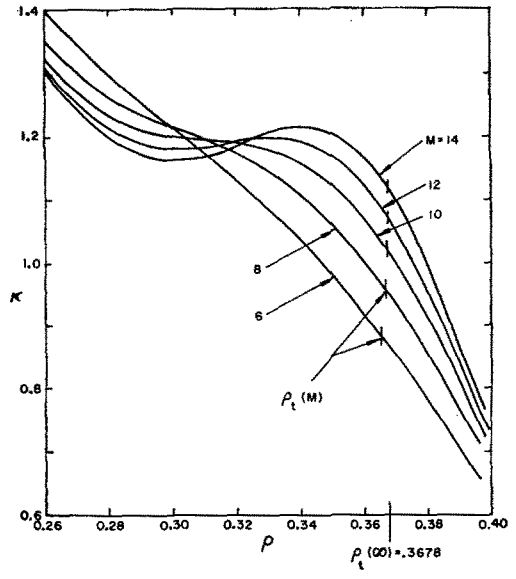


FIG. 6. The compressibility K vs ρ for the various spaces, indicating the tendency towards a peak (possibly infinite) at $\rho_t(\infty)$.

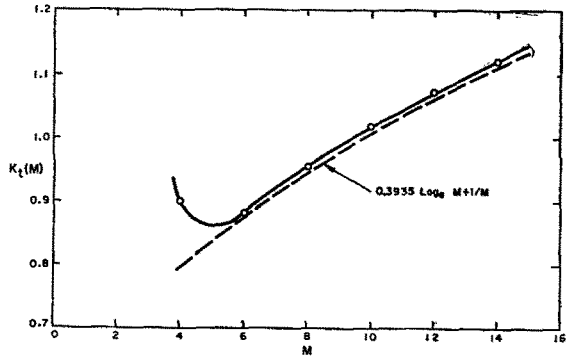


FIG. 7. The calculated values of the compressibility evaluated at $(\lambda_2/\lambda_1)_{max}$ for the spaces $4 \leq M \leq 14$. The dashed curve is plotted solely for comparison of the compressibility curves with a curve that diverges logarithmically with M.

inflection points in the $K(M)$ vs ρ curves. This fact explains the slow convergence of the results obtained by Runnels,[5] who used the maximum of the quantity $\rho^2 K$ as his criterion for closest approach to a phase transition rather than $(\lambda_2/\lambda_1)_{max}$.

The limit of $K_t(M)$ cannot be determined with certainty. In Fig. 7 it is plotted vs. M together with the curve $A \ln M + 1/M$, where A is a constant chosen appropriately to provide a means of comparing the compressibility curve with a curve that diverges logarithmically as $M \to \infty$. The curves seem to follow a parallel path, indicating that a logarithmic divergence of $K_t(M)$ as $M \to \infty$ is a possibility as suggested by Ree.[6]

In Fig. 8 we present a plot of Γ vs ρ for the $M = 6, 10, 14$ systems. Except in the neighborhood of $\rho_t(\infty)$, the curve for the infinite system should be very much similar to the $14 \times \infty$ system. An infinite compressibility at $\rho_t(\infty)$ would require $d\Gamma/d\rho$ to vanish at that point for the infinite system, and, as stated above, this possibility cannot be excluded by our calculations.

[6] F. H. Ree (private communication).

With regards to the calculations, we wish to note that the only quantity that was computed by finite differences is the compressibility. The density for a given z was calculated exactly (to eight places) by computing the derivative of the secular determinant. The two largest roots were found by searching for the zeros of the secular determinant for given z. This process obviated diagonalizing the entire matrix. Further accuracy was assured by using the sensitive function η (see the end of Sec. III) to determine $z_t(M)$ and the associated thermodynamic quantities.

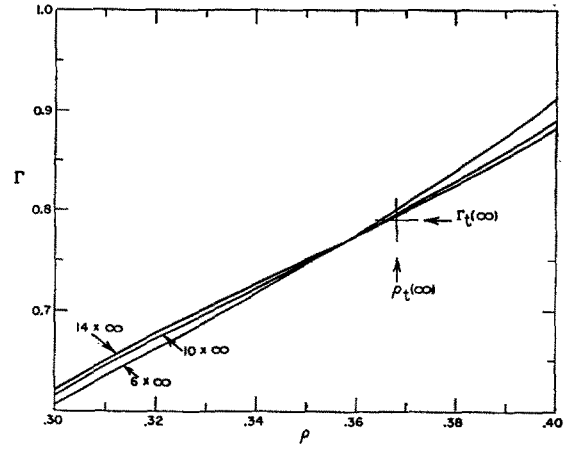


FIG. 8. The Γ vs ρ curve for the $14 \times \infty$ space compared with similar curves for smaller spaces in the vicinity of the transition point.

Extraction of Singularities from the S Matrix*

P. V. LANDSHOFF AND D. I. OLIVE
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England
(Received 13 July 1965)

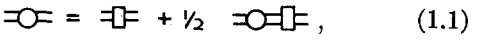
Two methods are described for extracting triangle singularities from matrix elements. The first is comparatively simple but involves the use of off-mass-shell amplitudes; the second is rather involved. A proof of the Cutkosky discontinuity formula is given independent of perturbation theory, and it is shown that the Riemann-sheet properties of the singularity in the physical region agree with perturbation theory. The connection between this and a causality requirement is discussed. The relevance of the work to practical computations is explained.

1. INTRODUCTION

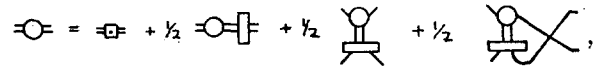
Practical Considerations

THERE are at least two practical reasons for attempting to extract singularities from matrix elements. The first is that knowledge of the singularity structure is necessary if one is to explore the unphysical sheets close to the physical region, which directly influence the physical amplitude in that they contain resonance poles and other singularities of dynamical origin. The second is that the extraction can provide a means of making approximate calculations of the amplitude.

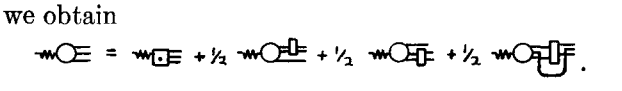
One way of performing such calculations is to use dispersion relations. If one has reason to believe that a certain singularity dominates in some part of the physical region, the discontinuity corresponding to that singularity is inserted into the dispersion relation and other discontinuities are neglected.[1] Another way, which does not involve dispersion integrals, is to extract from the amplitude not the discontinuity corresponding to a given singularity, but the complete structure of that singularity. For the case of two-particle normal threshold singularity, this extraction has been performed by Zimmermann[2] who shows that, if one defines an amplitude  in terms of the complete two-particle → two-particle scattering amplitude by means of the equation


(1.1)

then  will be free of the two-particle normal threshold singularity in the direct channel. If one makes some sort of guess as to its structure and inserts this in (1.1), one obtains an integral equation for the scattering amplitude whose solution will satisfy two-particle unitarity in the direct channel. For example, choosing  to be a constant yields the effective-range formula. Of course, a major weakness of (1.1) is that it contains no information from crossing. To overcome this, notice that, since the second term on the right-hand side of (1.1) contains the complete two-particle singularity structure in the direct channel, analytic continuations of it yield the corresponding structures in the two crossed channels. So, if we define a new amplitude  by


(1.2)

it is free of two-particle singularities in all three channels. One may now perform a crossing-symmetric model calculation[3] by choosing some simple form for it, such as a constant, so that (1.1) and (1.2) are two simultaneous integral equations for  and  . If we continue (1.2) analytically so that the left-hand side becomes the crossed amplitude  and, in order that this have physical meaning, replace the particle in the initial state by one whose mass is greater than the sum of the masses of the three particles in the final state, we obtain


(1.3)

This equation defines, in terms of the decay am-

* The research reported here has been sponsored in part by the Air Force Office of Scientific Research, OAR, under Grant AF EOAR 63-79 with the European Office of Aerospace Research, United States Air Force.

[1] The simple triangle singularity has received particular attention in this way because it is one of the few singularities that is actually an infinity in the amplitude and so may well be expected to dominate in the parts of the physical region close to it (if any). See P. V. Landshoff and S. B. Treiman, Phys. Rev. 127, 649 (1962).

[2] W. Zimmermann, Nuovo Cimento 21, 249 (1961).

[3] This calculation is, of course, not easy. It is being studied.

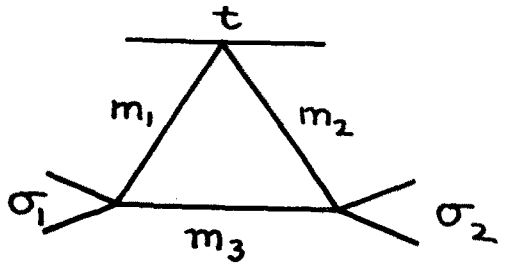
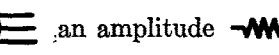
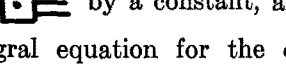


FIG. 1. The Landau–Cutkosky diagram under study.

plitude  an amplitude that is free of the two-particle thresholds in the three subenergies of pairs of particles in the final state. If one supposes that the decay is determined by the final state interactions alone, it is reasonable to approximate  by a constant, and so one obtains an integral equation for the decay amplitude.[4]

It should be stressed that each line in (1.1), (1.2), and (1.3) is on the mass shell and that the internal lines represent δ functions rather than propagators.[5] This has the effect that, unlike in dispersion theory, all the integrations are finite. For example, in (1.1) the integration implied in the last term is just like that encountered in two-particle unitarity: it is over physical angles at fixed energy. Thus, guesses as to high-energy behavior do not enter the calculations.

To obtain more realistic equations, it is evidently necessary to exhibit explicitly the structure of further singularities before putting the residual amplitude equal to a constant. That is partly the motivation of this paper, where we confine our attention to the triangle singularities[6] and, in particular, to the one corresponding to the Landau diagram appearing in Fig. 1, because this is the simplest nondegenerate triangle singularity occurring in a physical region.

Theoretical Considerations

It is believed that the singularities of matrix elements, or at least those singularities close to the physical regions, are very similar to those of finite-order perturbation theory.[7] The features of perturbation theory which suggest this are the hierarchical properties[8] (discussed in this section) which, roughly speaking, means that the behavior of the singularities is independent of the order of the diagram considered, and the demonstration by Cutkosky[9] that the perturbation-theory sum of the discontinuities across the cuts attached to Landau singularities can be expressed in a form independent of perturbation theory. Up to now the situation in S-matrix theories has been less clear. Polkinghorne[10] has shown that the singularities generated iteratively by unitarity and crossing must lie upon the Landau curves, but he was unable to show that the same parts of the Landau curves were singular, as in perturbation theory, without assuming that in certain limited regions the amplitudes enjoyed analytic properties like those known in perturbation theory. The idea, then, was to derive, for example, double dispersion relations for two-particle to two-particle amplitudes from single variable ones. Nowadays the object is to assume a much weaker form of analyticity assumption and derive the single-variable dispersion relations as well. Olive[11] has extended this idea of iteration of singularities to show that, if some sort of hierarchical property is true and if it is possible to determine which parts of the Landau curves are singular, then it is possible to build up an S-matrix theory from analyticity and physical unitarity postulates and prove the fundamental theorems of quantum field theory—crossing, TCP, Hermitian analyticity, etc.

In this paper, we show how the physical unitarity equations can indeed control the singularity structure in an unambiguous way, and guarantee the hierarchical property—at least in the physical region.

We should look at the physical region, because,

[4] This model has recently been investigated by M. Taha [Nuovo Cimento **42**, 201 (1966)].

[5] An equation rather like (1.2), but with the internal lines representing propagators, has been given by J. G. Taylor [Nuovo Cimento Suppl. **1**, 857 (1963)]. The analog of (1.1), when the internal lines represent propagators, is just the Bethe–Salpeter equation. Equation (1.3) contains similar information to the Khuri–Treiman dispersion relations [see I. J. R. Aitchison, Phys. Rev. **137**, B1070 (1965)] but, as is demonstrated in Ref. 4, is much easier to solve.

[6] The structure of the multiparticle normal thresholds is extracted in a recent paper by D. Branson, Ann. Phys. (N. Y.) **35**, 351 (1965).

[7] It is known that they cannot be exactly the same everywhere. For example, finite-order perturbation theory produces stable-particle poles and normal thresholds in the same positions on all Riemann sheets, which is forbidden by unitarity. See P. V. Landshoff, Nuovo Cimento **28**, 123 (1963); D. I. Olive, *ibid.* **28**, 1318 (1963).

[8] P. V. Landshoff, J. C. Polkinghorne, and J. C. Taylor, Nuovo Cimento **19**, 939 (1961). Another statement of the hierarchical principle that has been given is that the discontinuity associated with a given singularity contains none of the singularities corresponding to the reduced diagrams obtained by contracting out δ-function lines. As we show in studying integral (3.17), this is, in general, false. We hope to discuss this in subsequent work.

[9] R. E. Cutkosky, J. Math. Phys. **1**, 429 (1960); Rev. Mod. Phys. **33**, 448 (1961).

[10] J. C. Polkinghorne, Nuovo Cimento **23**, 360 (1962); **25**, 901 (1962). See also H. P. Stapp, Phys. Rev. **125**, 2139 (1962).

[11] D. I. Olive, Phys. Rev. **135**, B745 (1964).

naturally enough, this seems to be particularly favored, both in S-matrix theory—since it is there that the fundamental equations, the physical unitarity equations, operate—and in perturbation theory—since there the Feynman integrals have undistorted integration contours and are particularly easy to analyze. Landau[12] has shown that the singularities of a Feynman integral must lie on the "Landau curves" defined by the implicit equations,

$$\text{either } \alpha_i = 0 \text{ or } q_i^2 = m_i^2 \quad \text{for each internal line } i,$$
$$\sum_{i \in l} \alpha_i q_i = 0 \qquad \text{for each loop } l. \tag{1.4}$$

In the second equation it is understood that the internal momentum q_i of the ith internal line is measured in the sense of the loop. In the physical region, it is possible to augment these conditions to find necessary and sufficient conditions for singularities. As we show in the Appendix, it is a consequence of Feynman's rule that $-i\epsilon$ be attached to each internal mass, that we must have for singularity

$$\alpha_i \geq 0 \text{ for each internal line } i, \text{ each } q_i \text{ real.} \tag{1.5}$$

We now show that the physical-region hierarchical structure is a direct consequence of the singularity criteria (1.4) and (1.5). According to condition (1.5), the only points at which a Landau curve can cease being singular is a point where one (or more) of the α's vanish. When this happens, it can be shown[8] that the curve touches the lower-order curve obtained by contracting out the line with zero α. Thus, in a Feynman integral, the singularity of the Landau curve depends in no way upon the possible presence of higher-order curves, that is, those singularities corresponding to more internal lines being on the mass shell, but is affected only by points of tangency (sometimes called "effective intersections"[8]) with lower-order curves. Also, because of the conditions (1.4) and (1.5), the singular behavior of the Landau curve is the same, whatever Feynman integral it appears in, and we would therefore expect conditions (1.4) and (1.5) to be preserved in the perturbation sum, and hence apply to the complete amplitude. The properties just described constitute the physical region "hierarchical structure." In general, these properties may not be true outside the physical region and an important problem is to consider what modifications of the statement must be made for it to be valid generally, but we offer no discussion of this question.

The belief that conditions (1.4) and (1.5), and hence the hierarchical property, are applicable to the complete amplitude is further enhanced by a particularly beautiful physical interpretation proposed by Coleman and Norton,[13] which we now describe.

At any point of a physical-region Landau curve, the internal momenta of the corresponding diagram are real, on the mass shell, and have a definite sense. It follows that the diagram is "physical-looking" in that it looks like a physically realizable succession of scattering processes (and involves no decay amplitudes). Coleman and Norton's point is that it is attractive to interpret the quantity $x_i = \alpha_i q_i$ as a measure of the space–time traversed by the ith intermediate particle between the interactions. The second of the conditions (1.4) becomes $\sum x_i = 0$ and now means that rescattering only occurs if there is an actual space–time coincidence. The condition $\alpha_i > 0$ means that the particles with positive energy move forward in time. Thus, with this interpretation, the physical-region singularities occur only for values of the external momenta that allow a succession of intermediate point interactions with the particles participating in the interactions having physical momenta and physical (that is, positive) intermediate flight times. (The need to mention point interactions and microscopic times can be eliminated by saying instead that the singularity occurs when the over-all reaction can occupy a large volume of space–time, large compared with the range of the primitive forces, because of the possibility of there being physical intermediate particles in free flight.)

In a mass-shell S-matrix theory, one would imagine that information can be transmitted over large intervals of space and time only by real mass-shell particles. By *causality* one normally means that a signal cannot be received before it is transmitted. If all signals are to be conveyed by particles, as we have agreed, then this is equivalent to saying that only particles moving forward in time are observable. To the extent that we can say that the existence of singularities on physical-region Landau curves reveals the possibility of intermediate physical scattering processes, the $+\alpha$ condition for singularity is equivalent to saying that only causal processes occur. In other words, we

[12] For a review of the singularities of perturbation theory, see the lectures by R. J. Eden and J. C. Polkinghorne, *1961 Brandeis Summer School Lectures* (W. A. Benjamin, Inc., New York, 1961). A more up-to-date account, which also deals with the singularity structure outside the framework of perturbation theory, is given by R. J. Eden, P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne [*The Analytic S-Matrix* (Cambridge University Press, Cambridge, England, 1966)].

[13] S. Coleman and R. Norton, Nuovo Cimento **38**, 438 (1965).

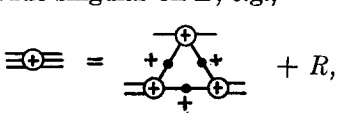
would like to suggest that the positive α condition provides a new way of formulating a causality requirement in S-matrix theory.

One of our aims in this paper is to show that the positive α condition (and hence causality and the hierarchical structure) does seem to be a consequence of the usual analyticity and unitarity assumptions of S-matrix theory.[11,14] At the same time we bear in mind the possibility that the positive α condition could be used to supplement a much weaker sort of analyticity assumption and prove, for instance, that single-particle thresholds are poles (as is suggested by the fact that this already follows from a crude notion of causality[15,16]).

Organization of the Paper

The theoretical considerations lead us to consider the same graph as before (see Fig. 1), because this gives rise to the simplest nondegenerate physical-region singularity which is not a normal threshold. Although, for simplicity, our unitarity equations are written down for an equal-mass theory, we label the masses in Fig. 1 with different values because the work is immediately generalizable. If we fix t at a physical value $[t < (m_1 - m_2)^2]$, the Landau curve L in the σ_1 and σ_2 variables is a hyperbola whose branch lying in the physical region is shown in Fig. 2.[17] L touches the σ_1 and σ_2 normal thresholds at A and B. It is the arc AB which corresponds to positive α's and which is supposed to be singular according to the criteria (1.5).

We discuss two methods of analyzing or extracting the singularity. That of Sec. 2 is the simpler and may be extended more readily to the analysis of further singularities, as we briefly show. The end result consists of several representations for the part of the amplitude singular on L; e.g.,



where the notation of Ref. 11 is used and where the

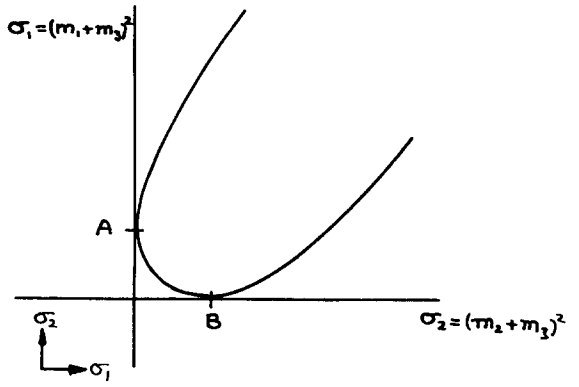
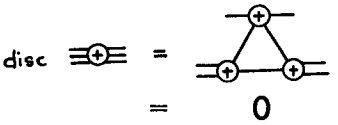


FIG. 2. Part of the singularity curve for Fig. 1 drawn in the real (σ_1, σ_2) plane for t fixed at a physical value. The arc AB corresponds to positive values of the parameters α.

internal lines represent propagators $(q^2 - m^2 + i\epsilon)^{-1}$. The drawbacks of this method are that it involves off-mass-shell amplitudes and that it makes the positive α criterion an assumption, although these difficulties could be eliminated at the cost of much greater complexity.

The method of Sec. 3 uses only the physical unitarity and analyticity requirements as stated in Ref. 11, and with a certain amount of mathematical manipulation involving Cayley determinants[18] leads to a rigorous proof of the Cutkosky formulas[9] for the discontinuity across L, in the form



(1.6)

Here, the internal lines represent factors $-2\pi i\,\delta^+(q^2 - m^2)$. Hence, the amplitude is nonsingular on the arcs ∞A and $B\infty$ and the positive α condition thus deduced. We stress that, throughout, the methods are independent of perturbation theory and involve no assumptions concerning crossing or Hermitian analyticity, etc.

In the Appendix, we state and prove a theorem giving necessary and sufficient conditions for the singularity of certain multidimensional integrals. This theorem is used in the text.

2. ELEMENTARY METHOD

One-Particle Structure

We first recall the derivation[11] of the one-particle structure that was mentioned in Sec. 1. For the six-point function, the one-particle singularities can

[14] See R. J. Eden, P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne, in Ref. 12.

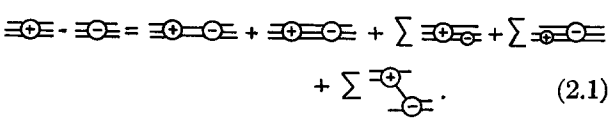
[15] D. Branson, Phys. Rev. 135, 1255 (1964).

[16] R. J. Eden and P. V. Landshoff, Ann. Phys. (N. Y.) 31, 370 (1965); G. Wanders, Helv. Phys. Acta 38, 142 (1965); H. P. Stapp, Phys. Rev. 139, B257 (1965); D. Iagolnitzer, J. Math. Phys. 6, 1576 (1965); A. Peres, Ann. Phys. (N. Y.) 37, 179 (1966).

[17] M. Fowler, P. V. Landshoff, and R. W. Lardner, Nuovo Cimento 17, 956 (1960). There is a printer's error in this paper; Fig. 1(b) should be rotated clockwise through 90°. [This is the figure that is relevant to the present discussion; it corresponds to the momentum transfer $(p_1 - p_4)^2$, where the momenta are labeled as in (2.12), being fixed at a physical value. Part of it is reproduced here as Fig. 2.]

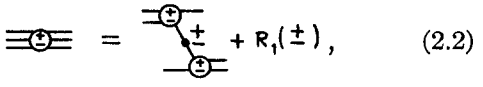
[18] See T. Regge and G. Barucchi, Nuovo Cimento 34, 106 (1964); also A. C. Aitken, *Determinants and Matrices* (Oliver and Boyd, Edinburgh, 1954).

enter the physical region and so their presence may be deduced from the physical unitarity conditions. Below the four-particle threshold this reads[11,14]



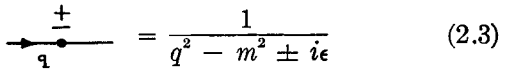
(2.1)

Here, (+) denotes a physical amplitude and (−) its Hermitian conjugate as defined in Ref. 11 or 14. The last term in (2.1) contains δ-function factors and may, together with a causality or analyticity requirement, be shown to require the existence of single-particle poles in the amplitude. So we may write



(2.2)

where



(2.3)

and $R_1(\pm)$ is free of the single-particle singularities.

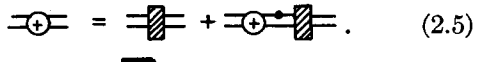
Notice that actually the pole terms are uniquely defined only at the poles $q^2 = m^2$. Away from the pole we have no unique prescription for the two scattering amplitudes that occur as factors in the residues of the poles; different prescriptions lead to different definitions of the background term $R_1(\pm)$, but, so long as they are analytic, $R_1(\pm)$ will be free of the one-particle singularities. However, each of the scattering amplitudes retains three of its momenta on the mass shell and these may be used to define the usual Mandelstam variables s, t, u. On the mass shell, we may express the scattering amplitude as a function of two of these variables, $F(s, t)$ say, since we have

$$s + t + u = \sum m^2. \tag{2.4}$$

We may *choose* to use the same function $F(s, t)$ off the mass shell even though (2.4) no longer applies. Although this is not altogether satisfactory, it at least avoids the introduction of a completely new function; the lack of elegance appears to be inherent in the theory.

Extraction of the Singularity

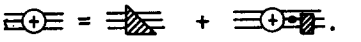
To extract the simple triangle singularity, it is convenient to use Branson's[15] alternative to (1.1), which reads



(2.5)

The new amplitude  defined by this equation is again free of the two-particle singularity in the direct channel. The difference between this equation and (1.1) is that one of the δ functions in the integration in the last term of the latter equation has been replaced by a pole. Although (2.5) involves off-mass-shell amplitudes, at most one line in each amplitude is off the mass shell and so, as we explained above, no new functions need be introduced.

In close analogy with (2.5), we define an amplitude  by the equation



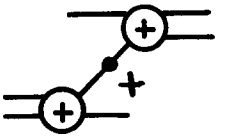
(2.6)

The last term in this equation displays a two-particle singularity in one of the right-hand subenergies; this singularity, which we refer to below as the X singularity, does not appear in  . To prove this, one may either use a close analogy to Branson's argument,[15] or prove directly that the corresponding discontinuity of  vanishes. Either method makes use of the expression[19]

(2.7)

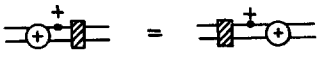
for the corresponding discontinuity of  .

According to (2.2) the left-hand side of (2.6) possesses the pole



(2.8)

The last term in (2.6) cannot contain such a pole; hence  must contain it. If we now "post-multiply" (2.6) by  and use, together with (2.5), the elementary property



(2.9)

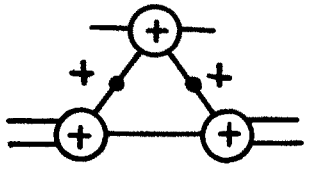
(whose validity is seen by simply writing down the integral that each symbol represents), we obtain

(2.10)

Hence, insertion of (2.8) into the right-hand side

[19] The proof of this is discussed by D. I. Olive, Nuovo Cimento **37**, 1422 (1965), and by R. C. Hwa, Phys. Rev. **134**, B1086 (1964). The method of Olive is more direct than that of Hwa.

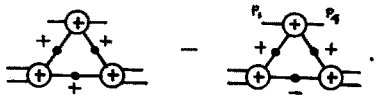
of (2.10) reveals that the last term in (2.6) has the simple triangle singularity whose structure is



(2.11)

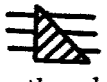
If we can show that the other term  on the right-hand side of (2.6) is free of this triangle singularity, (2.11) will also be the structure of the triangle singularity in the complete amplitude  .

Actually this is the case only for the parts of the triangle singularity occurring in the physical region, and we content ourselves with discussing this. Because the difference between the (+) and the (−) poles in (2.3) is $-2\pi i \times$ a δ function, we may write (2.11) as

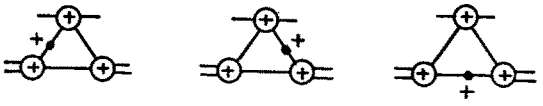


(2.12)

The first term is just what we expect from perturbation theory. As far as the sheet properties of the triangle singularity itself in this term are concerned (the term does contain other singularities because of the structure of the bubbles at the vertices), these are the same as the well-known[17,14] properties of the Feynman graph obtained by neglecting the structure of the amplitudes in the vertices. The singularity curve touches the normal threshold X; on one side of the contact, that side corresponding to positive α and positive free-flight times for the internal particles, it is singular, and in the other side it is not. The second term in (2.12) is singular nowhere in the physical region for, as we show in the Appendix, in order for it to yield a physical-region singularity, the α's must have the same signs as the (+) and the (−) labels on the poles. This is not the case anywhere in the physical region[17]; to reach such a part of the singularity curve from that part on which all the α's are positive, one would have to pass through the contact with the normal threshold in the momentum-transfer variable $(p_1 - p_4)^2$, which is out of the physical region.

Now, if  were to possess the triangle singularity in the physical region, it would have to be singular on both sides of the contact with the normal threshold X, because  does not contain the X-singularity, and so the X-cut is not available to "switch off" the triangle singularity. So, if  were tocontain the triangle singularity, the whole amplitude  would be singular in a part of the physical region where it would have to correspond to negative free-flight times for the internal particles. If we accept the Coleman–Norton assumption, this is forbidden. Therefore, we conclude that the complete triangle singularity structure of the whole amplitude  in the physical region is given by (2.11) or, equivalently, by the first term in (2.12).

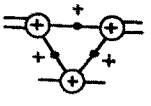
Either of these two equivalent results involves off-mass-shell amplitudes. Further, in these amplitudes, more than one line is off the mass shell, and so we cannot choose to define away the resulting arbitrariness by the same method we gave for the residues of the single-particle poles. This unsatisfactory feature may be alleviated by noting that either of the two forms is equivalent to any of the three forms



in the sense that they differ by functions that are not singular when the former are singular (though they are singular in parts of the physical region where the former forms are not singular). We *cannot* do better than this; just as in the case of the poles, it is only the *discontinuity* associated with the singularity that is completely and uniquely defined in terms of on-mass-shell amplitudes. In Sec. 3, we give an alternative analysis which shows that this discontinuity is given by the expression Cutkosky[9] obtained from perturbation theory.

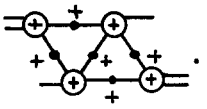
Other Singularities

The methods described above enable us to extract other singularities quite easily. For example, by exact analogy, we know that the amplitude  has the singularity



(2.13)

It follows from (2.6) that  also has this singularity. So, if we use (2.10) and then argue just as before, we find that  has the physical-region singularity

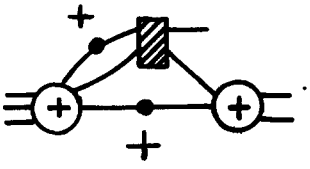
 (2.14)

(In the equal-mass case, this is degenerate; only one point is singular in the physical region. However, the arguments are not confined to the equal-mass case.)

Again, by exact analogy with the last term in (2.6), we know that the structure of the two-particle cut in a different right-hand subenergy is

 (2.15)

[There should be a "spot" on the upper internal line in (2.15).] Maneuvers, closely similar to, but a little more complicated than the previous ones, lead to the conclusion that  has a singularity whose structure is essentially

 (2.16)

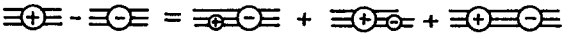
We hope to discuss this in a subsequent paper. (In the equal-mass case it is again degenerate.)

3. CUTKOSKY FORMULA

Use of the Unitarity Condition

In this section, we use the full unitarity relation (2.1) to determine the discontinuity of  associated with the physical-region singularity represented by the Landau–Cutkosky diagram of Fig. 1. The work is independent of that in Sec. 2 and does not involve off-mass-shell amplitudes. Its result will be that the discontinuity vanishes except for the part of the singularity curve L associated with positive α parameters, thus confirming the Coleman–Norton assumption, and for that part the result is the same as that derived by Cutkosky[9] from perturbation theory.

The unitarity relation (2.1) reads



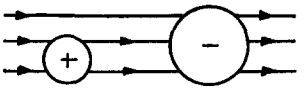
+ (terms analytic on L). (3.1)

The singularity L is generated in each of the terms displayed explicitly on the right-hand side of (3.1) as a result of the presence of the single-particle poles (2.2) in  . The singularity L is also self-regenerating: it occurs in each of the first two terms on the right-hand side of (3.1) as a direct result of its occurring in the amplitudes  appearing within those terms.

First Two Terms

If we denote the variables and masses as labeled in Fig. 1, the part of L lying in the physical region of the real (σ_1, σ_2) plane, for t fixed at a physical value, is[17] as drawn in Fig. 2. In this figure the straight lines are the normal thresholds $\sigma_1 = (m_1 + m_3)^2$, $\sigma_2 = (m_2 + m_3)^2$. It is the arc AB of L, between its contacts with the normal thresholds, that corresponds to positive α and which we aim to show represents a singularity of  .

We first consider how the singularity is generated by the term

 (3.2)

in (3.1) by the pole in  . The lines in (3.2) bear momenta as follows: reading from top to bottom, the external lines on the left bear momenta p_1, p_2, and p_3, respectively, those on the right, p_4, p_5, and p_6, and the internal lines, q_1 and q_3. In this term an integration is implied over the internal four momenta q_1 and q_3. The boundary of the region of this integration is expressible in terms of a Gram-determinant condition

$$G(q_3, p_1, p_4, p_5 + p_6) < 0$$

or, equivalently,

$$G(q_3, q_3 - p_5 - p_6, q_3 - p_4 - p_5 - p_6, q_3 + p_1 - p_4 - p_5 - p_6) < 0. \quad (3.3)$$

This last condition may be rewritten in terms of a Cayley determinant[18] as

$$M(u_2, v) \equiv \begin{vmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & \sigma_2 & s & \sigma_1 & m_3^2 \\ 1 & \sigma_2 & 0 & M_4^2 & t & u_2 \\ 1 & s & M_4^2 & 0 & M_1^2 & v \\ 1 & \sigma_1 & t & M_1^2 & 0 & m_1^2 \\ 1 & m_3^2 & u_2 & v & m_1^2 & 0 \end{vmatrix} > 0, \quad (3.4)$$

where $p_i^2 = M_i^2$, $q_i^2 = m_i^2$, $s = (p_4 + p_5 + p_6)^2$, $u_2 = (q_3 - p_5 - p_6)^2$, $v = (p_1 + q_1)^2$. The variables u_2, v may be used as the integration variables instead of q_1, q_3, and if we suppose that the v integration has been done, we obtain from (3.4) the end point of the remaining u_2 integration by solving the equations

$$M = \partial M/\partial v = 0. \quad (3.5)$$

The triangle singularity arises from a coincidence of this end point with the single-particle pole

$$u_2 = m_2^2 \quad (3.6)$$

of the amplitude  that occurs in the integrand. From (3.4) we see that

$$\partial M/\partial v = 2M_{4,6},$$

where $M_{i,j}$ is the (i, j) algebraic minor of the determinant M, so that Jacobi's identity[18]

$$M_{i,i}M_{j,j} - M_{i,j}M_{j,i} = MM_{ij,ij} \quad (3.7)$$

tells us that (3.5) implies either $M_{4,4} = 0$ or $M_{6,6} = 0$. The first possibility is the one that is of interest here (the other yields a second-type singularity[20]); it gives as the equation for L

$$R(m_1^2, m_2^2) = 0, \quad (3.8)$$

where

$$R(m_1^2, u_2) \equiv M_{4,4}(u_2) = \begin{vmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & \sigma_2 & \sigma_1 & m_3^2 \\ 1 & \sigma_2 & 0 & t & u_2 \\ 1 & \sigma_1 & t & 0 & m_1^2 \\ 1 & m_3^2 & u_2 & m_1^2 & 0 \end{vmatrix}. \quad (3.9)$$

This form of the equation for L is more useful here than the more familiar form.[17]

We calculate the discontinuity of (3.2) arising from this mechanism of generating the triangle singularity, by fixing t and taking σ_1, σ_2 around L on a path lying close to L, and in a plane normal to L in the four-dimensional complex (σ_1, σ_2) space. Thus, on the path, the displacement of (σ_1, σ_2) from L is given by

$$\begin{aligned} d\sigma_1 &= d\eta[(\partial/\partial\sigma_1)R(m_1^2, m_2^2)], \\ d\sigma_2 &= d\eta[(\partial/\partial\sigma_2)R(m_1^2, m_2^2)], \end{aligned} \quad (3.10)$$

with the derivatives evaluated on L. It may readily be ascertained that the signs of the derivatives are such that a real positive $d\eta$ corresponds to a displacement along the inward normal to L. For a displacement (3.10),

$$dR = Q\, d\eta + [(\partial/\partial u_2)R(m_1^2, u_2)]_{u_2 = m_2^2} \cdot du_2, \quad (3.11)$$

where

$$Q = [(\partial/\partial\sigma_1)R(m_1^2, m_2^2)]^2 + [(\partial/\partial\sigma_2)R(m_1^2, m_2^2)]^2, \quad (3.12)$$

so that the displacement of the integration end point from $u_2 = m_2^2$, which is given by $dR = 0$, is

$$du_2 = -Q\, d\eta/(\partial R/\partial u_2). \quad (3.13)$$

Now, in terms of a cofactor of the determinant (3.9),

$$\partial R/\partial u_2 = R_{3,5}$$

and Jacobi's identity (3.7) applied to R gives, on L,

$$(R_{3,5})^2 = R_{3,3}R_{5,5}.$$

Hence, $du_2/d\eta$ changes sign when either $R_{3,3} = 0$ or $R_{5,5} = 0$. The latter possibility does not occur on the branch of L we are discussing; the former occurs at the normal threshold $\sigma_1 = (m_1 + m_3)^2$, that is at A in Fig. 2. So $du_2/d\eta$ takes different signs on the arcs ∞A, AB∞ in Fig. 2, and it is easy to show that it is negative on the former and positive on the latter by calculating it at B using the fact that L always touches the σ_2 normal threshold which itself increases with m_2. Using Jacobi's theorem, and the fact that Cayley determinants of real external vectors are positive in the physical region, it follows that outside L in the σ_1, σ_2 plane, $M < 0$ when $u_2 = m_2^2$ so that the point $u_2 = m_2^2$ lies inside the integration region only for points inside L. Hence, the orientation of the contour in the u_2 plane for variations in η is as shown in Fig. 3.

It follows from the last column that the discontinuity in η of the integral, across L, is obtained from the integral (3.2) by replacing the factor with the pole by the residue of that pole times the factor $\pm -2\pi i\, \delta(u_2 - m_2^2)$, with the sign corresponding to arcs AB∞ and ∞A, respectively. Transforming back from the integration over invariants to the

[20] D. B. Fairlie, P. V. Landshoff, J. Nuttall, and J. C. Polkinghorne, J. Math. Phys. 3, 594 (1962).

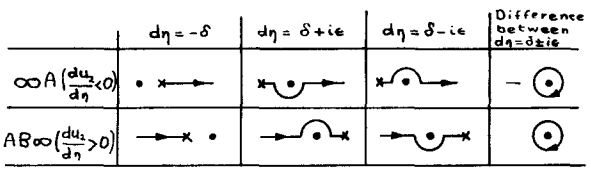
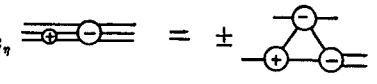


FIG. 3. Table representing orientation of the u_2-integration contour in (3.2) with respect to the pole $u_2 = m_2^2$. The column $d\eta = -\delta$ corresponds to points (σ_1, σ_2) on the physical sheet outside L; the next two columns correspond to continuations to points inside L by paths in Im $\eta > 0$ and Im $\eta < 0$, respectively. The last column represents the contour appropriate to the discontinuity associated with the singularity generated by the pole.

loop integration, the answer can be written unambiguously,[21]

$$\text{disc}_\eta \quad = \pm \quad (3.14)$$



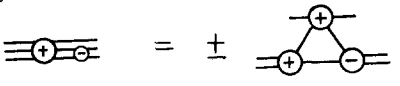
with + on AB∞ and − on ∞A (the subscript η is to emphasize that we take the discontinuity in the variable η). This integral is to be interpreted in terms of the unitarity rules [Eq. (2.7) of Ref. 11].

An analogous mechanism works for the term

(3.15)

in the unitarity relation (3.1). The resulting discontinuity is

$$\text{disc}_\eta \quad = \pm \quad (3.16)$$

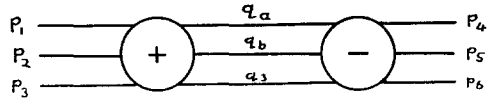


with + on ∞AB and − on B∞.

In obtaining (3.14) and (3.16), we have ignored the regeneration effect mentioned earlier, whereby the integrals are singular on L because the larger bubbles are. We take account of this in the final stage of the argument.

Third Term

We next consider the discontinuity arising from the insertion of single-particle poles in each of the amplitudes  in the term



(3.17)

of the unitarity condition (3.1).

Part of the boundary of the integration region is given in terms of a Gram determinant as

$$G(q_3, q_3 - p_5 - p_6, q_3 - p_2 - p_3, -q_b) < 0, \quad (3.18)$$

which, in terms of a Cayley determinant,[12] reads

$$N(u_1, u_2, \lambda, \mu_1, \mu_2) \equiv \begin{vmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & \sigma_2 & \sigma_1 & \lambda & m_3^2 \\ 1 & \sigma_2 & 0 & t & \mu_2 & u_2 \\ 1 & \sigma_1 & t & 0 & \mu_1 & u_1 \\ 1 & \lambda & \mu_2 & \mu_1 & 0 & m_b^2 \\ 1 & m_3^2 & u_2 & u_1 & m_b^2 & 0 \end{vmatrix} > 0, \quad (3.19)$$

where $\lambda = (q_b + q_3)^2$, $\mu_1 = (p_1 - q_a)^2$, $\mu_2 = (p_4 - q_a)^2$, and $u_1 = (q_3 - p_2 - p_3)^2$; and $u_2 = (q_3 - p_5 - p_6)^2$, $\sigma_1 = (p_2 + p_3)^2$, $\sigma_2 = (p_5 + p_6)^2$, and $t = (p_4 - p_1)^2$ as before. The variables u_1, u_2, λ, μ_1, and μ_2 may be used as the integration variables in (3.17). If the λ, μ_1, μ_2 integrations are done, (3.19) produces as boundary of the remaining u_1, u_2 integrations the curve obtained by solving the equations

$$N = \partial N/\partial\lambda = \partial N/\partial\mu_1 = \partial N/\partial\mu_2 = 0. \quad (3.20)$$

These equations are equivalent to

$$N = N_{2,5} = N_{3,5} = N_{4,5} = 0,$$

and the Jacobi identity (3.7) applied to N results in

$$R(u_1, u_2) \equiv N_{5,5} = 0 \quad (3.21)$$

as the equation for the boundary. The triangle singularity arises from this boundary curve passing through the intersection of the singularities

$$u_1 = m_1^2, \qquad u_2 = m_2^2 \quad (3.22)$$

of the integrand, as follows from the discussion of the Appendix. So the equation of L is obtained as

$$R(m_1^2, m_2^2) = 0,$$

which, since it may readily be seen that the determinants R defined in (3.21) and (3.9) are the same, is the same equation as we previously obtained in (3.8).

If necessary, to avoid a singularity of the integrand, the hypercontour of the (u_1, u_2)-integration may be distorted into complex (u_1, u_2)-space. The boundary may likewise be distorted so long as it remains on the complex surface $R(u_1, u_2) = 0$. This has the result that the arc AB is not a singularity of (3.17) on the physical sheet, as we now show. The Jacobi identity (3.7) applied to the determinant (3.21) that defines R gives

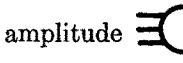
$$\frac{\partial R}{\partial u_1} \Big/ \frac{\partial R}{\partial u_2} = \frac{\partial R}{\partial \sigma_1} \Big/ \frac{\partial R}{\partial \sigma_2} \quad (3.23)$$

[21] Except for the important question of this sign, this result was obtained by Polkinghorne in the second paper of Ref. 10.

when σ_1 and σ_2 are on L, and when (3.22) applies. Hence, for σ_1, σ_2 on the arc AB in Fig. 2, the gradient of the curve $R = 0$ in the real (u_1, u_2) plane is negative in the neighborhood of $u_1 = m_1^2$, $u_2 = m_2^2$. This means that the attached complex parts of $R = 0$ lie in the regions

$$\text{Im } u_1/\text{Im } u_2 < 0 \tag{3.24}$$

as may be seen by the standard searchline method of tracing complex surfaces attached to real curves.[8,12,22]

The left-hand amplitude  in (3.17) carries the label (+) and so, according to (2.2) and (2.3), its pole lies just below the real axis and we avoid it if we can distort the boundary of the region of integration into Im $u_1 > 0$. The right-hand amplitude  carries the label (−), so we avoid its pole if we can distort the boundary into Im $u_2 < 0$. Because of (3.24), these two distortions are simultaneously allowed for the arc AB, so that the critical configuration of hypercontour and singularities of the integrand is avoided and this arc is not singular. However, on the arcs ∞A and B∞, the complex surfaces attached to $R = 0$ lie in

$$\text{Im } u_1/\text{Im } u_2 > 0 \tag{3.25}$$

so that the distortions are not simultaneously possible and these arcs represent singularities of (3.17).

We now calculate the corresponding discontinuities, by again taking σ_1, σ_2 around L on a path of the type (3.10). The analysis is more difficult than before in that this time it is necessary to consider the distortion of the hypercontour in two complex variables u_1, u_2. Presumably, this can be done directly by the techniques of homology theory, but here we use a simplified method. The integral under study is of the form

$$\int_{R>0} du_1\, du_2 \frac{f(u_1, u_2)}{(u_1 - m_1^2 + i\epsilon)(u_2 - m_2^2 - i\epsilon)}. \tag{3.26}$$

If we take $d\eta$ in (3.10) small, only the part of the hypercontour in the neighborhood of $u_1 = m_1^2$, $u_2 = m_2^2$ changes critically as we go around L, so the resulting discontinuity of (3.26) must be equal to

$$f(m_1^2, m_2^2) \times \text{disc} \int du_1\, du_2 \frac{1}{(u_1 - m_1^2 + i\epsilon)(u_2 - m_2^2 - i\epsilon)}. \tag{3.27}$$

[22] J. Tarski, J. Math. Phys. 1, 149 (1960).

The critical part of the hypercontour is bounded by the curve $R = 0$, but since only a small portion of this curve plays a critical part, we may replace it by its tangent, whose equation is

$$(u_1 - m_1^2)\frac{\partial R}{\partial u_1} + (u_2 - m_2^2)\frac{\partial R}{\partial u_2} + \left[\frac{\partial R}{\partial \sigma_1} d\sigma_1 + \frac{\partial R}{\partial \sigma_2} d\sigma_2\right] = 0$$

or, using (3.10),

$$(u_1 - m_1^2)(\partial R/\partial u_1) + (u_2 - m_2^2)(\partial R/\partial u_2) = -Q\, d\eta. \tag{3.28}$$

Here, Q is again defined by (3.12) and all the derivatives are evaluated for $u_1 = m_1^2$, $u_2 = m_2^2$, and σ_1, σ_2 on L. We change the integration variables to

$$\begin{aligned} x &= -(u_1 - m_1^2)(\partial R/\partial u_1) - (u_2 - m_2^2)(\partial R/\partial u_2), \\ y &= (u_1 - m_1^2)(\partial R/\partial u_2) - (u_2 - m_2^2)(\partial R/\partial u_1), \end{aligned} \tag{3.29}$$

with Jacobian

$$\begin{aligned} \partial(x, y)/\partial(u, v) &= (\partial R/\partial u_1)^2 + (\partial R/\partial u_2)^2 \\ &= A, \quad \text{say.} \end{aligned} \tag{3.30}$$

The integration region now becomes the half plane $x \leq Qd\eta$ so that the integral in (3.27) becomes

$$-A \int^{Qd\eta} dx \int_{-\infty}^{\infty} dy \left[\left(-x\frac{\partial R}{\partial u_1} + y\frac{\partial R}{\partial u_2} + i\epsilon'\right) \times \left(x\frac{\partial R}{\partial u_2} + y\frac{\partial R}{\partial u_1} + i\epsilon'\right)\right]^{-1}. \tag{3.31}$$

The y integration can be closed by adding a large semicircle in the upper half-plane (or in the lower half-plane—the answer is the same) and then evaluated by the residue theorem. The answer is nonzero only if $(\partial R/\partial u_1)/(\partial R/\partial u_2) < 0$, in agreement with our previous discussion of (3.23). In this case (3.31) becomes

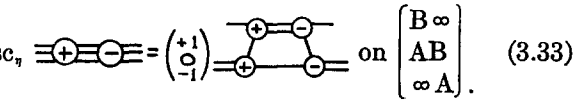
$$\pm 2\pi i A \int^{Qd\eta} \frac{dx}{Ax \pm i\epsilon''} = \pm 2\pi i \int^{Qd\eta} \frac{dx}{x \pm i\epsilon'''} \tag{3.32}$$

with the alternative signs applying according as it is $\partial R/\partial u_1$ or $\partial R/\partial u_2$ that is negative. We already know, from the discussion following (3.13), that the former holds on the arc ∞A, the latter on B∞. If we take $d\eta$ in an anticlockwise circuit round the origin, the upper end point of the integral performs

a similar circuit, so we pick up the discontinuity $\pm (2\pi i)^2$. Therefore, the discontinuity of (3.26) is

$$\binom{+1}{0}_{-1} \int_{R>0} du_1\, du_2\, f(u_1, u_2) \times [-2\pi i\, \delta(u_1 - m_1^2)][-2\pi i\, \delta(u_2 - m_2^2)]$$

with $+1$, 0, or -1 on B∞, AB, and ∞A, respectively. Applying this to (3.17) and using the known residues of the poles of the integrand, we have for the discontinuity in η across L

disc_η  on $\begin{bmatrix} \text{B}\infty \\ \text{AB} \\ \infty\text{A} \end{bmatrix}$. (3.33)

Comparison of the Unitarity Equations on Either Side of L

The object now is to calculate the discontinuity of the (+) amplitude across the Landau curve L. We propose to do this by comparing the versions of the unitarity equation (3.1) valid on either side of L, but, before we can do this, we must understand how the various terms are analytically related across L (if at all).

First, let us consider the (+) amplitude, the first term of (3.1). So far we do not know whether or not it is singular on the whole of L so that until we prove otherwise we suppose that all of L is singular. According to the general ideas of analyticity,[11] there is to be some way of continuing around L in order to relate analytically the physical (+) amplitudes defined on either side. Looking at the complex η plane (Fig. 4) at some point P of L, we see that L intersects it at one point (taken to be the origin) and that there are two ways of continuing from real points outside L ($\eta < 0$) to real point inside L ($\eta > 0$), either with an $(\eta + i\epsilon)$ detour or an $(\eta - i\epsilon)$ detour. As P moves around L, the detour chosen must produce the same result for each point of L, so that we must consistently choose an $(\eta + i\epsilon)$ distortion for all points or else an $(\eta - i\epsilon)$ distortion for all points. As we saw

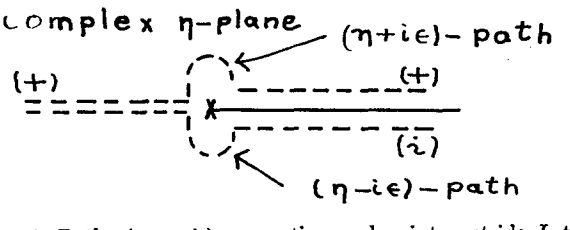


FIG. 4. Paths $(\eta \pm i\epsilon)$ connecting real points outside L to those inside L.

in Fig. 2, L touches the σ_1 and σ_2 normal thresholds at A and B, respectively, and so η, which is always measured inwards, agrees in sense with the variables σ_1 and σ_2 at the points A and B, respectively. In order to preserve the single valuedness of the physical amplitude,[11] it is necessary that the complex distortions chosen for two curves should agree when the curves touch. Since we have already chosen $(\sigma_1 + i\epsilon)$ and $(\sigma_2 + i\epsilon)$ distortions for the normal thresholds, it follows that L must have an $(\eta + i\epsilon)$ distortion (note that, because there are two points of tangency, there are two conditions which are fortunately consistent with each other: this indicates that we could have deduced one normal threshold $i\epsilon$ prescription from the other and hence weakened our analyticity assumption).

Similarly (or just by complex conjugation), we would obtain an $(\eta - i\epsilon)$ rule for the continuation linking the physical (−) amplitude.

The physical unitarity equation (3.1) is valid to either side of L, but is not an analytic continuation of itself across L. Consider the three terms (3.2), (3.15), and (3.17) appearing on the right-hand side of (3.1). We shall see that, in the continuation from outside to inside L, the arrangement of integration contours with respect to the singularities generating L depends on whether we follow an $(\eta + i\epsilon)$ or an $(\eta - i\epsilon)$ path, and, for each term, only one of these arrangements of contours corresponds to that understood in the corresponding term of the unitarity equation (3.1) as operating inside L. This particular distortion, giving rise to the "natural arrangement," we call the "natural distortion." It varies from term to term and from segment to segment of L, as we now see.

When the integral (3.2) is evaluated inside L ($\eta > 0$) the label (−) on the right-hand amplitude in the integral indicates that the integration contour in the μ_2 plane is depressed below the pole $u_2 = m_2^2$. Looking at Fig. 3, we see that such a contour is obtained by continuing the integral (3.2) defined outside L ($\eta < 0$) with an $(\eta + i\epsilon)$ path when the arc ∞A is traversed, and with an $(\eta - i\epsilon)$ path when the arc AB∞ is traversed. Thus, distortions $(\eta + i\epsilon)$ and $(\eta - i\epsilon)$ are "natural" on ∞A and AB∞, respectively.

A similar discussion can be made for the integral (3.15), but the integral (3.17), which is the third term on the right-hand side of (3.1), is slightly different. In this case, the $\pm i\epsilon'''$ in the denominator of (3.32) is equivalent to associating $\pm i\epsilon'''$ with the $d\eta$ in the end point and tells us that $(\eta + i\epsilon)$ and $(\eta - i\epsilon)$ distortions are natural on ∞A and B∞,

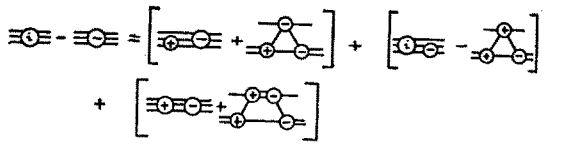
respectively. These results are tabulated in Fig. 5 for convenience.

So far we have mentioned only the generation of the triangle singularity L by the poles. We must also remember that L also occurs in each of the terms (3.2) and (3.15) as a result of its occurrence within the amplitudes  appearing in those terms. Accordingly, in the continuation of either of these terms across L, the amplitudes within the integrals must be continued along the same path as the term (as a whole).

Calculation of the Discontinuity in the (+) Amplitude across L

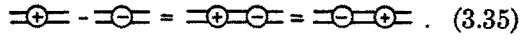
We now have enough information to continue the unitarity equation valid outside L into the region inside L. We shall follow an $(\eta - i\epsilon)$ path so that the $(-)$ amplitude is continued into the $(-)$ amplitude, while the $(+)$ amplitude is continued into a new region (i) separated from the physical $(+)$ boundary value by the cut attached to the singularity whose discontinuity we seek (see Fig. 4). This new equation will relate the (i) and $(-)$ amplitudes, in contradistinction to the physical unitarity equation operating inside L, which relates the $(+)$ and $(-)$ amplitudes. This fact indicates that the two unitarity equations (3.1), valid inside and outside L while looking similar, are not analytic continuations of each other. As we see, it is the discrepancy between them that yields the desired discontinuity.

We first suppose that the arc ∞A is traversed. Figure 4 shows that the $(\eta - i\epsilon)$ distortion is not the natural one for any of the terms (3.2), (3.15), and (3.17). We put this right by using the discontinuity formulas (3.14), (3.16), and (3.33), thus finding for the continuation of (3.1)



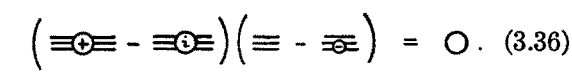
+ terms analytic on L. (3.34)

Using the two-particle unitarity relation



(3.35)

we find that the second, fourth, and sixth terms on the right-hand side of (3.34) cancel. If we subtract the result from the unitarity equation valid inside L, as given by (3.1), and rearrange, we have



(3.36)

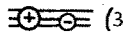
ARC OF L / INTEGRAL	∞A	AB	B∞
 (3.2)	$\eta+i\epsilon$	$\eta-i\epsilon$	$\eta-i\epsilon$
 (3.15)	$\eta+i\epsilon$	$\eta+i\epsilon$	$\eta-i\epsilon$
 (3.17)	$\eta+i\epsilon$	either	$\eta-i\epsilon$

FIG. 5. Table of paths of continuation from points outside L to points inside L giving "natural arrangement" of contours.

If we "postmultiply" this equation by the expression

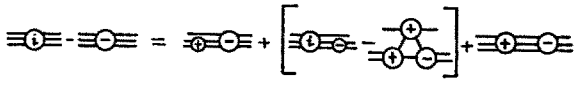
(3.37)

and again use (3.35) we obtain

(3.38)

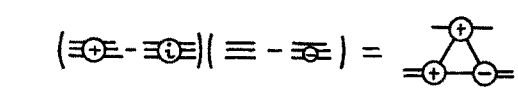
Hence, the arc ∞A is not a singularity of the $(+)$ amplitude.

The $(\eta - i\epsilon)$ continuation across the arc AB is unnatural only for the term (3.15) (according to Fig. 5), and so we find [using Eq. (3.16)]

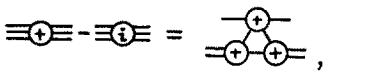


+ terms regular on L.

Subtracting from (3.1) and rearranging, we have



Postmultiplying by expression (3.37) and using (3.35) as before, we obtain finally



which is the formula predicted by Cutkosky.[9]

Repetition of the procedure yields (3.36) and hence (3.38) on B∞, so that our final result can be expressed in the form (1.6). Corresponding results for the $(-)$ amplitude can be found by considering an $(\eta + i\epsilon)$ path of continuation.

The analysis given applies only when (3.1) is valid, and, in particular, when $\sigma_1 < (3m)^2$ and $\sigma_2 < (3m)^2$, so that at higher energies more complicated equations must be studied. We could, however, extend the result by analytic continuation of the above formulas to higher energies. That the two methods should give the same result is yet another consistency requirement.

4. COMMENTS

We briefly mention some possible consequences or extensions of our fundamental result (1.6) which we hope to discuss at greater length in later papers.

(a) We think it must be possible to prove the positive-α singularity criterion and the Cutkosky discontinuity formulas for all physical-region singularities. We can see that new difficulties arise because a simplifying feature of our example was that it involved only single-particle thresholds.

(b) It is possible to show by analytic continuation arguments that our specific result (1.6) determines the sheet upon which the anomalous threshold singularity occurs in the two-particle amplitude. This has been a long-standing problem in S-matrix theory.[10]

(c) If arguments of type (b) can be extended, it may be possible to show that, since unitarity controls singularity structure in the physical region, it controls it everywhere. Furthermore, since the "hierarchical structure" discussed in the introduction holds in the physical region, there is presumably an analogous property elsewhere. It is just this sort of property that is needed to establish the "iteration of singularities" idea and complete the construction of an axiomatic S-matrix theory.[11]

(d) We assumed that the single-particle singularities were poles with a $(+i\epsilon)$ prescription. We think it possible to use our analysis to show that alternative possibilities are either self-contradictory or in contradiction with the causality requirement discussed earlier.

APPENDIX

Singularities in Multiple Integrals

Consider a multiple integral

$$I(p_1, p_2 \cdots p_n) = \int_A dk_1\, dk_2 \cdots dk_l \times f(p_1, p_2, \cdots, p_n; k_1, k_2, \cdots, k_l)$$

or, for short,

$$I(p) = \int_A dk\, f(p, k). \tag{A1}$$

The integration region A is supposed real and given by

$$B_j(p, k) \geq 0, \qquad j = 1 \cdots r,$$

where $B_j(p, k)$ are analytic functions. The integrand is supposed analytic except for singularities

$$S_i(p, k) = 0, \qquad i = 1 \cdots m,$$

appearing in the combination $S_i + i\epsilon_i$ so that the deformation of the contour past each one of these singularities is well defined.

The situation described is applicable to Feynman integrals and to unitarity integrals defined in the physical region for the relevant process. The theorem we prove (subject to certain provisos) is:

Theorem: In the neighborhood of the physical region, singularities of (A1) occur only at values of p satisfying the implicit equations

$$\alpha_i = 0 \qquad \text{or} \quad S_i = 0, \tag{A2}$$

$$\beta_i = 0 \text{ and } B_i > 0 \quad \text{or} \quad B_i = 0,$$

$$\sum \alpha_i \frac{\partial S_i}{\partial k} + \sum \beta_i \frac{\partial B_i}{\partial k} = 0, \tag{A3}$$

$$\alpha_i/\epsilon_i \geq 0. \tag{A4}$$

Proof[23]*:* Suppose that for given P there exists a point K (assumed isolated) in the space of integration variables such that

$$S_i(K, P) = 0, \qquad i = 1 \cdots M,$$

$$B_j(K, P) = 0, \qquad j = 1 \cdots R,$$

$$B_j(K, P) > 0, \qquad j = R + 1 \cdots r.$$

In order to make sure that the integral is well defined and analytic, we want to distort the integration contour away from an awkward point like this. Since the contour is initially real, we need only look for imaginary distortions given by δK at the point K. In order that, in the distortion, the contour does not intersect the singularity surfaces near K, we require

$$dS_i = i\sigma_i \text{ where } \sigma_i/\epsilon_i > 0, \qquad i = 1 \cdots M. \tag{A5}$$

The boundary of the integration region intersects the point K concerned and can be distorted provided it still lies on the same analytic manifolds. Hence, we require

$$dB_j = 0, \qquad j = 1 \cdots R. \tag{A6}$$

The $(R + M)$ variations dS and dB are given in terms of the l quantities δK by the differential relations

$$dS_i = \sum_{\lambda=1}^{l} \frac{\partial S_i}{\partial k_\lambda} dk_\lambda, \qquad dB_j = \sum_{\lambda=1}^{l} \frac{\partial B_j}{\partial k_\lambda} dk_\lambda.$$

If the rank of the matrix relating the differentials has rank $(R + M)$, then we can always find δK

[23] The argument used is a development of that of H. P. Stapp, Phys. Rev. 125, 2139 (1962), Appendix H.

giving rise to chosen dS and dB, and in particular to ones satisfying (A5) and (A6).

Difficulties may arise if the rank of the matrix is $(R + M - 1)$. Then we have a linear dependence of the rows so that

$$\sum \alpha_i \frac{\partial S_i}{\partial k_\lambda} + \sum \beta_i \frac{\partial B_i}{\partial k_\lambda} = 0, \qquad \lambda = 1, 2, \cdots, l. \tag{A7}$$

Ranks less than $(R + M - 1)$ are included in the consideration of smaller values of R and M and are not discussed further. Even if (A7) is satisfied it may still be possible to find differentials satisfying (A5) and (A6). Let us choose δK so that (A6) is satisfied. (This must be possible if $R \geq 1$, as we suppose.) Then, by (A6) and (A7),

$$\sum \alpha_i \, dS_i = 0$$

or, equivalently (since $dS_i = i\sigma_i$, with σ_i real),

$$\sum \alpha_i \sigma_i = 0. \tag{A8}$$

If (A4) is satisfied, it must be impossible to find σ_i satisfying (A5), for then each term in the sum in (A8) would be positive, contradicting the fact that the sum vanishes. Thus (A4) [in addition to (A2) and (A3)] implies singularity.

Conversely, we show that singularity implies (A4). Suppose $\alpha_1/\epsilon_1 > 0$ (if there exists no α for which this is true, we multiply all α's by -1). Then, we can choose δK so that (A6) is satisfied, $\sigma_1 \cdots \sigma_{M-1}$ satisfy $\sigma_i/\epsilon_i > 0$, and also that $\sum_{i=1}^{M-1} \alpha_i \sigma_i > 0$ (possible since $\alpha_1 \sigma_1 > 0$). It follows by (A8) that $\alpha_M \sigma_M < 0$. By hypothesis $\sigma_M/\epsilon_M < 0$ (for singularity), and so $\alpha_M/\epsilon_M > 0$. The same argument can be repeated for all α_i, thereby establishing the theorem.

Comments

(i) The condition (A3) means that the normals to the surfaces B and S must be linearly dependent at a point of intersection.

(ii) The singularity surfaces S_i and the boundary surfaces B_i enter the criterion on an equal footing until we come to the final one (A4) which involves only the α's and not the β's.

(iii) It is thought that (A2) and (A3) are necessary conditions for singularity away from the physical region when the contours may be distorted. The additional sufficient condition must be more complicated than (A4) since the manner of distortion must be understood and can no longer be specified in a simple way.

(iv) The result for integrals with unbounded contours has been known for some time.[24] When applied to Feynman integrals evaluated in the physical region, for which

$$S_i = q_i^2 - m_i^2$$

and $+i\epsilon$ is always associated with S_i, (A2), (A3), and (A4) reduce to

$$\text{either} \quad \alpha_i = 0 \quad \text{or} \quad q_i^2 - m_i^2 = 0;$$

$$\sum \alpha_i q_i = 0 \quad \text{around each loop},$$

$$\alpha_i \geq 0,$$

which are Eqs. (1.4) and (1.5) of the text. If integrals with mixed signs of ϵ's are considered, the positive-α condition is replaced by (A4).

(v) In the case of two integration variables k_1 and k_2, one singularity surface $S_1 \equiv k_1 - m_1$, and one boundary B, (A3) reduces to

$$\alpha_1 + \beta(\partial B/\partial k_1) = 0, \qquad \beta(\partial B/\partial k_2) = 0.$$

The second equation tells us that B touches S_1 at their point of intersection. Thus this point is an extremum (or end point) in the k_1 integration. This is the sort of situation encountered in the study of integrals (3.2) and (3.15) in the text.

If there are two singularity surfaces and one boundary surface in two dimensions, (A3) is automatically satisfied when the three surfaces intersect, since three vectors (the three normals) are automatically linearly dependent in two dimensions.

---

[24] P. V. Landshoff and J. C. Polkinghorne (unpublished); J. C. Polkinghorne, Nuovo Cimento **23**, 360 (1962); M. Fowler, J. Math. Phys. **3**, 936 (1962).

Generalized Ward Identity and Unified Treatment of Conservation Laws

D. LURIÉ
Dublin Institute for Advanced Studies, Dublin, Ireland
AND
Y. TAKAHASHI*
Matscience, The Institute of Mathematical Sciences, Madras, India
AND
H. UMEZAWA
Istituto di Fisica Teorica, Napoli, Italy
(Received 3 January 1966)

A technique for deriving conservation laws directly from field equations without recourse to the Lagrangians or Noether's theorem is reviewed and extended. The method allows a simple treatment of the so-called "generalized" conservation laws including Lipkin's "zilch." An interesting feature which results from our approach is the existence of conserved currents for discrete as well as continuous symmetries. It is also pointed out that conservation laws do not always follow from the invariance of equation of motion if it is not derivable from a Lagrangian. Finally, we show how our method can be applied to the normalization of wavefunctions of composite particles such as Bethe–Salpeter wavefunctions.

1. INTRODUCTION

In this paper, we review and extend a technique due to Takahashi and Umezawa[1] for deriving conservation laws for fields of arbitrary spin directly from field equations without explicit recourse to the Lagrange formalism and Noether's theorem. Our approach makes crucial use of a configuration-space version of the generalized Ward identity[2] relating the propagator and vertex. Using this identity, we give a single treatment of both the usual free-field conservation laws and the so-called generalized conservation laws including Lipkin's "zilch"[3,4] for the electromagnetic field. We also extend our method to the case of interacting fields. Finally, we apply the technique to the normalization theory[5,6] of Bethe–Salpeter wavefunctions.[7,8]

A new feature which is brought out by our approach is the existence of conserved currents for *discrete* symmetry transformations as well as continuous ones. We illustrate this point by explicit reference to the case of parity and charge conjugation.

Although the Lagrangian is never used explicitly in our approach, it should be stressed that, for interacting fields, invariance properties of the field equations do not, in general, give rise to conservation laws *unless* the field equations can be derived from a Lagrangian.

The generalized Ward identity is derived in Sec. 2. In Sec. 3, we discuss its application to the derivation of conservation laws for free fields, including discrete symmetries and generalized conservation laws. In Sec. 4, we discuss the extension of the method to interacting fields, and in Sec. 5 we give the application to Bethe–Salpeter wavefunctions.

2. GENERALIZED WARD IDENTITY

Consider a multicomponent field $\phi_r(x)$ ($r = 1, 2, \cdots, n$) representing a field of arbitrary spin, and satisfying a relativistic wave equation of the form

$$\Lambda_{rs}(\partial)\phi_s(x) = j_r(x) \tag{2.1}$$

or in matrix notation

$$\Lambda(\partial)\phi(x) = j(x). \tag{2.2}$$

For the Dirac field, for example, $\Lambda(\partial)$ is simply $\gamma\partial + m$. For a field with arbitrary spin, we assume

$$\Lambda(\partial) = \Lambda_0 + \Lambda_\mu\partial_\mu + \Lambda_{\mu_1\mu_2}\partial_{\mu_1}\partial_{\mu_2} + \cdots + \Lambda_{\mu_1\cdots\mu_N}\partial_{\mu_1}\cdots\partial_{\mu_N}, \tag{2.3}$$

* On leave of absence from Dublin Institute for Advanced Studies, Dublin, Ireland.

[1] Y. Takahashi and H. Umezawa, Nucl. Phys. 51, 193 (1964).

[2] Y. Takahashi, Nuovo Cimento 6, 371 (1957); H. S. Green, Proc. Phys. Soc. (London) 66, 873 (1953); K. Nishijima, Phys. Rev. 119, 485 (1960).

[3] D. M. Lipkin, J. Math. Phys. 5, 696 (1964).

[4] T. A. Morgan, J. Math. Phys. 5, 1659 (1964); T. W. B. Kibble, *ibid.* 6, 1022 (1965); D. M. Fradkin, *ibid.* 6, 879 (1965); D. B. Fairlie, Nuovo Cimento 37, 898 (1965); D. J. Candlin, *ibid.* 37, 1390 (1965); R. F. O'Connell and D. R. Tompkins, *ibid.* 38, 1088 (1965).

[5] R. E. Cutkosky and M. Leon, Phys. Rev. 135, B1445 (1964).

[6] D. Lurié, A. J. Macfarlane, and Y. Takahashi, Phys. Rev. 40, B1091 (1965).

[7] M. Gell-Mann and F. E. Low, Phys. Rev. 82, 350 (1951).

[8] E. Salpeter and H. Bethe, Phys. Rev. 82, 309 (1951).

where we can obviously assume the $\Lambda_{\mu_1\cdots\mu_l}$ to be symmetric with respect to all indices when $l > 1$. Summation on repeated indices is assumed unless otherwise stated.

We now construct the following differential operator[9]

$$\Gamma_\mu(\partial_1 - \overleftarrow{\partial}) = \sum_{l=\partial}^{N-1} \sum_{i=0}^{l} \Lambda_{\mu\mu_1\cdots\mu_l} \times \partial_{\mu_1} \cdots \partial_{\mu_i}(-\overleftarrow{\partial}_{\mu_{i+1}}) \cdots (-\overleftarrow{\partial}_{\mu_l}) = \Lambda_\mu + \Lambda_{\mu\nu}(\partial_\nu - \overleftarrow{\partial}_\nu) + \Lambda_{\mu\nu\rho}(\partial_\nu \partial_\rho - \partial_\nu \overleftarrow{\partial}_\rho + \overleftarrow{\partial}_\nu \overleftarrow{\partial}_\rho) + \cdots, \quad (2.4)$$

whose raison d'être is that it satisfies the identity

$$(\partial_\mu + \overleftarrow{\partial}_\mu)\Gamma_\mu(\partial, -\overleftarrow{\partial}) = \Lambda(\partial) - \Lambda(-\overleftarrow{\partial}). \quad (2.5)$$

To see the meaning of this identity, consider the case of a Dirac field; then $\Gamma_\mu(\partial, -\overleftarrow{\partial})$reduces to γ_μ and the above identity is just the transcription into configuration space of the generalized Ward identity[2]

$$S_F^{-1}(p) - S_F^{-1}(q) = i(p - q)_\mu \gamma_\mu.$$

The identity (2.5) is the basis of our approach to the conservation laws. In the following sections, we exhibit the manner in which it may be used to derive conservation laws without recourse to Lagrangians or Noether's theorem.

3. CONSERVATION LAWS FOR FREE FIELDS

For free fields the equations of motion reduce to

$$\Lambda(\partial)\phi(x) = 0. \quad (3.1a)$$

Let us assume the existence of a nonsingular matrix η[10] such that

$$\eta\Lambda(\partial) = \Lambda^\dagger(-\partial)\eta^\dagger. \quad (3.2)$$

We can then define an adjoint field

$$\bar{\phi}(x) = \phi^\dagger(x)\eta \quad (3.3)$$

which satisfies

$$\bar{\phi}(x)\Lambda(-\overleftarrow{\partial}) = 0. \quad (3.1b)$$

We now suppose that the field equations (3.1a) and (3.1b) are invariant under the substitutions

$$\phi(x) \to F[x] \quad \bar{\phi}(x) \to G[x],$$

where $F[x]$ and $G[x]$ are some functionals of the field operators ϕ, $\bar{\phi}$, and their derivatives. This implies that

$$\Lambda(\partial)F[x] = 0, \quad (3.4a)$$

$$G[x]\Lambda(-\overleftarrow{\partial}) = 0. \quad (3.4b)$$

If we now sandwich the identity (2.5) between $G[x]$ on the left, and $F[x]$ on the right, and use Eqs. (3.4a, b), we get the conservation law

$$G(\partial_\mu + \overleftarrow{\partial}_\mu)\Gamma_\mu F = \partial_\mu(G\Gamma_\mu F) = G\Lambda(\partial)F - G\Lambda(-\overleftarrow{\partial})F = 0 \quad (3.5)$$

for the current

$$J_\mu(x) = G[x]\Gamma_\mu(\partial, -\overleftarrow{\partial})F[x]. \quad (3.6)$$

Let us apply this to some examples.

A. Symmetry under Infinitesimal Transformations

We assume that the field equation is invariant under some transformation

$$\phi(x) \to \phi'(x'). \quad (3.7)$$

This implies that

$$\Lambda(\partial)\phi'(x) = 0, \quad (3.8)$$

i.e., both $\phi(x)$ and $\phi'(x)$ satisfy the free equation (3.1a). Setting

$$\delta\phi(x) \equiv \phi'(x) - \phi(x), \quad (3.9)$$

where $\delta\phi$ is the local variation, we can take $F = \delta\phi$, $G = \bar{\phi}$. This gives the conserved current

$$J_\mu(x) = \bar{\phi}(x)\Gamma_\mu(\partial, -\overleftarrow{\partial})\,\delta\phi(x). \quad (3.10)$$

For example, if the field equation is invariant under an infinitesimal phase transformation

$$\delta\phi(x) = +i\alpha\phi(x), \quad \delta\bar{\phi}(x) = -i\alpha\bar{\phi}(x), \quad (3.11)$$

we get the conserved current

$$J_\mu(x) = \bar{\phi}(x)\Gamma_\mu(\partial, -\overleftarrow{\partial})\phi(x), \quad (3.12)$$

which, for a Dirac field, is just the usual $\bar{\psi}\gamma_\mu\psi$ and, for a complex scalar field, $\partial_\mu\phi^\dagger\phi - \phi^\dagger\partial_\mu\phi$.[11]

[9] The differential operator Γ_μ was first introduced in Ref. 1 as a basic tool for deriving conservation laws and for quantizing fields of arbitrary spin. Later, this operator was used in particular instances by various authors (Ref. 4) in their derivations of conservation laws.

[10] See Ref. 2, where Λ, Γ, and η are given explicitly for the following cases: (a) Klein–Gordon spin-0, (b) Dirac spin-$\frac{1}{2}$, Schwinger spin-$\frac{3}{2}$. Note that the assumption that η exists is actually equivalent to assuming the existence of a Lagrangian; we can in fact construct the Lagrangian $\mathcal{L} = -\bar{\phi}(x)\,\Lambda(\partial)\,\phi(x)$ to obtain (3.1a) and (3.1b) where $\bar{\phi}$ is given by (3.3). Our point is, however, that this Lagrangian need not be used *explicitly*. Our method, therefore, is applicable when neither the Lagrangian nor η exist, as in the Bethe–Salpeter case treated in Sec. 4.

[11] The conserved current (3.12) can, of course, be obtained by trivially setting $F = \phi$ and $G = \bar{\phi}$.

As another example, we consider the infinitesimal translation

$$\delta\phi(x) = -\alpha_\mu \partial_\mu \phi(x). \tag{3.13}$$

Then, as is easily verified, the invariance of the field equation gives the conserved quantity

$$T_{\mu\nu} = \tfrac{1}{2}\epsilon\{\bar{\phi}(x)\Gamma_\mu(\partial, -\overleftarrow{\partial})\partial_\nu\phi(x) - \partial_\nu\bar{\phi}(x)\Gamma_\mu(\partial, -\overleftarrow{\partial})\phi(x)\}, \tag{3.14}$$

where

$$\epsilon = \begin{cases} 1 & \text{for complex fields,} \\ \frac{1}{2} & \text{for real fields.} \end{cases}$$

In writing down (3.14) we have taken care to ensure Hermiticity; $T_{\mu\nu}$ as given by (3.14) is, to within a divergence term, the usual energy-momentum tensor of the field. In a similar manner, we can derive the conserved angular momentum density tensor by considering infinitesimal proper Lorentz transformations.

B. Discrete Symmetries

It is often thought that only *continuous* symmetry transformations can give rise to conserved currents. We see, however, that our method allows us to construct conserved currents corresponding to *discrete* symmetries as well. We illustrate our argument for the case of parity and charge conjugation transformations.

The invariance of the field equations under a parity transformation implies that there exists a matrix P such that [we use the matrix $x_\mu = (\mathbf{x}, it)$]

$$P\Lambda(-\partial^*) = \Lambda(\partial)P. \tag{3.15}$$

The transformed function $\phi'(x) = P\phi(-x^*)$ then satisfies Eq. (3.1a), i.e.,

$$\Lambda(\partial)P\phi(-x^*) = 0. \tag{3.16}$$

If we now take $G = \bar{\phi}$ and $F = P\phi(-x^*)$, we see that we get a conserved current

$$J_\mu(x) = \bar{\phi}(x)\Gamma_\mu(\partial, -\overleftarrow{\partial})P\phi(-x^*), \tag{3.17}$$

corresponding to the symmetry under space reflection. The space integral of the fourth component,

$$S = \int d\sigma_\mu(x)\, J_\mu(x) = \int d\sigma_\mu(x)\, \bar{\phi}(x)\Gamma_\mu(\partial, -\overleftarrow{\partial})P\phi(-x^*), \tag{3.18}$$

is therefore conserved in time. To see its meaning we consider the example of a Dirac field. Then $P = i\gamma_4$ and

$$S = i\int d\sigma_\mu(x)\, \bar{\psi}(x)\gamma_\mu\gamma_4\psi(-x^*). \tag{3.19}$$

If we now set

$$P \equiv e^{-i\frac{1}{2}\pi S}, \tag{3.20}$$

it is straightforward to show that P is the parity operator satisfying

$$P\psi(x)P^{-1} = i\gamma_4\psi(-x^*),$$
$$P\bar{\psi}(x)P^{-1} = -i\bar{\psi}(-x^*)\gamma_4.$$

As we can see above, any linear combination of $\psi(x)$ and $i\gamma_4\psi(-x^*)$ satisfies the free equation of motion. However, the combination which can be expressed as

$$e^{i\lambda S}\psi(x)e^{-i\lambda S} \tag{3.21}$$

is nonlocal in general. It is noticed that, for the special values $\lambda = 0$ and $\frac{1}{2}\pi$, the quantity (3.21) becomes $\psi(x)$ and $i\gamma_4\psi(-x^*)$, respectively, which are now local. The transformation for $\lambda = 0$ is 1 which is trivial, and that for $\lambda = \frac{1}{2}\pi$ is the parity transformation.

Similarly, the invariance of the field equations under charge conjugation implies the existence of a unitary matrix C with the property

$$[\eta\Lambda(\partial)]^T = \rho C^{-1}\eta\Lambda(-\partial)C,$$

where

$$\rho = \begin{cases} +1 & \text{for fields with integral spin,} \\ -1 & \text{for fields with half-odd-integral spin.} \end{cases}$$

Then the charge conjugate fields

$$\phi^c(x) = C\phi^*(x), \quad \overline{\phi^c} = \phi C^\dagger\eta$$

satisfy Eqs. (3.1a, b), respectively. Now, if we take

$$\bar{\phi}(x)\Gamma_\mu(\partial, -\overleftarrow{\partial})\phi^c(x),$$

we find that it actually vanishes identically, due to the spin-statistics relation.

However, we can take

$$J_\mu = i\{\bar{\phi}^{(+)}\Gamma_\mu\phi^{(-)c} + \phi^{(-)c}\Gamma_\mu\phi^{(+)}\}, \tag{3.22}$$

where $\phi^{(+)}$ and $\phi^{(-)}$ are the positive and negative frequency parts of ϕ, respectively. The space integral

$$R = \int d\sigma_\mu(x)\, J_\mu(x)$$

is therefore conserved in time. R is closely connected with the charge conjugation operator in Hilbert space. The charge conjugation operator is then

$$C = e^{i\frac{1}{2}\pi(R-T)}, \tag{3.23}$$

where

$$T = i \int d\sigma_\mu(x)[\bar{\phi}^{(+)}(x)\Gamma_\mu\phi^{(+)}(x) + \rho\bar{\phi}^{(-)}\Gamma_\mu\phi^{(-)}]$$

and commutes with R. It is straightforward to check that

$$\phi^c(x) = C\phi(x)C^{-1}.$$

We see therefore that conserved currents can be constructed for discrete transformations. This is a rather novel feature which results from our approach to the conservation laws.

C. Generalized Conservation Laws

Since the discovery of a conserved "zilch" for the electromagnetic field,[3] generalized conservation laws have been considered by several authors.[4] In particular, O'Connell and Tompkins,[4] restricting attention to fields satisfying the Bhabha equation,[12] showed that *all* generalized conservation laws as well as the usual conservation laws associated with the Bhabha field can be summarized in the form

$$\partial_\mu(\bar{\psi}'\alpha_\mu\psi'') = 0,$$

where $\bar{\psi}'$ and ψ'' must satisfy Bhabha's equations

$$(\alpha_\mu\,\partial_\mu + m)\psi'' = 0, \qquad \bar{\psi}'(-\alpha_\mu\,\overleftarrow{\partial}_\mu + m) = 0,$$

but are otherwise arbitrary.

It is now clear that the above result follows as a special case of our general procedure. In fact, we can immediately generalize O'Connell and Tompkin's result to fields of arbitrary spin given in Bhabha's form or otherwise. If we allow our functionals $F[x]$ and $G[x]$ satisfying (3.4a) and (3.4b) to be general higher-order tensors, we can form a current

$$J_{\mu\sigma_1\cdots\sigma_m,\rho_1\cdots\rho_n}(x) = G_{\sigma_1\cdots\sigma_m}[x]\Gamma_\mu(\partial, -\overleftarrow{\partial})F_{\rho_1\cdots\rho_n}[x] \tag{3.24}$$

satisfying

$$\partial_\mu J_{\mu\sigma_1\cdots\sigma_m,\rho_1\cdots\rho_n}(x) = 0 \tag{3.25}$$

due to the identity (2.5). For example, we can choose

$$F_{\rho_1\cdots\rho_n}[x] = \partial_{\rho_1}\cdots\partial_{\rho_n}\phi(x),$$
$$G_{\sigma_1\cdots\sigma_m}[x] = \partial_{\sigma_1}\cdots\partial_{\sigma_m}\bar{\phi}(x). \tag{3.26}$$

The two lowest members of the series are just the charge current density (3.12) for $n = m = 0$, and the energy-momentum density (3.15) for $n + m = 1$. It should be borne in mind, however, that generalized conserved currents will not always have a direct physical meaning. For instance, we have

$$\Lambda(\partial)\phi_{\text{in}}(x) = 0,$$
$$\bar{\phi}_{\text{out}}(x)\Lambda(-\overleftarrow{\partial}) = 0, \tag{3.27}$$

with $\phi_{\text{out}}(x)$ a complicated functional of $\phi_{\text{in}}(x)$. The "current"

$$J_\mu(x) = \bar{\phi}_{\text{out}}(x)\Gamma_\mu(\partial, -\overleftarrow{\partial})\phi_{\text{in}}(x) \tag{3.28}$$

is therefore conserved but has no direct physical significance.

Let us recover Lipkin's "zilch" for the free electromagnetic field by means of our method. It is convenient to take as independent field variables the $F_{\mu\nu}$ which satisfies

$$\Box F_{\mu\nu}(x) = 0. \tag{3.29a}$$

The dual tensor $\hat{F}_{\mu\nu} \equiv \frac{1}{2}\epsilon_{\mu\nu\sigma\rho}F_{\sigma\rho}$ also satisfies (3.29a) so that

$$\hat{F}_{\mu\nu}(x)\overleftarrow{\Box} = 0. \tag{3.29b}$$

We can therefore build the conserved current

$$J_{\mu,\alpha\beta} = \hat{F}_{\alpha\sigma}\Gamma_{\mu,\sigma\rho}F_{\rho\beta}$$
$$= \hat{F}_{\alpha\lambda}(\partial_\mu - \overleftarrow{\partial}_\mu)F_{\lambda\beta}. \tag{3.30}$$

Lipkin's zilch tensor $Z_{\mu,\alpha\beta}$, symmetric in α and β is obtained by taking[13]

$$Z_{\mu,\alpha\beta} = J_{\mu,\alpha\beta} + J_{\mu,\beta\alpha}$$
$$= \hat{F}_{\alpha\lambda}(\partial_\mu - \overleftarrow{\partial}_\mu)F_{\lambda\beta} + \hat{F}_{\beta\lambda}(\partial_\mu - \overleftarrow{\partial}_\mu)F_{\lambda\alpha}. \tag{3.31}$$

4. CONSERVATION LAWS FOR INTERACTING FIELDS

In this section, we exhibit two ways in which the usefulness of the identity (2.5) may be extended to yield conservation laws for interacting fields.

(1) Referring back to the general derivation of the conservation law (3.5), it may happen that certain functionals $G[x]$ and $F[x]$ do not satisfy the free equations (3.4a, b) but that, nevertheless, the difference

$$G[x]\Lambda(\partial)F[x] - G[x]\Lambda(-\overleftarrow{\partial})F[x] \tag{4.1}$$

vanishes; this weaker condition is sufficient to establish the conservation law (3.5) for $J_\mu = G\Gamma_\mu F$.

(2) If the difference (4.1) does not vanish but happens to be equal to a divergence term

$$\partial_\nu K_\nu[x] \tag{4.2}$$

[12] J. Bhabha, Rev. Mod. Phys. **17**, 300 (1945); *ibid.* **21**, 451 (1949).

[13] See T. W. B. Kibble, in Ref. 4, especially Eq. (9).

with $K_\nu \neq G\Gamma_\nu F$, then we still get a conserved current by taking

$$J'_\mu = J_\mu - K_\mu = G\Gamma_\mu F - K_\mu. \tag{4.3}$$

These two possibilities are illustrated by the following examples. As an instance where the difference (4.1) vanishes, consider a Dirac field ψ in interaction with an electromagnetic field

$$\Lambda(\partial)\psi = (\gamma\cdot\partial + m)\psi = ie\gamma_\mu A_\mu\psi,$$
$$\bar\psi\Lambda(-\overleftarrow{\partial}) = \bar\psi(-\gamma\cdot\overleftarrow{\partial} + m) = ie\bar\psi\gamma_\mu A_\mu. \tag{4.4}$$

Setting $G = \bar\psi$ and $F = \psi$, we obtain

$$\bar\psi(\gamma\cdot\partial + m)\psi - \bar\psi(-\gamma\cdot\overleftarrow{\partial} + m)\psi = \partial_\mu(\bar\psi\gamma_\mu\psi)$$
$$= ie\bar\psi\gamma_\mu\psi A_\mu - ie\bar\psi\gamma_\mu\psi A_\mu = 0. \tag{4.5}$$

We thus recover the usual result that the current

$$J_\mu = \bar\psi\gamma_\mu\psi$$

is conserved even in the presence of the interaction.

As an example of the second possibility above, we consider the interaction of charged spin-0 mesons with an external electromagnetic field, $A_\mu(x)$. The equations of motion are (with $\partial_\mu A_\mu = 0$)

$$\Lambda(\partial)\phi = (\Box - \mu^2)\phi = 2ieA_\mu\partial_\mu\phi + e^2A_\mu A_\mu\phi,$$
$$\bar\phi\Lambda(-\overleftarrow{\partial}) = \bar\phi(\overleftarrow{\Box} - \mu^2) = -2ie\bar\phi\overleftarrow{\partial}_\mu A_\mu + e^2\bar\phi A_\mu A_\mu. \tag{4.6}$$

Evaluating the difference (4.1)

$$\partial_\mu[\bar\phi(\partial_\mu - \overleftarrow{\partial}_\mu)\phi] = \bar\phi(\Box - \mu^2)\phi - \bar\phi(\overleftarrow{\Box} - \mu^2)\phi$$
$$= 2ie\bar\phi(\partial_\mu + \overleftarrow{\partial}_\mu)\phi\cdot A_\mu = 2ie\,\partial_\mu[\bar\phi\phi A_\mu],$$

we obtain the conserved current

$$J'_\mu = ie[\bar\phi(\partial_\mu - \overleftarrow{\partial}_\mu)\phi - 2ie\bar\phi\phi A_\mu]. \tag{4.7}$$

As another example, we consider the interaction of a Dirac field with a neutral pseudo-scalar meson field. The equations of motion are

$$(\gamma\cdot\partial + m)\psi = ig\gamma_5\psi\phi,$$
$$\bar\psi(-\gamma\cdot\overleftarrow{\partial} + m) = ig\phi\bar\psi\gamma_5,$$
$$(\Box - \mu^2)\phi = -ig\bar\psi\gamma_5\psi. \tag{4.8}$$

Setting $G = \bar\psi$, $F = \partial_\nu\psi$, and using (4.8), we get

$$\partial_\mu[\bar\psi\gamma_\mu\,\partial_\nu\psi]$$
$$= \bar\psi(\gamma\cdot\partial + m)\,\partial_\nu\psi - \bar\psi(-\gamma\cdot\overleftarrow{\partial} + m)\,\partial_\nu\psi$$
$$= ig\bar\psi\gamma_5\psi\,\partial_\nu\phi = -(\Box - \mu^2)\phi\cdot\partial_\nu\phi$$
$$= \partial_\mu[-\partial_\mu\phi\,\partial_\nu\phi + \tfrac{1}{2}\delta_{\mu\nu}(\partial_\lambda\phi\,\partial_\lambda\phi + \mu^2\phi\phi)], \tag{4.9}$$

so that we obtain the conservation of the energy–momentum tensor

$$T_{\mu\nu} = \bar\psi\gamma_\mu\,\partial_\nu\psi + \partial_\mu\phi\,\partial_\nu\phi - \tfrac{1}{2}(\partial_\lambda\phi\,\partial_\lambda\phi + \mu^2\phi\phi). \tag{4.10}$$

The last two terms are, of course, the energy–momentum tensor of the meson field. In fact, on the basis of this simple example, we see that, for interacting fields, the second possibility is much more likely to arise than the first, since the various interacting particles will generally each contribute a piece to the conserved current.

In summary, whenever the difference (4.1) vanishes as a consequence of the field equations, we can take the conserved current for free fields over to the case of interacting fields. If, on the other hand, the difference (4.1) turns out to be a nonzero divergence term, the conserved current for interacting fields differs from its free-field value by $-K_\mu[x]$. We stress that we do not require the *explicit* use of a Lagrangian.

It is worth making one final remark. When dealing with interacting fields, it does seem necessary to assume the existence of a Lagrangian if one wishes to derive the conservation laws from invariance principles. Although we cannot prove this conjecture in general, we do observe that all the examples treated in this section have this property and that all are based on Lagrangians. On the other hand, simple equations can be exhibited[14] which are translation invariant, but fail to give rise to a corresponding conservation law; these equations have no Lagrangian formulation.

5. APPLICATION TO BETHE–SALPETER WAVEFUNCTIONS

An interesting application of our method can be made to the theory of Bethe–Salpeter wavefunctions.[7,8] Let us consider as an illustration the Bethe–Salpeter wavefunctions

$$\chi_P(x_1, x_2) = \langle 0|\, T(\psi_A(x_1)\psi_B(x_2))\,|P\rangle, \tag{5.1a}$$
$$\bar\chi_P(x_1, x_2) = \langle P|\, T(\bar\psi_A(x_1)\bar\psi_B(x_2))\,|0\rangle, \tag{5.1b}$$

for a bound state of two fermions A and B. P is the center-of-mass momentum of the pair. As usual, we have

$$\chi_P(x_1, x_2) \sim e^{iPX}\chi_P(x), \tag{5.2a}$$
$$\bar\chi_P(x_1, x_2) \sim e^{-iPX}\bar\chi_P(x), \tag{5.2b}$$

where X is the center-of-mass coordinate $\frac{1}{2}(x_1 + x_2)$ and x the relative coordinate $x_1 - x_2$. Going over

[14] S. Kamefuchi and Y. Takahashi, Nuovo Cimento 44, 1 (1966).

to the Fourier transforms $\chi_P(p)$ and $\bar{\chi}_P(q)$, we write the Bethe–Salpeter equations for χ and $\bar{\chi}$ in the operator form[15]

$$[I(P) + G(P)]\chi_P = 0, \tag{5.3a}$$

$$\bar{\chi}_P[I(P) + G(P)] = 0. \tag{5.3b}$$

Let us work with the configuration-space equations, where P is replaced by $-i\partial/\partial X$:

$$[I(-i\,\partial/\partial X) + G(-i\,\partial/\partial X)]\Phi(X) = 0, \tag{5.4a}$$

$$\bar{\Phi}(X)[I(i\,\overleftarrow{\partial}/\partial X) + G(i\,\overleftarrow{\partial}/\partial X)] = 0, \tag{5.4b}$$

which are formally similar to (3.1a,b). If we set $\Phi \sim e^{iPX}\chi_P$ and $\bar{\Phi} \sim e^{-iPX}\bar{\chi}_P$, these equations reduce to (5.3a,b). However, Eqs. (5.4a,b) are more general in that the antibound-state wavefunction also satisfies (5.4a,b). This is, of course, entirely analogous to the corresponding situation for, say, Dirac single-particle wavefunctions.

We now proceed as in the elementary-particle case. Expanding $I + G$ in powers of $\partial/\partial X$, we define Γ_μ by Eq. (2.4). We thus obtain the identity

$$I\left(-i\frac{\partial}{\partial X}\right) + G\left(-i\frac{\partial}{\partial X}\right) - I\left(i\frac{\overleftarrow{\partial}}{\partial X}\right) - G\left(i\frac{\overleftarrow{\partial}}{\partial X}\right)$$
$$= \left(\frac{\partial}{\partial X_\mu} + \frac{\overleftarrow{\partial}}{\partial X_\mu}\right)\Gamma_\mu\left(-i\frac{\partial}{\partial X}, i\frac{\overleftarrow{\partial}}{\partial X}\right). \tag{5.5}$$

Sandwiching this identity between $\bar{\Phi}(x)$ and $\Phi(x)$ and using (5.2), we get the conservation law

$$\partial_\mu J_\mu(X) = \frac{\partial}{\partial X_\mu}\left\{\bar{\Phi}(X)\Gamma_\mu\left(-i\frac{\partial}{\partial X}, i\frac{\overleftarrow{\partial}}{\partial X}\right)\Phi(X)\right\} = 0. \tag{5.6}$$

The space integral of the fourth component

$$-i\int J_4(X)\,d^3X$$
$$= -i\int d^3X\,\bar{\Phi}(X)\Gamma_4(-i\,\partial/\partial X, i\,\overleftarrow{\partial}/\partial X)\Phi(X) \tag{5.7}$$

is therefore conserved in time.

To see the physical meaning of this conserved quantity we set

$$\Phi \sim e^{iPX}\chi_P \quad \text{and} \quad \bar{\Phi} = e^{-iPX}\bar{\chi}_P.$$

Then

$$\Gamma_\mu(-i(\partial/\partial X), i(\overleftarrow{\partial}/\partial X)) \to \Gamma_\mu(P, P) \tag{5.8}$$

when sandwiched between $e^{iPX}\bar{\chi}_P$ and $e^{iPX}\chi_P$, and (5.5) goes over into the differential form of the generalized Ward identity:

$$(\partial/\partial P_\mu)[I(P) + G(P)] = i\Gamma_\mu(P, P). \tag{5.9}$$

Hence (5.7) reduces to

$$-\bar{\chi}_P(\partial/\partial P_0)[I(P) + G(P)]\chi_P, \tag{5.10}$$

which we recognize to be the usual normalization integral for Bethe–Salpeter wavefunctions.[5,6] In other words, if one could show that the integral on the right-hand side of (5.7) is real, one could use (5.7) to give a direct proof of the normalization condition for Bathe–Salpeter wavefunctions, using the technique of Ref. 1. Unfortunately, we have been unable to establish the reality of (5.7) directly, (although *a posteriori*, of course, we know it to be real as a result of the work of Refs. 5 and 6). Nevertheless, our approach here does serve to underline the very close correspondence which exists between Bethe–Salpeter and elementary-particle wavefunctions.[16]

ACKNOWLEDGMENT

One of the authors (Y.T.) thanks Professor Ramakrishnan of Matscience for the kind hospitality extended to him while in Madras.

[15] We use the notation of Ref. 6. I is essentially the product $S_F^{-1}\,S_F^{-1}$ of the inverse propagators for the two particles which bind together to form the bound state; G is the interaction function corresponding to the sum of all Bethe–Salpeter irreducible graphs.

[16] This has also been stressed in Ref. 6.

Lattice Statistics of Hydrogen Bonded Crystals. I. The Residual Entropy of Ice*

J. F. NAGLE†
Yale University, New Haven, Connecticut
(Received 26 May 1965)

The lattice statistical problem of calculating the residual entropy of ice has been considered in some detail for the hexagonal and cubic ice lattices as well as for a two-dimensional icelike lattice. Even for the two-dimensional lattice, this problem appears to be intractable using exact methods, so an approximation method is in order. The series method of DiMarzio and Stillinger has been developed so that the series is completely characterized by the numbers of various kinds of cycles on the lattice. The first five terms of the series have been evaluated and used to extrapolate values of the residual entropy $S(0)$ within rather narrow limits for all practical purposes. The result for hexagonal ice and cubic ice is $S(0) = .8145 \pm .0002$ cal/deg/mole which agrees with experiment even better than Pauling's original approximation. Some other methods are also discussed, and their results tend to confirm the series results.

1. INTRODUCTION

THE problem to be discussed in this paper is: Given any regular undirected graph (such as a crystal lattice with the atoms as vertices and the bonds as edges) with N vertices and with coordination number equal to four, compute the number W_N of ways to make the graph into a directed graph (that is, each edge is assigned a direction) such that each vertex is the terminus of exactly two directed edges. In particular, we are interested in computing

$$W \equiv \lim_{N\to\infty} (W_N)^{1/N}, \tag{1}$$

where $N \to \infty$ means that the graph becomes infinite in extent in all possible directions. Figure 1 shows one configuration of directed edges satisfying the conditions of the problem for the lattice graph consisting of a finite square net.

This problem arises in connection with the theoretical explanation of the residual entropy of ice at low temperatures.[1-3] Each oxygen in the ice crystal, which has a wurtzite hexagonal structure, is hydrogen bonded to four other oxygens. The hydrogens sit off center on the bond and two hydrogens sit close to each oxygen. Apparently, all such configurations of the hydrogens have nearly equal energies and the crystal remains disordered at low temperatures.[1] We may describe the oxygen–hydrogen arrangement in the crystal in terms of a directed graph, so that O—H——O becomes •——◄——•. The rule that two hydrogens sit close to each oxygen, which we refer to as the ice rule, is just the condition that each vertex is the terminus of two directed edges. Therefore, assuming that all arrow configurations obeying the ice rule have the same energy, we have

$$S(0) = Nk \log W_N^{1/N} = Nk \log W, \tag{2}$$

where $S(0)$ is the residual entropy of ice.

Pauling first estimated that $W = \frac{3}{2}$ using a zero-order approximation equivalent to the mean field or random mixing approximations.[2] Onsager showed that Pauling's result gave a lower bound for W.[1] Takahasi,[4] and recently DiMarzio and Stillinger,[5] discussed series approximations for W and derived the first two terms by somewhat laborious and not very general methods while leaving undiscussed the character of the general term. In Sec. 2 we review

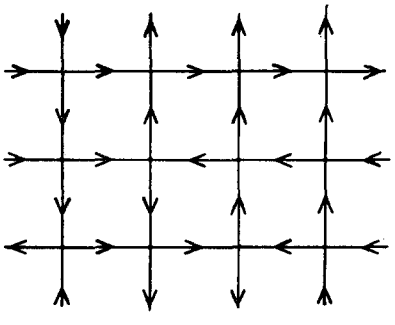


FIG. 1. One possible configuration of arrows.

* This paper is based on Part I of a dissertation submitted to Yale University in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

† Present address: Department of Chemistry, Cornell University, Ithaca, New York.

[1] L. Onsager and M. Dupuis, Rend. Scuola Intern. FS., X Corso, 294 (1960).

[2] L. Pauling, J. Am. Chem. Soc. 57, 2680 (1935). For a more recent review see L. Pauling, *The Nature of the Chemical Bond* (Cornell University Press, Ithaca, New York, 1960), 3rd ed.

[3] W. F. Giauque and J. W. Stout, J. Am. Chem. Soc. 58, 1144 (1936).

[4] H. Takahasi, Proc. Phys. Math. Soc. (Japan) 23, 1069 (1941).

[5] E. A. DiMarzio and F. H. Stillinger, Jr., J. Chem. Phys. 40, 1577 (1964).

the DiMarzio and Stillinger transformation of the expression for W_N. In Sec. 3 we prove a characterization of the general term in the series. In Sec. 4 we describe the procedures used to evaluate the terms in the series and give the results for the first five terms for the infinite square net, the infinite diamond lattice, and the infinite hexagonal ice lattice. In Sec. 5 these series are extrapolated to obtain estimates for W. In Sec. 6 we briefly describe other methods which give estimates for W.

2. TRANSFORMATION OF W_N

We conform to the notation of DiMarzio and Stillinger.[5] Let us designate the arrangement of

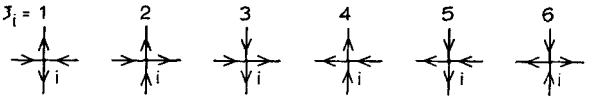
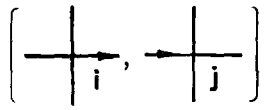
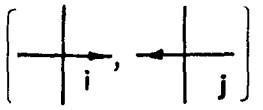


FIG. 2. The vertex configurations ξ_i of the ith vertex.

arrows incident to vertex i by ξ_i where ξ may take on six values corresponding to the six arrangements compatible with the ice rule (see Fig. 2). For two neighboring vertices i and j, we define a compatibility function,

$$A(\xi_i, \xi_j) = \begin{bmatrix} 1, \text{ if } \xi_i \text{ and } \xi_j \text{ are compatible} \\ 0, \text{ if } \xi_i \text{ and } \xi_j \text{ are incompatible} \end{bmatrix}.$$

For  in square ice, A  = 1 and A  = 0,

where the unspecified arrows may take on any directions compatible with the ice rule at i and j.

Now, it is clear that

$$W_N = \sum_{[\xi]} \prod_{i<j} A(\xi_i, \xi_j), \tag{3}$$

where $\sum_{[\xi]}$ means to sum over the set of all 6^N different combinations of ξ arrangements at each vertex i, and $\prod_{i<j}$ is the product over nearest neighbors with each pair taken once.

We next define a new compatibility function.

$$a(\xi_i, \xi_j) = \begin{bmatrix} +1, \text{ if } \xi_i \text{ and } \xi_j \text{ are compatible} \\ -1, \text{ if } \xi_i \text{ and } \xi_j \text{ are not compatible} \end{bmatrix}.$$

Then, it is an identity to write

$$W_N = (3/2)^N \sum_{[\xi]} \prod_{i<j} (1/6)[1 + a(\xi_i, \xi_j)]. \tag{4}$$

The series expansion will involve no a's in the zeroth term, one a in the first term, and so on.

It will be convenient to associate products of a's with subgraphs drawn on the ice lattice. For example,

$$\sum_{[\xi]} a(\xi_2, \xi_6)a(\xi_6, \xi_7)a(\xi_6, \xi_{10})a(\xi_7, \xi_{11}) \cdot a(\xi_{10}, \xi_{11})a(\xi_{11}, \xi_{12})a(\xi_{11}, \xi_{15})$$

would correspond to the subgraph of square ice illustrated in Fig. 3. Henceforth, we refer to the above sum as the contribution from the graph shown in Fig. 3. Now, it is possible to eliminate many graphs from our consideration because of the following two identities:

$$\sum_{\xi_i} a(\xi_i, \xi_j) = 0, \qquad \sum_{\xi_i} a(\xi_i, \xi_j)a(\xi_i, \xi_k)a(\xi_i, \xi_m) = 0. \tag{5}$$

These imply that any graph with any vertices with either one or three incident edges (such as vertices 2 and 6 in Fig. 3) gives a zero contribution. The remaining graphs have vertices with either two or four incident edges. Connected graphs with the property of having an even number of edges incident to each vertex can be traced in such a way that one traces each edge once and only once without lifting one's pen from the paper and one ends where one began. Such graphs are called simple or Eulerian cycles in the notation of graph theory.[6]

3. EVALUATION OF CYCLE WEIGHTS

The contribution of a cycle will be called its weight. DiMarzio and Stillinger calculated the weights of some of the smaller cycles for square ice and hexagonal ice.[5] Their method involved explicit matrices for the a's. Not only was the method cumbersome but it was necessary to repeat it for

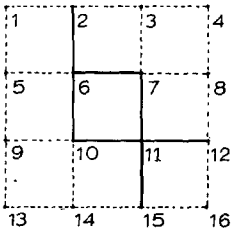


FIG. 3. A subgraph of the square lattice.

[6] C. Berge, *The Theory of Graphs and Its Applications*, translated by Alison Doig (John Wiley & Sons, Inc., New York, 1962).

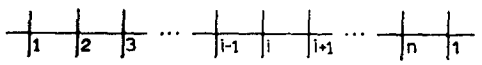


FIG. 4. A deformed elementary n cycle.

each different kind of ice lattice. Also, DiMarzio and Stillinger did not calculate a general expression for the cycle weights. In this section, we evaluate the weight of any cycle for any ice lattice in terms of its numbers of edges and vertices.

It should be mentioned that the term, cycle, includes subgraphs in which some edges must be retraced before returning to the origin; such cycles are called composite cycles; it follows from the last section that they have zero weight. To complete our cycle terminology we subdivide the class of Eulerian cycles. Those with only two edges incident to each vertex are called elementary cycles. For those Eulerian cycles with four edges incident to some vertices we introduce the term, crossover cycles.

We may deform an elementary n cycle as shown in Fig. 4.

It is elementary to show that

$$\sum_{\xi_2} a(\xi_1, \xi_2)a(\xi_2, \xi_3)$$

$$= \left[\begin{array}{llll} +2, & \text{if} & \text{1} & \text{and } \text{3} \\ +2, & \text{if} & \text{1} & \text{and } \text{3} \\ -2, & \text{if} & \text{1} & \text{and } \text{3} \\ -2, & \text{if} & \text{1} & \text{and } \text{3} \end{array}\right],$$

where by  we mean any one of the three possibilities, $\xi_1 = 2$, 3, or 6. Notice that the sign is + or − according as the designated arrows on vertices 1 and 3 are in the same or opposite directions. Proceeding by induction it can easily be shown that

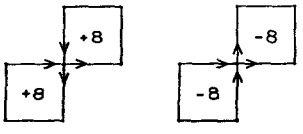


FIG. 5. The two cases for figure eight crossovers.

$$\sum_{\xi_2,\cdots,\xi_{i-1}} a(\xi_1, \xi_2) \cdots a(\xi_{i-1}, \xi_i)$$

$$= \left[\begin{array}{llll} +2^{i-2} & \text{if} & \text{1} & \text{and } \text{i} \\ +2^{i-2} & \text{if} & \text{1} & \text{and } \text{i} \\ -2^{i-2} & \text{if} & \text{1} & \text{and } \text{i} \\ -2^{i-2} & \text{if} & \text{1} & \text{and } \text{i} \end{array}\right].$$

Finally, we take $i = n + 1 = 1$. Then, by summing over ξ_1, we get

$$\sum_{\xi_1,\cdots,\xi_n} a(\xi_1, \xi_2) \cdots a(\xi_n, \xi_1) = (4 - 2)2^{n-1} = 2^n.$$

We must take into account the $(1/6)^{\frac{1}{2}}$ factors. These always give $(1/6)^m$, where m is the number of vertices in the cycle. Hence, the weight for elementary cycles with $m = n$ vertices and n edges is $(2/6)^n = (1/3)^n$.

Crossover cycles present an additional problem which, however, is easily solved once we see what to look for. Let us illustrate with a simple figure eight crossover. We first specify the arrangement of the arrows incident to the crossover vertex and then sum over the noncrossover vertex arrangements. We know form the last paragraph that the noncrossover vertices give factors ± 2. In Fig. 5 are shown representatives of the two important cases. Each loop gives a factor ± 8, but the product of the two loop factors is always + 64. Since all six of the arrangements of the crossover vertex give the same result, we get $(6 \times 64)/6^7 = (1/3)^6$. Hence, a figure eight crossover cycle with eight edges, six ordinary vertices, and one crossover vertex has the same weight as an elementary cycle with 6 edges. This leads one to suspect that crossover cycles have weight $(1/3)^{m-c} = (1/3)^{n-2c}$, where c is the number of crossover vertices and $m - c$ is the number of ordinary vertices.

The general computation of crossover cycle weights follows the same line as the computation of figure eight crossover cycle weights. We specify an arbitrary arrangement of arrows incident to the crossover vertices and sum over all the noncrossover vertex arrangements. It suffices to show that we get $+2^{m-c}$ rather than -2^{m-c}. Consider the graph

in Fig. 6 and its reduced graph. Each edge in the reduced graph has two arrows on it, one from the crossover vertex at either end. We label an edge S if the arrows point in the same direction and D if they point in opposite directions. We know that each S edge yields a positive factor and each D edge yields a negative factor. We must prove that there is an even number of D edges. We first subdivide the D edges into two classes. ⟵⟶ is a D_1 edge, and ⟶⟵ is a D_2 edge. Now, a necessary condition for the ice rule to hold at each crossover vertex is that the number of arrows pointing into the set of crossover vertices equals the number pointing out. Each S edge in the reduced graph contributes one arrow which points into the set of crossover vertices and one which points out. A D_1 edge contributes two arrows into the set of crossover vertices, and a D_2 edge contributes two arrows out. Obviously, the number of D_1 edges must equal the number of D_2 edges in order to satisfy the necessary condition for the ice rule to hold at each crossover vertex. Therefore, the total number of D edges must be even. Therefore, the cycle weight for crossover cycles and ipso facto for all Eulerian cycles is

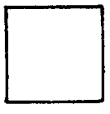
$$2^{m-c}6^c/6^m = (1/3)^{m-c} = (1/3)^{n-2c},$$

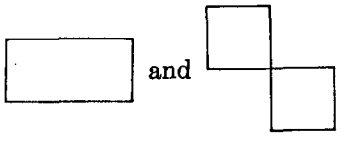
where n is the number of edges, c is the number of crossover vertices, and m is the total number of vertices.

4. EVALUATION OF COEFFICIENTS IN THE SERIES

It is now possible to begin to write the series. For example, for square ice we have

$$W_N = (3/2)^N\left(1 + \frac{N}{3^4} + \frac{4N}{3^6} + \frac{22N + (1/2)N(N-9)}{3^8} + \cdots\right). \quad (6)$$

The zeroth term corresponds to no cycles on the lattice. The first term, $N/3^4$, arises from square cycles,  . In the limit of large N there are N of these. Better yet, if we impose periodic boundary conditions, there are precisely N of these. The second term corresponds to four space types of cycles,



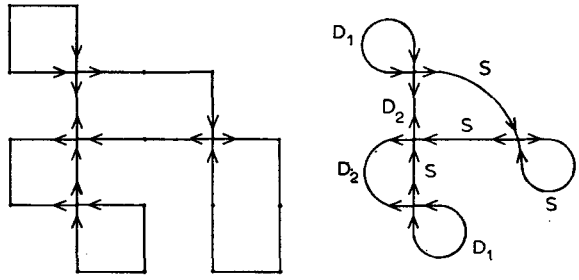


FIG. 6. A crossover cycle and its reduced graph.

Each space type occurs $2N$ times on the lattice. The $22N/3^8$ part of the third term arises from larger cycles. The $(1/2)N\ (N-9)/3^8$ part comes from two unconnected squares. The first one can be chosen in N ways, but the second can not touch the first one; hence, it can occupy only $N-9$ positions. The $\frac{1}{2}$ factor is due to the indistinguishability of the two squares. Next, we remember that we really want $W = (W_N)^{1/N}$, not W_N. To get W from W_N formally for a lattice with periodic boundary conditions, one simply replaces N wherever it appears in W_N by 1.[7] Hence, for square ice,

$$W = (3/2)\left(1 + \frac{1}{3^4} + \frac{4}{3^6} + \frac{22-4}{3^8} + \cdots\right). \quad (7)$$

Before proceeding further some observations are in order. One notices a resemblance between this series and the Ising model[8] high-temperature series when the temperature—exchange ratio, K satisfies $\tanh K = \frac{1}{3}$. However, in the ice series crossover cycles are weighted more heavily than in the Ising series. Consequently, approximation using the Ising series are too low. In connection with this, it might be mentioned that various techniques for solving the two dimensional Ising model have been of no avail in solving the two-dimensional ice problem. In particular, Stillinger found a dimerization of the ice problem which almost worked, but not quite, and investigation of the matrix method has failed to disclose any information, except to give the successive approximations described in Sec. 6. Therefore, it seems necessary to use series methods to get any reliable information at all, especially since we are interested mostly in ice in three dimensions, for which dimensionality exact solutions to problems seem very hard to come by.

The results of counting Eulerian cycles on the square net, the diamond lattice (cubic ice), and the

[7] C. Domb, Advan. Phys. 9, 149 (1960), Sec. 3.6.1, Eq. (144).

[8] For example, see G. F. Newell and E. W. Montroll, Rev. Mod. Phys. 25, 353 (1953), or the review given in Ref. 7.

TABLE I. Summary table of series expansion for square ice.[a]

n	0c	1c	2c	3c	4c	5c	≥6c	u.c.	ϕ_n	ϕ_{n+2}/ϕ_n
4	1								1	4
6	2	2							4	4.5
8	7	8	6	0	1			−4	18	5 1/9
10	28	40	36	22	4	8	2	−48	92	5.630
12	124	208	220	184	103	48	89	−458	518	

[a] The notation is explained in the text.

hexagonal ice lattice (ordinary ice) are given in Tables I, II, and III, respectively. To explain the notation we note that

$$W = (3/2)[1 + \sum_n \phi_n(1/3)^n]. \tag{8}$$

The column headings give the number of crossover vertices in the cycles of nth order and u.c. abbreviates unconnected cycles.

The procedure used to find the entries in Table I for square ice was to simply draw the various cycles on paper. Of course, this is especially liable to small errors in the highest terms.

For cubic ice one may also proceed by simply listing the cycles and this has been done for $n = 6, 8, 10$, and 12. However, another method developed by Domb and Fisher[9] for the high-temperature series for the Ising model problem allows us to compute the $n = 14$ term in addition to the lower terms. There are two basic formulas. For cubic ice and even n, they are

$$q_n = 2 + r_n - 3nr_{n-2} + \frac{3^2}{2!}n(n-3)r_{n-4} - \frac{3^3}{3!}n(n-4)(n-5)r_{n-6} + \frac{3^4}{4!}n(n-5)(n-6)(n-7)r_{n-8} \pm \cdots \tag{9}$$

and

$$r_n = [(n/2)!]^2 \sum_{s=0}^{n/2} \frac{(2s)!\,(n-2s)!}{(s!)^4[(n/2-s)!]^4}, \tag{10}$$

where r_n is the number of returns to the origin in a random walk of n steps, and q_n is the number of returns to the origin in a no immediate reversal random walk of n steps. Also, there are no "tadpoles" in q_n. A tadpole is a composite cycle in which the first and last steps are in opposite directions. Every space type of elementary cycle of weight $(1/3)^n$ will be counted $2n$ times in q_n, because these cycles may be walked in two directions and there are n possible starting points. The first complication is the crossover cycles. Single crossovers are counted $4n$ times in q_n, because each cycle may be walked four ways starting at any of the n-2 ordinary sites and eight ways starting from the crossover site. All these must be subtracted from q_n and the single crossovers must be promoted to weight $(1/3)^{n-2}$. Thus far, the only way devised to count the crossovers is to look at the lattice and count them visually. Fortunately, the crossovers are not as important in the coefficients of low order in the cubic ice series as they are for square ice. Otherwise, this method would be of no more use than it is for square ice, which, because of the abundance of crossovers, is more easily handled using the completely visual approach. The second complication is that some composite cycles can be walked which have zero weight in the ice expansion. These also must be subtracted from q_n. The total of these subtractions is listed in Table II under the column heading, sub. terms.

The procedure for hexagonal ice is the same as the one for cubic ice. First, r_n, the number of returns in walks of n steps, must be computed. Now, it is pointed out in Ref. 7 (p. 316) that there exist proofs that r_n is the same for the hexagonal close packed lattice and the face centered cubic lattice. These are respectively the lattices which one derives from the hexagonal ice lattice and the cubic ice (diamond) lattice by suppressing one of the next nearest neighbor sublattices. (By this same operation one derives the triangular lattice from the plane hexagonal lattice.) From this it can be shown fairly easily that r_n is the same for the cubic and hexagonal ice lattices. However, since this proof does not exist in the literature, we have sketched a direct proof in the Appendix that r_n is the same for cubic and hexagonal ice.[10]

One notices that the entries in Table III are nearly equal to the entries in Table II, but that

[9] C. Domb and M. E. Fisher, Proc. Cambridge Phil. Soc. 54, 48 (1958). For tables and explicit formulas for r_n see Ref. 7, Appendix II to Sec. 5. In this reference the formula for q_n is given in Sec. 5.2.3.

[10] I wish to thank M. F. Sykes and M. E. Fisher for informing me of the results in this paragraph and Professor Fisher for outlining the proof in the Appendix.

TABLE II. Summary table for cubic ice series.[a]

n	$q_n/2n$	sub. terms	0c	1c	2c	3c	all c	u.c.	ϕ_n	ϕ_{n+2}/ϕ_n
6	2		2				2		2	1.5
8	3		3				3		3	12.0
10	24		24	12			36		36	3.083
12	143	49	94	48	4		146	−35	111	9.622
14	918	336	582	456	156	12	1206	−138	1068	

[a] The notation is explained in the text.

it seems that W_{hex} is slightly larger than W_{cubic}. This is in agreement with a proof given by L. Onsager that $W_{\text{hex}} \geq W_{\text{cubic}}$.[11]

5. ANALYSIS OF THE SERIES

We wish to use the series expansions calculated in the last section to compute estimates for W. Obviously, we can just truncate the series after the first few terms. However, by making some straightforward guesses as to the behavior of the entire series from its first few terms, one is able to get reasonable estimates for the remainder of the series. Of course, one does not know whether one's guess as to the behavior of the entire series is correct. However, for square ice a straightforward guess of the series behavior will yield a value of W which is in good agreement with the independent estimate made in Sec. 6. It is plausible that a straightforward guess for the series behavior of cubic or hexagonal ice will also yield a reasonable estimate of W.

We turn to square ice. Adding up the terms which we have computed yields

$$W = (3/2)(1 + .02311) = 1.5347.$$

To analyze the remainder of the series, we employ a technique which has proved useful for the analogous Ising model series.[12] In Fig. 7 we have plotted ϕ_{n+2}/ϕ_n versus $1/n$. It can be seen that this gives a fairly smooth curve. By straightforwardly extrapolating this curve, we can compute estimates for ϕ_n, for $n > 12$, which gives

$$1.5387 = (1.5)(1 + .02311 + .00216 + .00050) < W$$

and

$$W < (1.5)(1 + .02311 + .00216 + .00206) = 1.5410,$$

where the .02311 comes from the terms $n = 4$–12; the .00216 comes from the extrapolated values for the terms $n = 14$–26; .00050 comes from lumping all the higher n terms together with a ratio of $(7\frac{1}{4})/9$; and .00206 comes from lumping all the higher n terms together with a ratio of $(8\frac{1}{2})/9$. The rather crude estimates for $n > 26$ are shown by dashed lines in Fig. 7. Fortunately, for our desired accuracy the terms in the series soon become small enough so that one needs not be too fastidious about the tail end of the series as long as the limit ratio is safely less than 9. This is to be contrasted with the Ising model for which one wishes to know the limit ratio, which is the y intercept of the curve, with precision in order to locate the transition temperature.

We proceed in the same way for cubic and hexagonal ice as we did for square ice. The ratios, ϕ_{n+2}/ϕ_n, are plotted in Fig. 8. Unlike square ice, the ratios for cubic and hexagonal ice are alternately high and low. The obvious procedure is to extrapolate the high sequence and the low sequence independently. Although this seems especially risky with only two points in each sequence, the series converges fast enough that any reasonable error is relatively small. As with square ice we will extrap-

TABLE III. Summary table for hexagonal ice series[a].

n	0c	1c	2c	3c	all c	u.c.	ϕ_n	ϕ_{n+2}/ϕ_n
6	2				2		2	1.500
8	3				3		3	12.000
10	24	12			36		36	3.167
12	94	48	7		149	−35	114	9.474?
14	582?	450	168	18	1218?	−138	1080?	

[a] The notation is explained in the text.

[11] J. F. Nagle, Ph.D. thesis, Yale University (1965).
[12] See Ref. 7, Sec. 4.5.3.

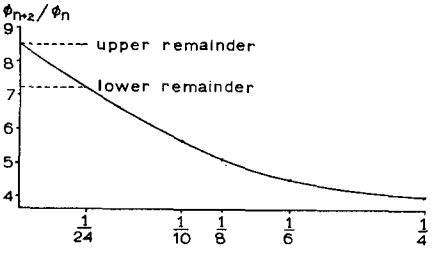


FIG. 7. The term ratios ϕ_{n+2}/ϕ_n versus $1/n$ for square ice.

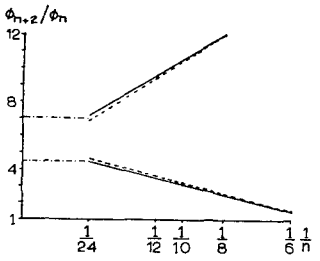


FIG. 8. The term ratios ϕ_{n+2}/ϕ_n versus $1/n$ for cubic ice (solid lines) and hexagonal ice (dashed lines). The dot and dash lines represent the upper and lower estimates for the remaining terms after $n = 24$.

olate the $n = 16$–26 terms and then lump all the remaining terms together with a high value to give an upper estimate and then with a low value to give a lower estimate. For cubic ice we have

$$1.50681 = (1.5)(1 + .004241 + .000287 + .000014) < W_{\text{cubic}} < (1.5)(1 + .004241 + .000287 + .000047) = 1.50686.$$

For hexagonal ice we have

$$1.50683 = (1.5)(1 + .004250 + .000290 + .000013) < W_{\text{hex}} < (1.5)(1 + .004250 + .000290 + .000039) = 1.50687.$$

The second addend in the brackets is the sum of the $n = 6$–14 terms. The third addend is the estimated sum of the $n = 16$–26 terms. The fourth addend is either the high or low estimate for the remainder of the series. The inequalities are not to be taken too seriously, since we have not assigned an estimated error to the third addend. $A \pm .000100$ estimated error assignation to this term seems fairly liberal. Within the limits of this error W_{cubic} and W_{hex} are equal. Referring to them both as W_{ch}, the final estimate is

$$1.5067 < W_{\text{ch}} < 1.5070.$$

6. OTHER METHODS[13]

As mentioned earlier it is possible to get an independent estimate for square ice using the matrix method.[14] One wraps a square net, which is finite in one direction, on a cylinder; this imposes periodic boundary conditions. One examines the matrix which "builds" the arrow configurations step by step along the cylinder. As is well known, the largest eigenvalue, λ_m, of this matrix is related to W as $W(m) = (\lambda_m)^{1/m}$, where m is the number of vertices in each row around the cylinder. $W(m)$ has been computed for $m = 1, \cdots, 7$. It is useful to plot $W(m)$ versus $1/m$. One observes an odd–even effect, but the sequences for only odd m and also for only even m both extrapolate to $W(\infty) = 1.540$, in excellent agreement with the results in Sec. 5.

Another method, proposed by Onsager, uses the series development as a starting point, but instead of evaluating individual terms in the series it attempts to sum the series in one step.[11] The method is based on a recurrence formula for the random walk with no immediate backtracks on a lattice. This formula is summed, Fourier transformed, and solved algebraically. The result is in the form shown below for cubic ice.

$$W = (3/2)\left[1 + (3/4)\left(\frac{1}{2\pi}\right)^3 \int_{-\pi}^{\pi}\int\int dk_1\, dk_2\, dk_3 \cdot \int_0^1 \frac{c - (1/3)u}{9 + 2u(1 - 2c) + u^2}\, du\right], \qquad (11)$$

where

$$c(k_1k_2k_3) = \cos k_1 \cos k_2 + \cos k_2 \cos k_3 + \cos k_3 \cos k_1.$$

This integral and the analogous one for square ice have been evaluated numerically. The difficulty with this method is that it is not exact in principle; some of the cycles are counted too much, others too little so that it is even impossible to tell *a priori* whether the result is too high or too low. Nevertheless, the results are just slightly too large for both square ice and cubic ice, namely,

$$W \approx 1.543 \text{ for square ice and}$$

$$W \approx 1.5077 \text{ for cubic ice.}$$

Although these methods are not of direct value in evaluating the residual entropy of real ice, they do lend support to the approximate results obtained in Sec. 5.

7. SUMMARY

(1) $W_{\text{square ice}}$ is estimated at $1.540 \pm .001$ from the combined results of the matrix method and the series expansion method.

(2) $W_{\text{real ice}}$ is estimated at $1.50685 \pm .00015$ using only the series expansion method.

The latter result gives us the residual entropy of real ice

$$S(0) = .8145 \pm .0002 \text{ cal/deg/mole.}$$

For comparison the experimental results are

$$S(0) = .82 \pm .05 \text{ cal/deg/mole,}^{2}$$

[13] See Ref. 11 for a more detailed account of these methods.
[14] See. Ref. 8 for a review of the matrix method as applied to the Ising model.

and Pauling's approximation is

$$S(0) = .805 \text{ cal/deg/mole.}^3$$

Therefore, our improvement in the calculation of the residual entropy of ice does not destroy the agreement with experiment but, if anything, enhances it.

ACKNOWLEDGMENTS

I wish to thank Dr. L. K. Runnels and Dr. F. H. Stillinger, Jr., for helpful discussions and Professor M. E. Fisher for comments on the manuscript. Especially, I wish to thank Professor L. Onsager for his guidance and considerable help in bringing this work to fruition. Also, a National Science Foundation Cooperative Fellowship is gratefully acknowledged.

APPENDIX: PROOF THAT R_n FOR THE HEXAGONAL ICE LATTICE IS THE SAME AS FOR THE DIAMOND LATTICE

The number of returns to the origin after n steps r_n is the coefficient of unity in the expression for the generating function for the lattice. (The general method of generating functions for random walks on lattices is discussed in Ref. 7, Sec. 5.2.8 and Appendix II.) We may deform the diamond lattice and the hexagonal ice lattice into layers of two dimensional brick lattices. The difference between these two lattices is in the orientation of successive layers. All the layers have the same orientation in the diamond lattice. Since each brick layer has two kinds of vertices there are two kinds of vertices in the diamond lattice. In the hexagonal ice lattice, successive layers have opposite orientations and there are four kinds of vertices. First, we describe the generating functions for steps within a layer, of which there are two depending upon the vertex type. These are $\phi_a = x + x^{-1} + y$ and $\phi_b = x + x^{-1} + y^{-1}$. The generating functions for steps between layers are just $\tau_a = z$ and $\tau_b = z^{-1}$. The generating functions for walks with n steps for the diamond lattice consists of the sum of all products of ϕ's and τ's such that the subscripts a and b alternate. For example, one such term is

$$\phi_b\tau_a\phi_b\tau_a\phi_b\phi_a\tau_b\phi_a\phi_b\phi_a\tau_b. \tag{A1}$$

The total generating function for walks with n steps for the hexagonal lattice also consists of a sum of products of ϕ's and τ's. However, the rule that the subscripts a and b alternate must be changed slightly so that subscripts on the ϕ's in the even numbered layers are the opposite of what they would be in the corresponding term in the diamond lattice generating function. Thus, the hexagonal ice lattice term corresponding to Eq. (1), is

$$\phi_b\tau_a\phi_a\tau_a\phi_b\phi_a\tau_b\phi_b\phi_a\phi_b\tau_b. \tag{A2}$$

We observe that the term Eq. (1) equals the term in Eq. (2). This is so in general provided that the terms represent paths which begin and end in the same layer. The proof of this is immediate upon considering the ϕ's in each layer and observing that there must be the same number of ϕ_a's and ϕ_b's in each layer except the first. Thus, within the class of terms in the generating functions which permits returns to the origin, there is a one to one correspondence between terms which are identically equal for the diamond and hexagonal ice lattices, and therefore, r_n is the same for these lattices.

Lattice Statistics of Hydrogen Bonded Crystals. II. The Slater KDP Model and the Rys F-Model*

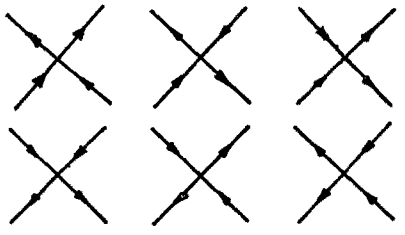
J. F. NAGLE†
Yale University, New Haven, Connecticut
(Received 26 May 1965)

The Slater KDP model, a simple hydrogen bonded ferroelectric model, and the Rys F-model, a simple hydrogen bonded antiferroelectric model, have been treated using both high- and low-temperature series for the partition function. The high-temperature series is a modification of the residual entropy of ice series discussed in I. For each model a temperature is found at which the high-temperature series and the low-temperature series are identically equal. For the KDP model this equality gives a transition temperature and a latent heat is easily calculated, both of which are exact. It so happens that these exact results agree with previous analyses which used only mean field approximations. For the F-model the formal equality of the series gives the first evidence for a phase transition. Although the latent heat calculation throws some doubt on the existence of a transition, after further discussion of the series it is concluded that there most probably is a phase transition.

1. INTRODUCTION

IN this paper, the question of phase transitions in simple models of hydrogen bonded ferroelectrics and antiferroelectrics will be discussed. The models will be similar to the model of ice discussed in the previous paper,[1] in so far as one deals with the question of arranging arrows on the edges of a four coordinated lattice such that the ice rule holds, that is, such that precisely two of the four arrows incident to each vertex point towards that vertex. However, we now introduce vertex energies into the models; each vertex is assigned an energy depending only on the arrows incident to it and the total energy of the lattice is the sum of the vertex energies.

TABLE I. Vertex energies for vertex configurations for various models.



ice	0	0	0
2-dimensional KDP model	0	ϵ	ϵ
F-model	ϵ	0	ϵ

This is illustrated for the square lattice in Table I. Unlike the ice model which remains disordered at all temperatures, the two-dimensional KDP model and the F-model will order at low temperatures. One of the two completely ordered configurations is shown for the two-dimensional KDP model in Fig. 1 and for the F model in Fig. 2.

The physical problems to which these models are related are the phase transitions in KH_2PO_4 (KDP), and isomorphous crystals, which are ferroelectric, and $NH_4H_2PO_4$ (ADP) and isomorphous crystals, which are antiferroelectric.[2] The theory of the ferroelectric transition in KDP has been discussed by many authors.[3-7] In 1941 Slater introduced a model for KDP which is a three-dimensional analogue on a diamond lattice of the two-dimensional KDP model which we have just illustrated.[3] Slater solved this model using essentially a mean field approximation and found a first-order transition, whereas the observed transition is second order. Nevertheless, the observed transition is very narrow, the observed heat of transition is comparable to the latent heat of the model, and both the theoretical and experimental dielectric constants follow the Curie–Weiss law fairly well, so it seemed to be a matter of introducing small changes in the model rather than

* Part of this paper is based on Part II of a dissertation submitted to Yale University in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

† Present address: Department of Chemistry, Cornell University, Ithaca, New York.

[1] J. F. Nagle, J. Math. Phys. 7, 1484 (1966); referred to as I.

[2] (a) W. Känzig, *Solid State Physics*, N. B. Hannay, Ed. (Reinhold Publishing Corporation, New York, 1959), Vol. 4, p. 1; (b) F. Jona and G. Shirane, *Ferroelectric Crystals* (The Macmillan Company, New York, 1962); see especially Chap. III.

[3] J. C. Slater, J. Chem. Phys. 9, 16 (1941).

[4] Y. Takagi, J. Phys. Soc. (Japan) 3, 273 (1948).

[5] E. A. Uehling, *Lectures in Theoretical Physics* (Interscience Publishers, Inc., New York, 1963), p. 138.

[6] H. Takahasi, Proc. Phys. Math. Soc. (Japan) 23, 1069 (1941).

[7] For further references see F. Jona and G. Shirane, in Ref. 2, or Ref. 5.

discarding it altogether. In our opinion the most relevant modification of the Slater model was that of Takagi who relaxed the ice rule to allow one or three of the four arrows incident to a vertex to point towards that vertex.[4] A higher energy is assigned to these vertex configurations; this energy becomes a parameter in fitting the data, which can be done reasonably well. In particular, the transition becomes second order. Other authors have modified the Slater model in other ways, thereby obtaining different parameters with which to fit the data. A few have used different basic models including some with long range forces. However, to our knowledge all have used essentially mean field statistical mechanics with the exception of Takahasi who, using a simple qualitative argument, proved that there is an order disorder transition at Slater's transition temperature.[6] In contrast to the situation for KDP, there has been very little theoretical work done directly related to ADP, probably because of the more complicated experimental behavior. (For example, upon passing through the transition, the crystal shatters and there is thermal hysteresis.) However, a model of cooperative phenomena, the F-model, was proposed and studied by Rys.[8] Apparently unknown to Rys, this model can be regarded as a simple hydrogen bonded antiferroelectric model. Rys used two methods to study the F-model but neither seemed to indicate a phase transition.

The aspect of the theory of hydrogen bonded phase transitions with which this paper is concerned is the statistical mechanical problem of improving upon the mean field approximation. As usual, better statistics seem to restrict one to the simpler and less realistic models. In particular, our methods seem to require models which satisfy the ice rule, such as the two dimensional KDP model and the F-model introduced in the first paragraph. In Sec. 2 the high temperature series is developed for both these models. In Sec. 3 the usual low temperature series are presented. In Sec. 4 the basic results pertaining to the transition in the KDP models are derived. The more challenging question of the existence of a phase transition in the F-model is discussed in Sec. 5.

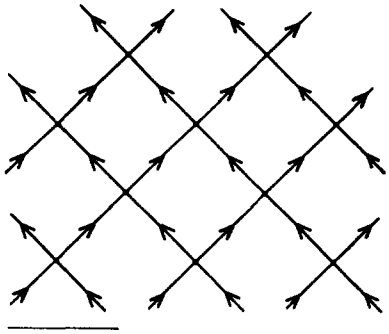


FIG. 1. One of the two completely ordered states for the two-dimensional KDP model.

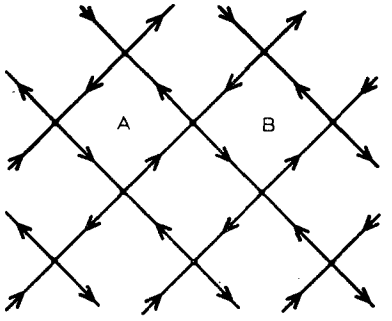


FIG. 2. One of the two completely ordered states for the F-model.

2. HIGH TEMPERATURE SERIES

These series may be developed very much like the residual entropy of ice series in I. A nearest neighbor compatibility function is defined as

$$a(\xi_i, \xi_j) = \begin{cases} 1, \text{ if } \xi_i \text{ and } \xi_j \text{ are compatible} \\ -1, \text{ if } \xi_i \text{ and } \xi_j \text{ are not compatible} \end{cases},$$

where ξ_i, the configuration of the ith vertex, may be any one of six configurations conforming to the ice rule. Next, let ϵ_i be the energy of the ith vertex when it has the configuration ξ, and let $K = \epsilon_i/kT$. Then, we define $B(\xi_i) \equiv \exp(-\epsilon_i/kT) = 1$ or e^{-K} depending upon ξ and the model under consideration. For example,

$$B\left[\quad\right] = \begin{cases} 1 & \text{for the KDP model} \\ e^{-K} & \text{for the F-model} \end{cases}.$$

Then, since $\prod_i B(\xi_i) = \exp(-E/kT)$ where E is the total configuration energy, it is clear that the partition function is

$$Z_H(T) = \sum_{\text{states}} \exp(-E/kT) = \sum_{[\xi]} \prod_{i<j} \tfrac{1}{2}[1 + a(\xi_i, \xi_j)] \prod_i B(\xi_i), \quad (1)$$

where the sum is over all configurations of all N vertices and the first product is over all nearest neighbor edges. Compared to Eq. (4) in I the only new feature is the $B(\xi_i)$ factors.

As was the case with ice it is easy to show that for both the KDP model and the F-model,

$$\sum_{\xi_i} a(\xi_i, \xi_j)B(\xi_i) = 0,$$

$$\sum_{\xi_i} a(\xi_i, \xi_j)a(\xi_i, \xi_k)a(\xi_i, \xi_m)B(\xi_i) = 0,$$

so only products corresponding to Eulerian cycles

[8] F. Rys, Helv. Phys. Acta 36, 537 (1963).

give nonzero contributions to the series, which may be written

$$Z_H(T) = (\tfrac{1}{4})^N(2 + 4e^{-K})^N[1 + \text{cycle contributions}]. \quad (2)$$

In Paper I it was shown how the cycle contributions could be evaluated for any type of Eulerian cycle. Essentially, it was shown that each crossover vertex contributes a factor, 1, and each noncrossover vertex contributes a factor, 1/3, to the cycle contribution. Analogous results may be proven in the same way for the models under consideration here, only now there are several kinds of noncrossover vertices. The results are listed in Table II. To see how the first entry comes about for the F-model, one should perform the sum in Eq. (3) where the vertices i, j, and k are arranged as 

. The

$$\sum_{\xi_i} a(\xi_i, \xi_i)B(\xi_i)a(\xi_i, \xi_k) = \pm(4e^{-K} - 2) \quad (3)$$

sign in Eq. (3) depends upon ξ_i and ξ_k. Negative signs always occur in pairs around a cycle. One must then divide by $2 + 4e^{-K}$ in order to maintain the $(2 + 4e^{-K})^N$ factor in Eq. (2).

Now that all cycle contributions are known we may proceed as in Sec. 4 of I. The series for $Z_h^{1/N}$ for KDP may be written as

$$Z_H^{1/N} = (\tfrac{1}{2} + e^{-K})\left[1 + \sum_{m,n} a_{m,n}\frac{(2e^{-K} - 1)^m}{(2e^{-K} + 1)^n}\right], \quad (4)$$

TABLE II. Vertex factors for two-dimensional KDP and F-models.

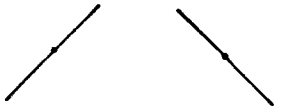
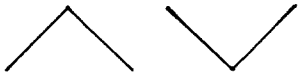
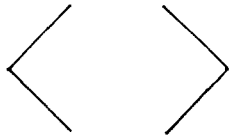
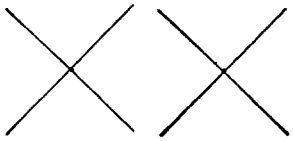
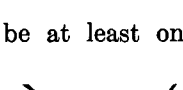
Type of Vertex	Vertex Factors KDP	F-model
	$\frac{1}{2e^{-K}+1}$	$\frac{2e^{-K}-1}{2e^{-K}+1}$
	$\frac{2e^{-K}-1}{2e^{-K}+1}$	$\frac{1}{2e^{-K}+1}$
	$\frac{1}{2e^{-K}+1}$	$\frac{1}{2e^{-K}+1}$
	1	1

TABLE III. $b_{m,n}$ for the F-model.

m	0	2	4	6	8
n					
4	1				
6	2	2			
8	3	12	3		
10	2	50	36	4	
12	−9	212	230	80	5
14	−56				
16	−170 (approx.)				

and for the F-model as

$$Z_H^{1/N} = (\tfrac{1}{2} + e^{-K})\left[1 + \sum_{m,n} b_{m,n}\frac{(2e^{-K} - 1)^m}{(2e^{-K} + 1)^n}\right]. \quad (5)$$

For our purposes the following observation concerning the $a_{m,n}$ coefficients will suffice: Since there must be at least one  vertex and one  vertex, $a_{m,n} \neq 0$ only if $m \geq 2$.

However, more detailed information concerning the F-model series is desirable. For $n \leq 12$ the cycle listings for square ice used to compile Table I in Paper I may again be used with little extra effort to find the $b_{m,n}$. For reasons which will become clear in Sec. 4, it was also desirable to find $b_{0,14}$ and to make an approximation for $b_{0,16}$. The approximation for $b_{0,16}$ involved extrapolating by a ratio method the number of connected cycles from the lower-order results. All these results are summarized in Table III.

3. LOW TEMPERATURE SERIES

The low temperature partition function for the KDP model can be evaluated quite easily as follows

$$Z_L(T) = 2[1 + N^{\frac{1}{2}} \exp[(\log 2 - \epsilon/kT)N^{\frac{1}{2}}] + \cdots]. \quad (6)$$

The factor 2 arises because there are two perfectly ordered states. Within the bracket, the first term 1 corresponds to a perfectly ordered state. Next, we reverse one of the arrows, but this is impossible without reversing an entire chain of arrows which extends from the bottom to the top of the crystal. We may start the chain at $N^{\frac{1}{2}}$ vertices on the bottom surface, the chain may proceed in two ways at each step for $N^{\frac{1}{2}}$ steps, and the total energy difference is $\epsilon N^{\frac{1}{2}}$. This explains the second term. Higher terms will have two or more reversed chains. Now, we notice that when $e^{-K} < \frac{1}{2}$, this term is infinitesimal as are all higher terms. Hence, $Z_L = 1$ for $e^{-K} < \frac{1}{2}$.

When $e^{-K} = \frac{1}{2}$, the exponential becomes unity, and also the interchain coupling vanishes. Thus the maximum term corresponds to exactly half the chains reversed. Since there are $N^{\frac{1}{2}}$ terms in all, one finds

$$Z_L < N^{\frac{1}{2}} \frac{N^{\frac{1}{2}}!}{[(\frac{1}{2}N^{\frac{1}{2}})!]^2}.$$

Therefore, $\lim_{N\to\infty} Z_L^{1/N} = 1$. Hence, for our purposes $Z_L^{1/N} = 1$ for $e^{-K} \leq \frac{1}{2}$.

For the F-model, unlike the KDP model, it is possible for only a few vertices to acquire the higher energy ϵ. For example, in Fig. 2 we can reverse all the arrows in square A to give a total energy of 4ϵ. Next, we could reverse all the arrows in both squares A and B to give a total energy of 6ϵ, and so on. We can now begin a low-temperature series in powers of e^{-K}, where $K = \epsilon/kT$. The coefficient of e^{-nK} will be the number of cycles on the graph which satisfy the following conditions: (1) n is the number of noncrossover vertices, and (2) every cycle is a corner cycle, which means that one may trace the cycle by making a turn at every visited vertex.

Rys in his study transforms the cycles on the square lattice in a one to one way to cycles on a Diagonalgitter.[8] This enables him to discuss comparisons with the Ising model somewhat more easily. However, for our purposes, the foregoing characterization is much more useful because it shows that the low temperature cycles are equivalent to the subclass of high-temperature cycles with $m = 0$. That is, after taking the Nth root of Z_L we have

$$Z_L^{1/N} = 1 + \sum b_{0,n} e^{-nK}, \tag{7}$$

where the $b_{0,n}$ are given in Table III. Our results agree with Rys' except for $b_{0,14}$.[8]

4. THE TRANSITION IN THE KDP MODEL

Now that we have descriptions of both the high- and the low-temperature phases, we investigate where they connect, that is, we find the transition temperature T_c. This will be given when the free energies of the two phases are equal. Since $F = -kT \log Z$, we may consider only the equality of the partition functions. This is particularly easy for the KDP model. Defining T_c as the temperature for which $e^{-K} = \frac{1}{2}$, we see that

$$Z_L^{1/N}(T_c) = 1 = Z_H^{1/N}(T_c). \tag{8}$$

The high-temperature equality in Eq. (8) follows from Eq. (4) because $a_{m,n} \neq 0$ only for $m \geq 2$.

To facilitate the exposition we have been dealing with the two-dimensional KDP model. For Slater's three-dimensional KDP model, the $a_{m,n}$ are changed but the condition for $a_{m,n} \neq 0$ is still $m \geq 2$. The low-temperature series discussion is qualitatively the same.[3,6] Therefore, this discussion applies virtually unchanged to Slater's KDP model.

Slater discussed the transition using $Z_H^{1/N} = (e^{-K} + \frac{1}{2})$. Our $Z_H^{1/N}$ would reduce to his if we neglected correlations around cycles. The unusual feature is that the cycle correlations vanish precisely at the critical temperature, thereby producing agreement between the exact results and the mean field results.

It is also easy to see that the exact latent heat, calculated in Eq. (9) below, is the same as Slater's latent heat, because after differentiation, the cycle contributions still retain the common factor $(2e^{-K}-1)$ which vanishes at T_c;

$$U_H - U_L = U_H = -NkT_c^2[Z_H(T_c)]^{-1/N} \frac{\partial Z_H^{1/N}}{\partial T}\bigg|_{T=T_c}$$
$$= \tfrac{1}{2}N\epsilon. \tag{9}$$

However, as Takahasi also pointed out,[6] the specific heat, which involves two differentiations, would be changed in an exact analysis. The other observable which is usually discussed is the dielectric constant, for which a series may also be written.[9] A first-order approximation, in which cycles are ignored, recovers the usual Slater result. Just as with the specific heat, more precise results would involve a series analysis such as the one done in I for the residual entropy of ice. These problems will be set aside for another communication.

It might also be pointed out that the high-temperature series can also be applied to Takagi's model of KDP with little extra effort. Unfortunately, the low-temperature series becomes difficult to work with, so that the prospects of results more exact than Takagi's seem remote.

5. ON THE EXISTENCE OF A PHASE TRANSITION FOR THE F-MODEL

We notice that for $e^{-K} = \frac{1}{2}$ Eqs. (5) and (7) yield

$$Z_H^{1/N} = 1 + \sum_{n=4}^{\infty} b_{0,n}(\tfrac{1}{2})^n = Z_L^{1/N}. \tag{10}$$

Therefore, it seems that the transition temperature for the F-model is also given by the relation $\epsilon = kT_c \log 2$ just as for the KDP model.

Assuming a transition at T_c we proceed to compute the latent heat as follows:

$$U_H - U_L = -NkT_c^2 \frac{\partial}{\partial T}[\log Z_H^{1/N} - \log Z_L^{1/N}]_{T=T_c}$$
$$= N\tfrac{1}{2}(1 - 3\Lambda), \tag{11}$$

[9] J. F. Nagle, Ph.D. thesis, Yale University (1965).

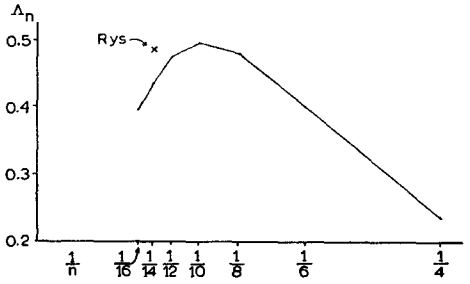


FIG. 3. Λ_n versus $1/n$ for the F-model [see Eq. (12)].

where

$$\Lambda = \sum_{n=4}^{\infty} n b_{0,n} (\tfrac{1}{2})^n \Big[\sum_{n=4}^{\infty} b_{0,n} (\tfrac{1}{2})^n \Big]^{-1}. \qquad (12)$$

We define Λ_n to be the value for Λ when the series in Eq. (12) are truncated after the nth term. The values of Λ_n versus $1/n$ are shown in Fig. 3. Rather disturbingly, for the largest available n, $\Lambda_n > \frac{1}{3}$. If $\lim_{n\to\infty} \Lambda_n > \frac{1}{3}$, Eq. (11) would yield a negative latent heat. Of course, this would mean that the phases should be reversed and that the high-temperature phase should be stable at temperatures slightly below T_c and the low-temperature phase should be stable immediately above T_c. Since $Z_H^{1/N} = \infty$ when $e^{-K} = 0$ and $Z_L^{1/N}$ does not seem to converge even asymptotically when $e^{-K} = 1$, this suggestion would require two more transition temperatures and cannot be taken seriously. The obvious resolution of this point is for $\lim_{n\to\infty} \Lambda_n \leq \frac{1}{3}$. Figure 3 suggests that this might be the case, but the most "natural" continuations of the curve leave the issue undecided. Since it seems so hard to resolve the problem, perhaps $\Lambda = \frac{1}{3}$ and the transition is second order.

The preceding discussion suggests that a more thorough examination of the convergence of the series is needed, since nonconvergence of either series in the neighborhood of T_c might invalidate the formal transition which has been found. There are three adverse cases which should be discussed:

(1) $Z_L^{1/N}$ does not converge for some $e^{-K} \leq \frac{1}{2}$,
(2) $Z_H^{1/N}$ does not converge for some e^{-K} such that $\frac{1}{2} < e^{-K} \leq 1$,
(3) $Z_H^{1/N}$ does not converge for all $e^{-K} < \frac{1}{2}$. Thus, it might be possible that $Z_H^{1/N} < Z_L^{1/N}$ whenever $Z_H^{1/N}$ exists. This would allow $\Lambda \geq \frac{1}{3}$.

Even if all the terms were positive in the low-temperature series, it would still seem to converge for $e^{-K} = \frac{1}{2}$, so that case (1) seems quite unlikely. To examine cases (2) and (3), the series in Eq. (5) was evaluated for various values of e^{-K}. The series was first summed over m and then over n. Over this temperature range it behaves nicely as a function of e^{-K} and as a function of n, as shown in Fig. 4. Thus, the evidence suggests that none of these adverse cases occurs and that both the $Z_L^{1/N}$ series and the $Z_H^{1/N}$ series converge through T_c. This would imply either that Λ must be smaller than $\frac{1}{3}$ and there is a first-order phase transition or that Λ equals $\frac{1}{3}$ and there is a second- or higher-order phase transition.

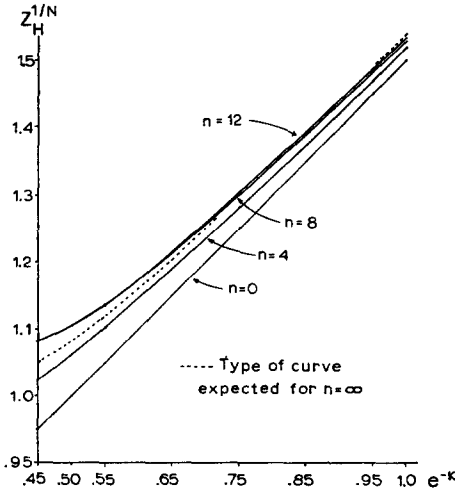


FIG. 4. Truncated $Z_H^{1,N}$ series versus e^{-K}. For the various n values all the $b_{m,n'}$ terms with $n' \leq n$ were used in the truncated series.

Strictly speaking, a more detailed series analysis is in order for the KDP model also. If this is done one finds that $Z_H^{1/N} < 1$ for $e^{-K} < \frac{1}{2}$. Since there are two ordered states with zero energy, one must have $Z^{1/N} \geq 1$ so that $Z_H^{1/N}$ is definitely not valid below T_c. More important, our discussion of Z_L in Sec. 3 implies that there is a spontaneous disordering at T_c. This is the essential idea in Takahasi's analysis. In contrast, for the F-model it has not been possible thus far to find a qualitative proof that there is indeed a transition at T_c.

Note added in proof: Recently the author has found that the analysis of low-temperature series for the degree of order indicates very strongly that there is a phase transition for some $T_c \leq \epsilon/k \log 2$.

ACKNOWLEDGMENTS

I wish to thank Professor L. Onsager for suggesting a problem which led to the work reported in this paper and for reading the manuscript. Also, I wish to thank Dr. H. N. V. Temperley and Dr. F. H. Stillinger, Jr., for bringing Rys's paper to my attention. Finally, a Socony-Mobile Fellowship and some additional Yale University financial support are gratefully acknowledged.

Nonrelativistic Quantum Theory of an Electron in an Arbitrarily Intense Laser Field

H. H. NICKLE
Robert College, Istanbul, Turkey
(Received 6 December 1965)

An essentially exact treatment of the time-dependent Schrödinger equation for a Bloch electron (or a free electron) in the presence of an arbitrarily intense laser field is described. Expressions for the wavefunction, current density, and energy of the electron state are presented in closed form for the case when the effective mass approximation is valid. The limitations of an "almost exact" solution of very simple form are investigated, the corrections to the almost exact solution being determined by the WKB approximation method. The exact solution for the wavefunction turns out to be *quite different* from that given by perturbation theory. However, the changes in the values of the current density and energy due to the presence of the laser field turn out to be, within the limitations imposed by the nonrelativistic nature of the Schrödinger theory, linear and quadratic in the field amplitude, and therefore agree with the results of perturbation theory.

1. INTRODUCTION

COHERENT optical frequency Maxwell waves of high intensity are now obtainable from lasers. A large number of experiments[1] and numerous calculations[2] have been made concerning the interaction of such a radiation field with matter. The present article concerns the quantum eigenstates of a free electron and, by an extension indicated later, the eigenstates of a Bloch electron in the presence of an arbitrarily intense coherent Maxwell field with a single frequency $(\omega/2\pi)$ and a single propagation wave vector $\mathbf{K}$, the electromagnetic field being treated as a classical field.

One might expect to be able to treat the interaction between an electron and a laser field by the use of time-dependent perturbation theory, since the fine structure constant is certainly small compared to unity. However, it turns out that the coupling constants appropriate for describing the interaction of an electron with a laser field are

$$g_1 \equiv (e/mc)(\mathbf{p}\cdot\mathbf{A})/\hbar\omega \quad \text{and} \quad g_2 \equiv e^2A^2/2mc^2\hbar\omega,$$

where $\mathbf{A}$ is the vector potential of the Maxwell field and $\mathbf{p}$ is the unperturbed momentum of the electron.

First we note that

$$(g_1^2/g_2) = 4\{(\mathbf{p}\cdot\hat{\mathbf{a}})^2/2m\}(\hbar\omega)^{-1},$$

where $\mathbf{p}\cdot\hat{\mathbf{a}}$ is the component of the momentum in the direction of the vector potential. For optical masers, $\hbar\omega \sim 1$ eV, and so (g_1^2/g_2) will be of order unity, unless we consider circumstances in which the unperturbed kinetic energy of the electron is either much greater or much smaller than one electron volt. Next we note that $g_2 \ll g_1$ as long as $(eA) = e(\lambda/2\pi)\,|\mathcal{E}|$ is very small compared to $2\{(2mc^2)(\mathbf{p}\cdot\hat{\mathbf{a}})^2/2m\}^{\frac{1}{2}}$ (here λ is the wavelength in vacuum at frequency $\omega/2\pi$ and $\mathcal{E}$ is the electric field vector). In other words, g_2 is negligible compared to g_1 provided the electric field intensity (in V/cm) is very small compared to $(4\pi/\lambda)\{10^3T\}^{\frac{1}{2}}$, where T is the unperturbed kinetic energy of the electron expressed in electron volts. We also note that if $\hbar\omega = 1$ eV, then $g_2 = 1$ when $|\mathbf{A}| = (\lambda/2\pi)\,|\mathcal{E}| = 10^3$ V. This implies an electric field intensity on the order of 10^8 V/cm, which corresponds to an energy flux of 2.6×10^{13} W/cm^2 in free space. It should be possible to obtain electric field strengths of this magnitude or greater by focusing the output beam of lasers which will be developed in the near future. For example, Geusic and Scovil[3] suggest that it may be possible to obtain output powers of 10^2 to 10^3 MW in a single mode by Q-switching a diffraction-limited pulsed ruby oscillator.

The present article describes a particular method for solving the Schrödinger equation for one electron in the presence of an arbitrarily intense monochromatic plane wave field. In this connection, we wish to emphasize that some care must be exercised in any attempt to apply the theory developed here

[1] See, for example, *Proceedings of the Third International Conference on Quantum Electronics*, P. Grivet and N. Bloembergen, Eds. (Columbia University Press, New York, 1964), Vols. 1 and 2. Also see the following review article: J. E. Geusic and H. E. D. Scovil, *Reports on Progress in Physics* (The Institute of Physics and the Physical Society, London, 1964), Vol. XXVII, pp. 241–327.

[2] Here we mention only a few of the theoretical articles on nonlinear optical phenomena: J. A. Armstrong, N. Bloembergen, J. Ducuing, and P. S. Pershan, Phys. Rev. 127, 1918 (1962); P. A. Franken and J. F. Ward, Rev. Mod. Phys. 35, 23 (1963); E. Adler, Phys. Rev. 134, A728 (1964); J. F. Ward, Rev. Mod. Phys. 37, 1 (1965); also see Ref. 1.

[3] *Proceedings of the Third International Conference on Quantum Electronics*, P. Grivet and N. Bloembergen, Eds. (Columbia University Press, New York, 1964), Vol. 2, p. 1218.

to certain experimental situations, e.g., to the case of an electron interacting with the output beam of a pulsed laser. Here we only consider an artificial situation in which the electron is *always* interacting with the laser field, which is assumed to be a perfect plane wave and also perfectly monochromatic.

The solution of the corresponding classical problem is well known.[4] Also, the problem of a Dirac electron in the presence of plane wave electromagnetic radiation has been solved by Volkov.[5] In principle, the solution of the Schrödinger problem can be obtained by taking the nonrelativistic limit of an appropriately chosen solution for the Dirac electron. However, the method described here is much less complicated. There is no physical justification for treating the electron relativistically, since a consideration of the classical problem indicates that relativistic effects will become important only when g_1 and/or g_2 are no longer small in comparison with $\gamma^{-1} \cong 2mc^2/\hbar\omega \sim 10^6$ [the parameter γ is defined by Eq. (2.23)].

The organization of the article is as follows: In Sec. 2 the fundamental equations are derived. We use the same method as in ordinary time-dependent perturbation theory[6]; that is, we assume that an exact eigenfunction of the Schrödinger equation can be expressed as an expansion in terms of the eigenfunctions of the unperturbed Schrödinger equation, where the expansion coefficients must depend on the time. Thus we obtain the usual set of coupled differential equations for the expansion coefficients. However, unlike perturbation theory, we find an essentially exact solution for the expansion coefficients.

The solution obtained in Sec. 3 is in closed form and valid provided g_1 and/or g_2 are small compared to $\gamma^{-\frac{1}{2}} \sim 10^3$. This solution should be quite useful for making certain types of calculations. For example, by using this solution one may obtain a formula for the "free carrier" absorption of the field energy by Bloch electrons which is valid for large values of g.[7] However, the solution obtained in Sec. 3 is not quite exact, so the corrections to it are calculated in Sec. 4 as an expansion in powers of γg (with g the larger of the two coupling constants). Since γg ceases to be small at very large fields indeed (electric field intensity on the order of 10^{11} V/cm or larger) when, in any case, relativistic effects would appear, the calculation so far as it has been carried in this paper is essentially complete. In other words, although one could carry the calculations of Sec. 4 to higher powers of γg, there is no reason to do so since the Schrödinger equation is not valid at such large field intensities. The results of Secs. 3 and 4 are presented in the form of explicit formulas for the Schrödinger eigenfunctions, the current density, and the average energy of the electron.

[4] See, for example, L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts), 2nd ed., pp. 128, 129.
[5] D. M. Volkov, Z. Physik 94, 250 (1935); Zh. Eksperim. i Teor. Fiz. 7, 1286 (1937).
[6] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1955), 2nd ed., Sec. 29.
[7] P. J. Price (to be published).

2. FUNDAMENTAL EQUATIONS

In this section we describe a method of solving the Schrödinger equation when the Hamiltonian is given by

$$H = H_0 + (1/2m)\{|\boldsymbol{\Pi}|^2 - |\mathbf{p}|^2\}, \quad (2.1)$$

where

$$\boldsymbol{\Pi} = \mathbf{p} + (e/c)\mathbf{A}, \quad (2.2)$$

$$\mathbf{A} = \mathbf{A}_\omega e^{i(\omega t - \mathbf{K}\cdot\mathbf{r})} + \text{complex conjugate (c.c.)} \quad (2.3)$$

(a c number in the problem considered here) is the vector potential of a monochromatic plane wave of frequency $\omega/2\pi$ and wave vector $\mathbf{K}$, $-e$ is the electron charge, H_0 is the unperturbed Hamiltonian, and the gauge is chosen such that the divergence of $\mathbf{A}$ is equal to zero and $\mathcal{E} = -(1/c)(\partial\mathbf{A}/\partial t)$ (that is, the scalar potential of the Maxwell field vanishes). For the unperturbed Hamiltonian H_0, we consider two cases: (a) $H_0 = (p^2/2m)$ (the free-electron case), (b) $H_0 = (p^2/2m) + V(\mathbf{r})$, where $V(\mathbf{r})$ is a periodic lattice potential (the Bloch electron case).

It is clear from (2.1) that the perturbation operator,

$$H_F \equiv H - H_0, \quad (2.4)$$

has the form

$$H_F = (e/2mc)\{\mathbf{p}\cdot\mathbf{A} + \mathbf{A}\cdot\mathbf{p}\} + (e^2/2mc^2)A^2, \quad (2.5)$$

where $\mathbf{p} = -i\hbar\nabla$. Since the divergence of $\mathbf{A}$ is zero, we can write

$$H_F = -(ie\hbar/mc)\mathbf{A}\cdot\nabla + (e^2/2mc^2)A^2. \quad (2.6)$$

If we apply H_F to the unperturbed wavefunction $\psi_0 \sim \exp[i(\omega_0 t - \mathbf{k}\cdot\mathbf{r})]$ of a free electron with energy $E \equiv \hbar\omega_0$ and momentum $\hbar\mathbf{k}$, we obtain

$$H_F\psi = (e\hbar/mc)(\mathbf{k}\cdot\mathbf{A})\psi + (e^2/2mc^2)A^2\psi. \quad (2.7)$$

Thus, the appropriate coupling constants which characterize the disturbance of an electron by the Maxwell field are

$$g_1 \equiv \{\mathbf{k}\cdot\mathrm{Re}\,\mathbf{A}_\omega\}(e\hbar/mc)(1/\hbar\omega), \quad (2.8)$$

$$g_2 \equiv \{\mathrm{Re}\,\mathbf{A}_\omega^2\}(e^2/2mc^2)(1/\hbar\omega), \quad (2.9)$$

where $\mathbf{k}$ is the wave vector of the unperturbed electron. These are the same definitions as in the Introduction, except that we are now using a complex vector potential. In connection with the definition of g_1, let us write the vector potential in the form

$$\mathbf{A} = \hat{\mathbf{a}}A \cos(\omega t - Kz).$$

Then $\nabla\cdot\mathbf{A} = 0$ implies that the unit vector $\hat{\mathbf{a}}$ lies in the xy plane. Note that the perturbation operator H_F commutes with p_x and p_y; in other words, the component of the electron's wave vector perpendicular to $\mathbf{K}$ is a constant of the motion.

Perturbation theory[6] would give the state of the electron as a power series in g_1 and g_2. The object of the present calculation, however, is to obtain a solution whose validity is *not* limited to small values of the coupling constants. It should also be noted that the spin term $\mu_0\boldsymbol{\sigma}\cdot\mathbf{B}$ has been omitted from the Hamiltonian (2.1) because it is negligible in comparison with the average value of the term $(e/mc)\mathbf{A}\cdot\mathbf{p}$, at least for an electron whose unperturbed kinetic energy is large compared to 2.5×10^{-7} eV (see Appendix A).

Let the unperturbed state with wave vector $\mathbf{k} + n\mathbf{K}$ ($n = 0, \pm1, \pm2, \pm3, \cdots$) have normalized wavefunction $\phi_n \exp(-i\omega_n t)$, with

$$H_0\phi_n = E_n\phi_n, \qquad E_n \equiv \hbar\omega_n, \tag{2.10}$$

and let us look for a solution to the Schrödinger equation of the form

$$\psi = \sum_{n=-\infty}^{+\infty} a_n(t)\phi_n. \tag{2.11}$$

Substitution of (2.11) into the time-dependent Schrödinger equation, $i\hbar\dot{\psi} = (H_0 + H_F)\psi$, leads to the following infinite system of coupled differential equations[8]:

$$i\hbar\dot{a}_n = \hbar\omega_n a_n + \sum_m a_m(\phi_n\,|H_F|\,\phi_m). \tag{2.12}$$

For a free electron (in this case the unperturbed eigenfunctions are plane waves), the only nonvanishing off-diagonal matrix elements of H_F are

$$(n\,|H_F|\,n \pm 1) = \hbar\omega g_1 e^{\pm i(\omega t+\alpha)}, \tag{2.13a}$$

$$(n\,|H_F|\,n \pm 2) = \hbar\omega g_2 e^{\pm 2i(\omega t+\alpha)}, \tag{2.13b}$$

where the phase angle α is defined as the inverse tangent of the ratio of the imaginary and real parts of the complex vector $\mathbf{A}_\omega$.

[8] Unless explicitly indicated otherwise, summations run over all positive and negative integers (including zero) from $-\infty$ to $+\infty$.

Substitution of Eqs. (2.13) into (2.12) gives

$$\begin{aligned} i\hbar\dot{a}_n &= \hbar\omega_n a_n \\ &+ \hbar\omega g_1\{a_{n-1}e^{-i(\omega t+\alpha)} + a_{n+1}e^{i(\omega t+\alpha)}\} \\ &+ \hbar\omega g_2\{a_{n-2}e^{-2i(\omega t+\alpha)} + a_{n+2}e^{2i(\omega t+\alpha)}\}. \end{aligned} \tag{2.14}$$

The Schrödinger equation (2.14) may be solved by substituting

$$a_n = G_n \exp\{-i[n(\omega t + \alpha) + \omega_0 t]\}, \tag{2.15}$$

with the G_n independent of t. Then

$$\begin{aligned} \{n + [(\omega_0 - \omega_n)/\omega]\}G_n &= g_1(G_{n-1} + G_{n+1}) \\ &+ g_2(G_{n-2} + G_{n+2}). \end{aligned} \tag{2.16}$$

It is convenient to normalize the solution of (2.16) such that

$$\sum |G_n|^2 = \sum |a_n|^2 = 1. \tag{2.17}$$

For the case of a Bloch electron in a perfect crystal, the unperturbed eigenfunctions are no longer just plane waves, but are modulated by functions $U(\mathbf{r};\mathbf{k})$ having the periodicity of the crystal lattice:

$$\phi_n = U(\mathbf{r};\mathbf{k} + n\mathbf{K})e^{i(\mathbf{k}+n\mathbf{K})\cdot\mathbf{r}}. \tag{2.18}$$

However, one can still use Eqs. (2.13a) and (2.13b) for the Bloch electron case, provided $K \ll k$, and it is therefore a valid approximation to replace $(U(\mathbf{r};\mathbf{k} + n\mathbf{K}) \mid U(\mathbf{r};\mathbf{k} + (n \pm 1)\mathbf{K}))$ by unity. In this connection, we only consider the solution which is obtained when the interband terms of H_F are neglected. In other words, the electron state is assumed to belong to a single band. Furthermore, scattering by crystal imperfections and lattice vibrations is completely neglected. The envelope wavefunction for a Bloch electron may be defined as

$$\Psi(\mathbf{r}, t) \equiv \sum_n a_n(t)e^{i(\mathbf{k}+n\mathbf{K})\cdot\mathbf{r}}. \tag{2.19}$$

For free electrons, this is the actual Schrödinger wavefunction—the same as (2.11). In either case, we hereafter refer to it as *the wavefunction*.

We require the solution of (2.16) which reduces to $G_n = \delta_{n,0}$ and is nonsingular at $g_1 = g_2 = 0$. Physically, this is just the requirement that our solution to the Schrödinger equation must reduce to the unperturbed eigenfunction ϕ_0 when no Maxwell field is present. Using the effective mass approximation, let us replace $\hbar\omega_n \equiv E(\mathbf{k} + n\mathbf{K})$ by

$$\begin{aligned} \hbar\omega_n &= \hbar\omega_0 + nK_i(\partial E/\partial k_i)_0 \\ &+ \tfrac{1}{2}n^2K_iK_j(\partial^2 E/\partial k_i\,\partial k_j)_0. \end{aligned} \tag{2.20}$$

Substituting (2.20) into (2.16), we obtain

$$(n - \gamma n^2)G_n = g_1'(G_{n-1} + G_{n+1}) + g_2'(G_{n-2} + G_{n+2}), \quad (2.21)$$

where

$$g_i' \equiv g_i\left\{1 - \frac{K_i(\partial E/\partial k_i)_0}{\hbar\omega}\right\}^{-1}, \quad (2.22)$$

$$\gamma \equiv (1/2\hbar\omega)K_iK_j(\partial^2 E/\partial k_i\,\partial k_j)_0\left\{1 - \frac{K_i(\partial E/\partial k_i)_0}{\hbar\omega}\right\}^{-1} \approx \hbar\omega/2mc^2. \quad (2.23)$$

Note that the second term in the denominators of (2.22) and (2.23) is the ratio of the component of the unperturbed electron's velocity in the direction of wave propagation to the velocity c' of an electromagnetic wave in the medium, that is,

$$\frac{\mathbf{K}\cdot(\partial E/\partial \mathbf{k})_0}{\hbar\omega} = \frac{\mathbf{v}^{(0)}\cdot\hat{\mathbf{K}}}{c'}. \quad (2.24)$$

3. ALMOST EXACT SOLUTION

Now, we obtain the required solution of (2.21) with the approximation of neglecting the γn^2 term; that is, we obtain the solution of

$$nG_n = g_1'(G_{n-1} + G_{n+1}) + g_2'(G_{n-2} + G_{n+2}). \quad (3.1)$$

First, let us consider the case when $g_2 \ll g_1$, which corresponds to physical situations in which the electric field intensity (in V/cm) is very small compared to $(4\pi/\lambda)\{10^3T\}^{\frac{1}{2}}$, where λ is the wavelength of the laser radiation and T (expressed in electron volts) is the kinetic energy of the unperturbed electron. For $\lambda \sim 5 \times 10^{-5}$ cm and $T \sim 1$ eV, this corresponds to $\mathcal{E} \ll 8 \times 10^6$ V/cm. Neglecting the g_2 term completely, the resulting difference equation is

$$(n/g_1')G_n = G_{n-1} + G_{n+1}. \quad (3.2)$$

Equation (3.2) is identical to one of the two recursion relations satisfied by cylinder functions,[9] which are members of any of the sequences $\{C_n(x)\}$ satisfying

$$C_{n-1}(x) + C_{n+1}(x) = (2n/x)C_n(x), \quad (3.3a)$$

$$C_{n-1}(x) - C_{n+1}(x) = 2[dC_n(x)/dx]. \quad (3.3b)$$

Here, x and n take any complex values. For real integer values of n, it is clear that $G_n = C_n(2g_1')$ is a solution of Eq. (3.2). It can be shown that the general solution of Eqs. (3.3) is

$$C_n(x) = w_1(n)J_n(x) + w_2(n)Y_n(x), \quad (3.4)$$

where J_n and Y_n are Bessel functions of the first and second kinds, respectively, and $w_1(n)$ and $w_2(n)$ are arbitrary periodic functions of n with period equal to unity. In the present article, we confine our attention to that particular solution of Eq. (3.2) which satisfies the following two requirements: (1) It should satisfy Eq. (3.2) for any arbitrary positive real value of g_1', and the functional form of the solution's dependence of $2g_1'$ should not vary when g_1' changes; (2) $\psi \to a_0\phi_0$ as $g_1 \to 0$.[10] Requirement (1) means that the general solution of Eq. (3.2) is

$$G_n(2g_1') = c_1J_n(2g_1') + c_2Y_n(2g_1'), \quad (3.4')$$

where c_1 and c_2 are constants, independent of g_1'. Then requirement (2) necessitates that $c_2 = 0$, because $Y_n(2g_1')$ does not remain finite in the limit $g_1' \to 0$.

There is another ansatz which, at first glance, appears to give another solution to the Schrödinger equation. Instead of (2.15) let us assume

$$a_n = H_n \exp\{-i[n(\omega t + \alpha) + \Omega t + \omega_0 t]\}, \quad (3.5)$$

where the H_n are independent of t, and Ω is an arbitrary constant. Then (3.2) is replaced by

$$(n + \epsilon)H_n = g_1'(H_{n-1} + H_{n+1}), \quad (3.6)$$

where

$$\epsilon \equiv (\Omega/\omega)\left\{1 - \frac{K_i(\partial E/\partial k_i)_0}{\hbar\omega}\right\}^{-1} \quad (3.7)$$

is not, in general, an integer. Then formal solutions of (3.6) are

$$H_n = C_{n+\epsilon}(2g_1'). \quad (3.8)$$

[9] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge University Press, Cambridge, England, 1952), 2nd ed., Chap. III, pp. 82–84.

[10] One reason for believing that the solutions of physical interest should satisfy requirement (1) is that g_1 [defined by Eq. (2.8)] is, for a given value of the unperturbed electron's wave vector $\mathbf{k}$, directly proportional to the amplitude of the vector potential of the Maxwell field and inversely proportional to its frequency. One can imagine a series of *Gedanken* experiments in which, by choosing different intensities and/or frequencies of the Maxwell field, one could experimentally realize the situation corresponding to any given value of g between zero and some upper bound $g_{\max}$ (note: here g denotes g_1'). We therefore require that our solution be able to describe *any* of these situations, i.e., our solution should satisfy Eq. (3.2) for any and all real values of g in the interval $0 \le g \le g_{\max}$. If we only desired a solution of Eq. (3.2) for one particular value of g, then the coefficient c_2 in Eq. (3.4′) might depend on g, and it might be possible to satisfy requirement (2) without setting $c_2 = 0$ (we wish to thank the referee for pointing out this possibility). It is beyond the framework of the present calculation to discuss solutions of the latter type [i.e., solutions valid for particular values of g containing both $J_n(2g)$ and $Y_n(2g)$], except to say that they are probably not of physical importance.

Since only Bessel functions of the first kind and also of integer order are finite at the origin,[11] the solutions (3.8) do not give an admissible result for the state of an electron. Thus, the only acceptable solution of this type is

$$a_n(t) = J_n(2g_1') \exp\{-i[n(\omega t + \alpha) + \omega_0 t]\}. \quad (3.9)$$

There is one other point which should be mentioned. Cylinder functions satisfy (3.3b) as well as (3.3a), but there is no corresponding equation which the G_n must satisfy. In fact, it is known that Lommel's polynomials[12] satisfy only the single recurrence relation (3.3a). If (3.3a) is used to express $J_{n+m}(x)$ linearly in terms of $J_n(x)$ and $J_{n-1}(x)$, the coefficients define the Lommel polynomials $R_{m,n}(x)$:

$$J_{n+m}(x) = J_n(x)R_{m,n}(x) + J_{n-1}(x)R_{m-1,n+1}(x). \quad (3.10)$$

These polynomials satisfy the following recurrence relation:

$$R_{m-1,n}(x) + R_{m+1,n}(x) = 2(n+m)x^{-1}R_{m,n}(x). \quad (3.11)$$

Hence, $G_n = R_{n,0}(2g_1')$ will formally satisfy Eq. (3.2). However, $R_{n,0}(2g_1')$ is not bounded for all integer values of n in the limit $g_1' \to 0$ [for example, $R_{3,0}(g) = -4g^{-1}$)], so we do not consider this class of solutions any further.

We are left with a unique solution, (3.9), corresponding to the wavefunction

$$\Psi(\mathbf{r}, t) = e^{i(\mathbf{k}\cdot\mathbf{r} - \omega_0 t)} \sum_n J_n(2g_1')e^{-in(\omega t - \mathbf{K}\cdot\mathbf{r} + \alpha)}. \quad (3.12)$$

Using the Jacobi–Anger formula,[13]

$$e^{iz\sin u} = \sum_n e^{inu}J_n(z), \quad (3.13)$$

we can rewrite Eq. (3.12) in the form

$$\Psi = \exp\{i[(\mathbf{k}\cdot\mathbf{r} - \omega_0 t) + 2g_1' \sin(\mathbf{K}\cdot\mathbf{r} - \omega t - \alpha)]\}. \quad (3.14)$$

Now, let us obtain the solution to Eq. (3.1) without confining our attention to the case when g_2 can be set equal to zero. It is convenient to introduce the generating function

$$Q(z) \equiv \sum_{n=-\infty}^{+\infty} z^n G_n. \quad (3.15)$$

If the G_n satisfy (3.1), then

$$z(dQ/dz) = \{g_1'(z + z^{-1}) + g_2'(z^2 + z^{-2})\}Q, \quad (3.16)$$

$$Q(z) = \exp\{g_1'(z - z^{-1}) + (\tfrac{1}{2})g_2'(z^2 - z^{-2})\}. \quad (3.17)$$

It follows from Eqs. (2.15), (2.19), and (3.16) that the wavefunction is

$$\Psi = e^{i(\mathbf{k}\cdot\mathbf{r} - \omega_0 t)}Q(s), \quad (3.18)$$

where

$$s \equiv \exp[i(\mathbf{K}\cdot\mathbf{r} - \omega t - \alpha)]. \quad (3.19)$$

Although $Q(z)$ is formally defined for all complex values of z, the wavefunction contains $Q(z)$ only for z on the unit circle (values of z on the unit circle are hereafter denoted by s).

The differential equation (3.16) corresponds, by (3.18), to the Schrödinger equation. If the γn^3 term of (2.21) had been retained, it would have resulted in a term $(zd/dz)^2Q$ in (3.16), corresponding to the $-(\hbar^2/2m)\nabla^2\Psi$ term in the Schrödinger equation. It is, of course, clear that one can also directly transform the Schrödinger equation into the form (3.16) (plus the terms neglected in this section) without going through the intermediate step of the difference equations for G_n. This is done for the free-electron case in Appendix B.

We believe that the "almost exact" solution given by Eqs. (3.17), (3.18), and (3.19) may be quite useful in practical calculations. Therefore, we list four important properties of the $G_n(g_1', g_2')$, the time-independent coefficients satisfying Eq. (2.1).

Property 1:

$$G_p(g_1, g_2) = \sum_{n=-\infty}^{+\infty} J_{p-2n}(2g_1)J_n(g_2). \quad (3.20)$$

This can be proved by writing out the product of the two series

$$e^{g_1(z - z^{-1})} = \sum_m z^m J_m(2g_1),$$

$$e^{(1/2)g_2(z^2 - z^{-2})} = \sum_{n=-\infty}^{+\infty} z^{2n} J_n(g_2),$$

and collecting all terms in the double sum which have $m + 2n = p$. Note that $G_n(g_1, 0) = J_n(2g_1)$ and therefore $G_n(0, 0) = \delta_{n,0}$ since $J_n(0) = \delta_{n,0}$.

Property 2:

$$\sum_{m=-\infty}^{+\infty} G_m(f_1, f_2)G_{m+p}(g_1, g_2) = G_p(f_1 - g_1, f_2 - g_2). \quad (3.21)$$

[11] A concise review of the essential properties of Bessel functions is given in the following book: J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), Sec. 6.5.

[12] See Secs. 9.6 (pp. 294, 295) and 9.63 (pp. 298, 299) of Ref. 9.

[13] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill Book Company, Inc., New York, 1955), Vol. 2, p. 7.

Proof: Let $Q_1(z)$ and $Q_2(z)$ denote the generating functions for $G_n(f_1, f_2)$ and $G_n(g_1, g_2)$, respectively. Then

$$\sum_i z^i G_i(f_1 - g_1, f_2 - g_2) = Q_1(z)Q_2(1/z) = \sum_m \sum_n z^{m-n} G_m(f_1, f_2)G_n(g_1, g_2),$$

and (3.21) follows by collecting all terms in the double sum which have $m - n = p$.

Property 3:

$$\sum_n n\{G_n(g_1, g_2)\}^2 = 0. \tag{3.22}$$

Proof: Use Eq. (3.1) to write

$$\sum_n n\{G_n(g_1, g_2)\}^2 = g_1 \sum_n [G_n(g_1, g_2)G_{n-1}(g_1, g_2) + G_n(g_1, g_2)G_{n+1}(g_1, g_2)] + g_2 \sum_n [G_n(g_1, g_2) \times G_{n-2}(g_1, g_2) + G_n(g_1, g_2)G_{n+2}(g_1, g_2)].$$

Then use (3.21) to evaluate the sums:

$$\sum_n nG_n^2 = g_1[G_{-1}(0, 0) + G_1(0, 0)] + g_2[G_{-2}(0, 0) + G_2(0, 0)] = 0.$$

Property 4:

$$\sum_n n^2\{G_n(g_1, g_2)\}^2 = 2(g_1^2 + g_2^2). \tag{3.23}$$

Proof: Again use (3.1) to replace nG_n by a linear combination of $G_{n\pm1}$ and $G_{n\pm2}$. Then use (3.21) to evaluate the sums.

The wavefunction given by (3.17) and (3.18) is

$$\Psi = \exp\{i(\mathbf{k}\cdot\mathbf{r} - \omega_0 t) + 2ig_1' \sin(\mathbf{K}\cdot\mathbf{r} - \omega t - \alpha) + ig_2' \sin[2(\mathbf{K}\cdot\mathbf{r} - \omega t - \alpha)]\}. \tag{3.24}$$

According to (3.24) the probability density, $|\Psi|^2$ is constant, whereas, for a free electron, the current density

$$\mathbf{J} = -(e\hbar/2im)(\Psi^* \operatorname{grad}\Psi - \Psi \operatorname{grad}\Psi^*) - (e^2/mc)\mathbf{A}|\Psi|^2$$
$$= -(e\hbar/m)\{\mathbf{k} + 2g_1'\mathbf{K}\cos(\mathbf{K}\cdot\mathbf{r} - \omega t - \alpha) + 2g_2'\mathbf{K}\cos[2(\mathbf{K}\cdot\mathbf{r} - \omega t - \alpha)]\} - (e^2/mc)\mathbf{A} \tag{3.25}$$

is not. (In the effective mass approximation for an electron in a crystal lattice, one would replace m by the effective mass.) Hence, it is clear that the "almost exact" solution (3.24) *cannot* possibly satisfy the equation of continuity

$$\operatorname{div}\mathbf{J} - e(\partial/\partial t)|\Psi|^2 = 0. \tag{3.26}$$

In Sec. 4 we show that (3.26) is satisfied when the γ_n^2 term of (2.21) is taken into consideration.

It is possible to give a direct physical interpretation to the terms of the current density which are linear and quadratic in the field. Making the rough approximation that the eigenvalue of $\mathbf{\Pi}$ is $\hbar\mathbf{k} + (e/c)\mathbf{A}$, we see that the Lorentz force is proportional to $[\hbar\mathbf{k} + (e/c)\mathbf{A}] \times (\mathbf{K} \times \mathbf{A})$. Thus the g_1 term on the right-hand side of (3.25) represents the effect of the $\mathbf{k} \times (\mathbf{K} \times \mathbf{A})$ term in the Lorentz force.[14] The g_2 term on the right-hand side of (3.25) represents the effect of the $\mathbf{A} \times (\mathbf{K} \times \mathbf{A})$ term in the Lorentz force. The last term on the right-hand side of (3.25), $-(e^2/mc)\mathbf{A}$, represents the effect of direct acceleration of the electron by the electric field $\mathcal{E} = -(1/c)\dot{\mathbf{A}}$.

We also note that a linear superposition of wavefunctions (3.24), corresponding to "middle" states with different values of $\mathbf{k}$, is also a solution to the Schrödinger equation. In particular, the unperturbed momenta $\hbar\mathbf{k}$ of the "middle" states are not required to differ by integer multiples of $\hbar\mathbf{K}$. Such a superposition can be used to construct a wave packet of any desired form.

We also observe that the average energy of the electron is, by (3.22) and (3.23),

$$\langle E\rangle = E(k) + (dE/dk)_0 \sum nG_n^2 + \tfrac{1}{2}K^2(d^2E/dk^2)_0 \sum n^2G_n^2 = E(k) + 2(g_1g_1' + g_2g_2')\gamma\hbar\omega. \tag{3.27}$$

It is also important to note the following property of this "almost exact" solution. The effect of the laser field is to couple an unperturbed electron state with wave vector $\mathbf{k}$ to states with wave vectors $\mathbf{k} + n\mathbf{K}$ and $\mathbf{k} + 2n\mathbf{K}$, respectively, where n is any positive or negative integer. If we neglect the $(e^2/2mc^2)A^2$ term of the perturbation operator, we find that the amplitudes of these "coupled components" are proportional to $J_n(2g_1')$. The as-

[14] Note that $\mathbf{k} \times (\mathbf{K} \times \mathbf{A}) = \mathbf{K}(\mathbf{k} \cdot \mathbf{A}) - \mathbf{A}(\mathbf{k} \cdot \mathbf{K})$, but there is apparently no contribution to $\mathbf{J}$ arising from the second part of the Lorentz force. There is a simple explanation. The current density can be written as the real part of $\Psi^*[(ie\hbar/m) \operatorname{grad} - (e^2/mc)\mathbf{A}]\Psi$. Therefore, $\mathbf{J} \cdot \hat{\mathbf{a}} = \operatorname{Re}\Psi^*[(ie\hbar/m)\,\hat{\mathbf{a}} \cdot \operatorname{grad} - (e^2/mc)A]\Psi$. But the operator $\hat{\mathbf{a}} \cdot \mathbf{p}$ commutes with H_F, and therefore $\Psi^*\hat{\mathbf{a}} \cdot \operatorname{grad}\Psi$ gives the same result in the presence of the laser field as in its absence. It is also immediately evident from Eq. (2.19) that $\mathbf{J}\cdot\hat{\mathbf{a}} = \mathbf{J}^0\cdot\hat{\mathbf{a}} - (e^2/mc)\mathbf{A}|\Psi|^2$, where the superscript 0 denotes the value of $\mathbf{J}$ when the laser field is not present. Physically this result corresponds to the fact that the presence of a plane wave field does not change the eigenvalues of $\hat{\mathbf{a}}\cdot(-i\hbar \operatorname{grad})$, that is, the transverse (meaning in the plane of the vector potential) components of the electron's wave vector are unchanged by the presence of the field. Here it is important to remember the distinction between $\mathbf{\Pi}$ and $\mathbf{p} = -i\hbar \operatorname{grad}$.

ymptotic behavior of $J_n(2g_1')$ as $n \to +\infty$ through real positive values is given by[15]

$$J_n(2g) \sim (2\pi n)^{-\frac{1}{2}}(eg/n)^n,$$

where e denotes the base of natural logarithms. This means that the amplitudes of the "coupled components" begin to fall off very rapidly for $n > eg_1'$, and the effect of the "coupled components" is essentially negligible for $n \gg g_1'$. A similar result can be obtained when the $(e^2/2mc^2)A^2$ term of the perturbation is not neglected. In other words, the effect of the "coupled components" becomes negligible for $n \gg g_1'$ or g_2', whichever is smaller. This is why the present method of neglecting γn^2 works so well, as demonstrated in Sec. 4. The nature of the exact solution is such that the effect of the "coupled components" with $n \gg g_1$ or g_2 is negligible, but the parameter γ is so small ($\gamma \sim 10^{-6}$) that γn^2 is appreciable in comparison with n only for the "coupled components" which are unimportant. Hence, inclusion of the so-called "kinetic term" corrections does not change our answers very much.

4. EXACT SOLUTION

Let us return to Eq. (2.21), from now on omitting the primes on the g's. In order to satisfy the difference equation (2.21), the generating function $Q(z)$, defined by (3.15), must satisfy the following differential equation:

$$Q'' + (1 - \gamma^{-1})z^{-1}Q' + (1/\gamma) \times [g_1(z^{-1} + z^{-3}) + g_2(1 + z^{-4})]Q = 0. \quad (4.1)$$

It is convenient to eliminate the first derivative by introducing the new generating function

$$R(z) \equiv Q(z) \exp [\tfrac{1}{2}(1 - \gamma^{-1}) \ln z]. \quad (4.2)$$

One can easily verify that $R(z)$ must satisfy

$$R'' + I(z)R = 0, \quad (4.3)$$

where

$$I(z) \equiv (g_1/\gamma)(z^{-1} + z^{-3}) + (g_2/\gamma)(1 + z^{-4}) + (2\gamma)^{-2}(1 - \gamma^2)z^{-2}. \quad (4.4)$$

Furthermore, it is convenient to change the independent variable to

$$y \equiv -i \ln z. \quad (4.5)$$

The differential equation (4.3) then becomes

$$(d^2S/dy^2) + (u/4\gamma^2)S = 0, \quad (4.6)$$

where

$$S(y) = R(z)z^{-\frac{1}{2}} \quad (4.7)$$

and

$$u(y) \equiv 1 - 8\gamma g_1 \cos y - 8\gamma g_2 \cos 2y. \quad (4.8)$$

Note that Eqs. (4.2) and (4.7) imply that

$$Q(z) = S(y) \exp (iy/2\gamma). \quad (4.9)$$

Recalling the relation (3.18) between the wavefunction and the generating function Q, it is clear that any wavefunction given by the product of $\exp \{i[\mathbf{k}\cdot\mathbf{r} - \omega_0 t - (y/2\gamma)]\}$ and $S(y)$ is a valid solution of the present problem, provided we take y equal to $\mathbf{K}\cdot\mathbf{r} - \omega t - \alpha$. (From now on, y always denotes $\mathbf{K}\cdot\mathbf{r} - \omega t - \alpha$.)

The function $S(y)$ is normalized such that

$$\langle S(y)S(-y)\rangle_y = 1, \quad (4.10)$$

where $\langle \cdots \rangle_y$ denotes the average with respect to y; this means that our wavefunction Ψ is normalized such that the average (with respect to $y \equiv \mathbf{K}\cdot\mathbf{r} - \omega t - \alpha$) of $|\Psi|^2$ is equal to unity.

Equation (4.6) is a particular example of Hill's equation.[16] In the two special cases when it is a valid approximation to set either g_1 or g_2 equal to zero, Eq. (4.6) reduces to Mathieu's equation[17] and the exact solution of the Schrödinger equation can be expressed in terms of Mathieu functions. However, it is clear from the following analysis that the Wentzel–Kramers–Brillouin approximation[18] (WKB method) gives a quite adequate solution of (4.6); therefore, the solution in terms of Mathieu functions is not discussed here.

In order to see that the WKB method is a good approximation in this case, we recall that $\gamma \cong 10^{-6}$ and therefore $u(y)/4\gamma^2$ is large and positive for g_1, $g_2 \lesssim 10^5$; also $|u'/u| \ll 1$ for γg_1, $\gamma g_2 \ll 1$. The WKB method gives the following approximate solution of Eq. (4.6):

$$S(y) = au^{-\frac{1}{4}}e^{if(y)} + bu^{-\frac{1}{4}}e^{-if(y)}, \quad (4.11)$$

where[19]

$$f(y) \equiv (1/2\gamma) \int_0^y \{u(y')\}^{\frac{1}{2}}\, dy'. \quad (4.12)$$

[15] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, National Bureau of Standards, Applied Mathematics Series 55 (U. S. Government Printing Office, Washington, D. C., 1964), p. 365, Eq. (9.3.1).

[16] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge University Press, Cambridge, England, 1962), 4th ed., Sec. 19.12, p. 406.

[17] N. W. McLachlan, *Theory and Application of Mathieu Functions* (Clarendon Press, Oxford, England, 1947); J. Meixner and F. W. Schäfke, *Mathieusche Funktionen und Sphäroidfunktionen mit Anwendungen auf physikalische und technische Probleme* (Springer-Verlag, Berlin, 1954).

[18] C. Lanczos, *Linear Differential Operators* (D. Van Nostrand Company, Ltd., London, 1961), Sec. 7.10, pp. 374–376.

[19] In Eqs. (4.12), (4.16), and (4.17), the symbol y' denotes the dummy integration variable which is to be integrated from $y' = 0$ to $y' = y$.

As before, we seek that particular solution of the Schrödinger equation which reduces to $\Psi = \exp[i(\mathbf{k}\cdot\mathbf{r} - \omega_0 t)]$ in the limit $g_1, g_2 \to 0$. Since in this limit, $f(y) \to y/2\gamma$, we must set $a = 0$ in Eq. (4.11):

$$S(y) = bu^{-\frac{1}{4}} \times \exp\left[-(i/2\gamma)\int_0^y \{u(y')\}^{\frac{1}{2}}\,dy'\right], \quad (4.13)$$

where, by (4.10),

$$b^2\langle u^{-\frac{1}{2}}\rangle_v = 1. \quad (4.14)$$

It is clear that an exact solution of Eq. (4.6) can also be written in the same form as (4.13), namely,

$$S(y) = Nv^{-\frac{1}{4}} \exp[-ih(y)], \quad (4.15)$$

where

$$h(y) \equiv (1/2\gamma)\int_0^y \{v(y')\}^{\frac{1}{2}}\,dy', \quad (4.16)$$

and $v(y)$ satisfies

$$v = u\{1 - \gamma^2(v''/uv) + (5/4)\gamma^2(v'^2/uv^2)\}. \quad (4.17)$$

The WKB solution (4.13) neglects $\gamma^2(v''/v)$ and $\gamma^2(v'/v)^2$ compared to u. An improved WKB solution would be obtained if we used the solution (4.15) with

$$v = u\{1 - \gamma^2(u''/u^2) + (5/4)\gamma^2(u'^2/u^3)\}. \quad (4.18)$$

It is clear from (4.18) that the WKB solution (4.13) is correct to first order in γ, since $v = u\{1 + O(\gamma^3 g)\}$ and therefore

$$h(y) = f(y) + O(\gamma^2 g). \quad (4.19)$$

In principle, one can calculate $v(y)$ to any order in γ. For example, the next approximation would be obtained by substitution of the v given by Eq. (4.18) into the right side of Eq. (4.17). It is evident from (4.16) that in order to determine $S(y)$ correct to order γ^n, one must determine $v(y)$ correct to order γ^{n+1}. The final expressions for the wavefunction, current density, and energy of the electron state obtained in this section are all correct to first order in γ.

The exact wavefunction is given by

$$\Psi = N[v(y)]^{-\frac{1}{4}} \times \exp\{i[\mathbf{k}\cdot\mathbf{r} - \omega_0 t + (y/2\gamma) - h(y)]\}. \quad (4.20)$$

The WKB approximation (correct to first order in γ) is

$$\Psi_{\text{WKB}} = [u(y)]^{-\frac{1}{4}} \times \exp\{i[\mathbf{k}\cdot\mathbf{r} - \omega_0 t + (y/2\gamma) - f(y)]\}, \quad (4.21)$$

where

$$f(y) = (y/2\gamma) - 2g_1 \sin y - g_2 \sin 2y - 2\gamma(g_1^2 + g_2^2)y - \gamma[4g_1g_2 \sin y + g_1^2 \sin 2y + \tfrac{4}{3}g_1g_2 \sin 3y + \tfrac{1}{2}g_2^2 \sin 4y] + O(\gamma^2 g^3), \quad (4.22)$$

$$[u(y)]^{-\frac{1}{4}} = 1 + 2\gamma(g_1 \cos y + g_2 \cos 2y) + O(\gamma^2 g^2). \quad (4.23)$$

Here, we have taken $b = 1$, since the normalization condition (4.14) gives $b = 1 + O(\gamma^2 g^2)$. Equation (4.21) may be rewritten in the form

$$\Psi_{\text{WKB}} = \Psi_0 e^{2i\gamma(g_1^2+g_2^2)y}\{1 + 2\gamma(g_1 \cos y + g_2 \cos 2y) + i\gamma[4g_1g_2 \sin y + g_1^2 \sin 2y + \tfrac{4}{3}g_1g_2 \times \sin 3y + \tfrac{1}{2}g_2^2 \sin 4y] + O(\gamma^2 g^2, \gamma^2 g^3), \quad (4.24)$$

where Ψ_0 is the "almost exact" solution given by Eq. (3.24). The probability density is $u^{-\frac{1}{2}} + O(\gamma^2)$, and to first order in γ it is given by

$$|\Psi|^2 = 1 + 4\gamma(g_1 \cos y + g_2 \cos 2y). \quad (4.25)$$

The exact current density (for a free electron), evaluated by using the exact wavefunction (4.20), has the form

$$\mathbf{J} = -(e\hbar/m)\,|N|^2\,\{\mathbf{k} + (1/2\gamma)\mathbf{K}(1 - v^{\frac{1}{2}})\}v^{-\frac{1}{2}} - (e^2/mc)\mathbf{A}\,|N|^2\,v^{-\frac{1}{2}}. \quad (4.26)$$

To first order in γ, we obtain

$$\mathbf{J}_{\text{WKB}} = -(e\hbar/m)\{\mathbf{k} + 2\mathbf{K}(g_1 \cos y + g_2 \cos 2y)\} - 4(e\hbar/m)\gamma(g_1 \cos y + g_2 \cos 2y) \times \{\mathbf{k} + 3\mathbf{K}(g_1 \cos y + g_2 \cos 2y)\} - (e^2/mc)\mathbf{A}\{1 + 4\gamma g_1 \cos y + 4\gamma g_2 \cos 2y\}. \quad (4.27)$$

One can easily demonstrate that the wavefunctions obtained here actually do satisfy the equation of continuity.

From Eqs. (4.24) and (4.27), we see that the "almost exact" solution given in Sec. 3 is a good approximation, provided g_1 and g_2 are small compared to $\gamma^{-\frac{1}{2}} \approx 10^3$. On the other hand, the WKB solution (4.24) is a good approximation provided $g_1, g_2 \ll \gamma^{-1} \approx 10^6$: This inequality for g_2 corresponds to an electric field intensity $\mathcal{E} \ll 10^{11}$ V/cm. However, as we mentioned in the Introduction, relativistic corrections become important when $\gamma g_2 \approx 1$. Therefore there is *no* reason to obtain "better" solutions to the nonrelativistic Schrödinger equation than the usual WKB solution (4.24).

In order to evaluate the average energy of the electron state, we now calculate $\langle n\rangle$ and $\langle n^2\rangle$ using

the WKB solution. It follows from the definition of $Q(z)$ and Eqs. (4.2), (4.5), and (4.7) that $S(y)$ can be written in the form

$$S(y) = \sum_{n=-\infty}^{+\infty} e^{i[n-(1/2\gamma)]y} G_n. \tag{4.28}$$

It therefore follows that

$$S(-y)[-i(d/dy) + (1/2\gamma)]S(y) = \sum_n \sum_m n e^{i(n-m)y} G_m G_n, \tag{4.29}$$

$$S(-y)[-(d^2/dy^2) - (i/\gamma)(d/dy) + (1/4\gamma^2)]S(y) = \sum_n \sum_m n^2 e^{i(n-m)y} G_m G_n. \tag{4.30}$$

Therefore, in order to calculate $\langle n \rangle$ and $\langle n^2 \rangle$ to first order in γ, we calculate the left-hand sides of Eqs. (4.29) and (4.30), using the WKB approximation (4.13) for $S(y)$, then expand the results in powers of γ and then collect the y-independent parts. Thus, we obtain

$$\langle n \rangle = 6\gamma(g_1^2 + g_2^2) + O(\gamma^2 g^3), \tag{4.31}$$

$$\langle n^2 \rangle = 2(g_1^2 + g_2^2) + 24\gamma g_1^2 g_2 + O(\gamma^2 g^4). \tag{4.32}$$

By (4.31) and (4.32), the average energy of an electron state is

$$\langle E \rangle = E(k) + 6\gamma(g_1^2 + g_2^2)K(dE/dk)_0 + K^2(d^2E/dk^2)_0[g_1^2(1 + 12\gamma g_2) + g_2^2]. \tag{4.33}$$

It should be noted that the parameters g_1 and g_2 appearing in Eq. (4.33) are actually g_1' and g_2' defined by Eq. (2.22) (the primes on the g's have been omitted throughout this section).

5. CONCLUSIONS

The WKB solution described in Sec. 4 is an "essentially exact" solution of the nonrelativistic problem. By this statement we mean that it is as "good" a solution as one could desire, since the corrections to it are only important when relativistic effects become important, in which case the Schrödinger equation itself is not applicable.

Furthermore, the results of Sec. 4 indicate that the "almost exact" solution described in Sec. 3 is, in fact, a *very good* approximation.

We wish to stress the fact that the wave function obtained here [see Eqs. (3.24) and (4.24)] differs significantly from the wave function in the absence of the laser field, and the correct expression for the wavefunction is *not* given by low-order perturbation theory. However, the changes of physical quantities (current density and energy) due to the presence of the laser field are just linear and quadratic in the field amplitude, and therefore coincide with what would be obtained by the use of perturbation theory.

This last statement must be qualified somewhat. If we carry the calculations of Sec. 4 to higher order in γ, then the results for J and $\langle E \rangle$ will contain terms of third and higher orders in the field amplitude. However, these terms will not be given correctly, because the Schrödinger equation is not valid in the relativistic region. In other words, if we take the exact solution of the Schrödinger equation and expand the results for J and $\langle E \rangle$ in powers of (γg), then it is clear that only the terms of first order in γ are given correctly. But, since these terms in the expressions for J and $\langle E \rangle$ turn out to be linear and quadratic in the field amplitude, they *are also given correctly* by low-order perturbation theory.

Finally, we wish to point out that, in other quantum problems, it may be helpful to follow the general procedure of Secs. 2 and 3: (1) Go from the usual Schrödinger equation to an infinite set of coupled difference equations; (2) then make appropriate approximations in the difference equations (for example, we dropped γn^2 in comparison with n); (3) go back to a differential equation by introducing the appropriate generating function.[20] From an investigation of the difference equations (which are, of course, equivalent to the Schrödinger equation), one may hope to discover a good approximate solution, whereas the equivalent method of approximation might not be discovered upon examination of the usual form of the Schrödinger equation.

ACKNOWLEDGMENTS

This work was initiated, and a major portion of the work was completed, while the author held a 1962 summer appointment at the IBM Watson Laboratory in New York. The author wishes to thank International Business Machines Corporation for financial support of this project during the summer of 1962. Also, he wishes to thank Dr. Peter J. Price for suggesting this problem, for numerous helpful suggestions and discussions during the course of the work, and for comments on a preliminary version of this article.

[20] After developing the nonperturbation method of solving the Schrödinger equation described in this article, it was brought to the author's attention that the method used here is very similar to a general method for treating large perturbations proposed by H. A. Bethe, Phys. Rev. 54, 955 (1938). The author wishes to thank Dr. M. Lax for bringing this article to his attention.

APPENDIX A

The spin term $\mu_0 \mathfrak{d}\cdot\mathbf{B}$ was not included in the Hamiltonian (2.1). Here, we demonstrate that this term is negligible in comparison with the $(e/mc)\mathbf{A}\cdot\mathbf{p}$ term of the Hamiltonian. To see this, let us take

$$\mathbf{A} = A\cos(\omega t - Kx + \alpha)\mathbf{j},$$

which implies

$$\mathcal{E} = A(\omega/c)\sin(\omega t - Kx + \alpha)\mathbf{j},$$
$$\mathbf{B} = A(\omega/c)\sin(\omega t - Kx + \alpha)\mathbf{k}.$$

Then

$$\mu_0\sigma\cdot\mathbf{B} = (e\hbar/2mc)\sigma_z A(\omega/c)\sin(\omega t - Kx + \alpha),$$
$$(e/mc)\mathbf{A}\cdot\mathbf{p} = (e/c)Av_z\cos(\omega t - Kx + \alpha),$$

where v_z is the component of the unperturbed electron's velocity in the direction of the electric field. Thus, the ratio of these two terms is roughly $r = (\hbar\omega/2mcv_z)$, and hence

$$r^2 = (\hbar\omega)^2/(2mc^2)(2mv_z^2) \cong 10^{-6}/4(\tfrac{1}{2}mv_z^2),$$

where we have assumed that $\hbar\omega = 1$ eV (the kinetic energy of the unperturbed electron should also be expressed in electron volts, so that r will be a pure number). Thus $r^2 \ll 1$, provided we are interested in electrons whose kinetic energies are large compared to 2.5×10^{-7} eV.

APPENDIX B

Here, we present an alternative derivation of Eq. (4.6). The Schrödinger equation is

$$i\hbar\dot{\Psi} = H\Psi, \tag{B1}$$

where

$$H = H_0 + H_F \tag{B2}$$

and H_F is defined by Eqs. (2.5) or (2.6). If we choose the z axis to coincide with the direction of propagation of the Maxwell wave, then $\mathbf{A} = \hat{\mathbf{a}}A(\eta)$, where $\hat{\mathbf{a}}$ is a unit vector in the xy plane and $A(\eta)$ is a real periodic function of $\eta \equiv Kz - \omega t - \alpha$. (Note, η is identical to the quantity denoted by y in Sec. 4; however, we now use a different symbol in order to avoid confusion with the coordinate y.)

Let us look for a solution of the form

$$\Psi(\mathbf{r}, t) = \phi(\eta)u_m(r)e^{-i\omega_m t}, \tag{B3}$$

where $u_m(\mathbf{r})\exp(-i\omega_m t)$ is a solution of the unperturbed problem, and $\phi(\eta)$ must approach unity as $g_1, g_2 \to 0$.

Taking the time derivative of (B3) and multiplying by $i\hbar$, we obtain

$$i\hbar\dot{\Psi} = i\hbar(\omega\phi' - i\omega_m\phi)u_m(\mathbf{r})\exp(-i\omega_m t), \tag{B4}$$

where ϕ' denotes the first derivative of ϕ with respect to its argument η. Furthermore, substitution of (B3) into the right-side of Eq. (B1) yields

$$(H_0 + H_F)\phi(\eta)u_m e^{-i\omega_m t} = [H_0, \phi]u_m e^{-i\omega_m t} + H_0 u_m e^{-i\omega_m t} + (e/mc)\mathbf{A}\cdot\mathbf{p}\phi u_m e^{-i\omega_m t} + (e^2/2mc^2)A^2\phi u_m e^{-i\omega_m t}, \tag{B5}$$

where $[H_0, \phi]$ denotes the commutator. Hence

$$i\hbar\omega\phi' = u_m^{-1}[H_0, \phi]u_m + (e/mc)\phi u_m^{-1}(\mathbf{A}\cdot\mathbf{p})u_m + (e^2/2mc^2)A^2\phi. \tag{B6}$$

Now, let us confine our attention to the free-electron case, $H_0 = (p^2/2m)$. Then

$$[H_0, \phi] = (1/2m)(p_z[p_z, \phi] + [p_z, \phi]p_z) = (i\hbar K/2m)(p_z\phi' + \phi' p_z) = (i\hbar K/2m)(i\hbar K\phi'' + 2\phi' p_z). \tag{B7}$$

The unperturbed free electron eigenfunctions are simply plane waves; let us take

$$u_m(\mathbf{r}) = N\exp(i\mathbf{k}\cdot\mathbf{r}), \tag{B8}$$

where N is the normalization constant. It follows from Eq. (B8) that

$$u_m^{-1}\mathbf{p}u_m = \hbar\mathbf{k}. \tag{B9}$$

Hence, Eq. (B6) becomes

$$i\hbar\omega\phi' = (1/2m)\{-\hbar^2K^2\phi'' + 2i\hbar^2k_z K\phi'\} + (e\hbar/mc)\mathbf{k}\cdot\mathbf{A}\phi + (e^2/2mc^2)A^2\phi. \tag{B10}$$

We can eliminate the first derivative ϕ' by means of the transformation

$$\phi(\eta) = S(\eta)\exp\{-i[(mc^2/\hbar\omega) - (k_z/K)]\eta\}. \tag{B11}$$

One can easily verify that $S(y)$ satisfies the differential equation

$$S'' + \{(mc^2/\hbar\omega)^2 - 2(mc^2/\hbar\omega)(k_z/K) + (k_z/K)^2 - (2e/\hbar\omega)K^{-1}(\mathbf{k}\cdot\mathbf{A}) - (e/\hbar\omega)^2A^2\}S = 0, \tag{B12}$$

which is exactly equivalent to Eq. (4.6).

A New Cluster Scheme in Statistical Physics

G. V. RAMANATHAN*
Plasma Physics Laboratory, Princeton University, Princeton, New Jersey
(Received 3 November 1965)

A new cluster scheme is proposed, and some of its advantages over the Mayer expansion method are demonstrated. An explicit equation is derived for the s-particle correlation in a large equilibrium system of particles interacting through a two-body potential. This equation is solved to the leading order in the plasma limit. It is found that, in several systems, the equilibrium s-particle correlation is, to the leading order, a functional of the two-particle correlation, independent of the detailed form of the potential function.

INTRODUCTION

IT is the purpose of this paper to present a new systematic way of solving the BBGKY hierarchy and to demonstrate some of its advantages.

In the current literature the most widely used method is the Mayer expansion scheme. This expresses the distribution functions in terms of correlation functions in the following manner:

$$\begin{aligned} f_{12} &= f_1 f_2 + g_{12}, \\ f_{123} &= f_1 f_2 f_3 + \sum_{123} f_1 g_{23} + h_{123}, \\ f_{1234} &= f_1 f_2 f_3 f_4 + \sum_{1234} f_1 h_{234} + \sum_{1234} f_1 f_2 g_{34} + l_{1234}, \end{aligned} \tag{0.1}$$

and so on. On substituting these expressions for the distribution functions in the BBGKY hierarchy, one obtains a hierarchy of equations for the correlation functions g, h, l, etc. This hierarchy is usually truncated by neglecting a higher correlation function to obtain a closed system of equations.

This procedure, however, is not necessarily the best way of obtaining a closed system of equations. For instance, Kirkwood and Monroe[1] proposed the superposition approximation for truncating the hierarchy at the two-particle level. This method gave the pressure of a low-density gas correct to the third virial coefficient, and also predicted phase transition in a dense hard-sphere gas. Although Kirkwood's method is not very accurate at high densities, the fact that it showed a phase transition suggests the investigation into the possibilities of finding a better and more general expansion scheme.

In many problems involving three-particle and higher-order effects, it is cumbersome, and sometimes impossible, to use the Mayer expansion scheme. First of all, the derivation of the equations for the three-particle and higher correlations, even at equilibrium, is beset with difficulties of algebraic computation. Second, as an example of the situations in which the Mayer scheme breaks down, one could give the treatment of plasma kinetics when one considers long-wavelength correlations. The usual method predicts that the highest correlation in the hierarchy should decay only by Landau damping, whereas in fact it must be damped by collisional processes, which dominate over the former when the wavelengths are long.

In this paper, a new cluster expansion scheme is proposed and its application to large equilibrium systems is demonstrated.[2] As an illustration, the analysis for a plasma is presented. In this method a new set of functions are defined as correlations. These are related to the distribution functions in terms of product expansions rather than the Mayer-type sum expansions. This is a natural consequence of viewing the correlations in an equilibrium system as potentials of average forces. On substituting the expressions for the distribution functions in the BBGKY hierarchy, one can obtain the equations for the correlation functions. The first distinguishing feature of this method is that, for a large equilibrium system, not only the equation for the s-particle correlation function can be derived explicitly, but the equation can also be solved in various asymptotic limits. The natural and consistent method of ordering these new correlations is illustrated in the plasma limit. It is shown that these correlations are infinitesimally small everywhere in the phase space. This is in contrast to the Mayer expansion scheme in which the s-particle correlation is finite when all the s-particles are close. If the potential

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

[1] J. G. Kirkwood and E. Monroe, J. Chem. Phys. 9, 514 (1941).

[2] This scheme by itself is not entirely new. It appears in the early papers of J. E. Mayer and E. W. Montroll, J. Chem. Phys. 9, 2 (1941), and has been used by R. E. Nettleton and M. S. Green, *ibid.* 29, 1365 (1958). But in this paper, we redefine the correlations and order them in a systematic manner.

between two particles is one of infinitely large repulsion as the distance between them approaches zero, the Mayer scheme gives

$$\left.\frac{g_{12}}{f_1 f_2}\right|_{|\mathbf{x}_1-\mathbf{x}_2|=0} = -1,$$

$$\left.\frac{h_{123}}{f_1 f_2 f_3}\right|_{\substack{|\mathbf{x}_1-\mathbf{x}_2|=0\\|\mathbf{x}_1-\mathbf{x}_3|=0}} = 2,$$

$$\left.\frac{l_{1234}}{f_1 f_2 f_3 f_4}\right|_{\substack{|\mathbf{x}_1-\mathbf{x}_2|=0\\|\mathbf{x}_1-\mathbf{x}_3|=0\\|\mathbf{x}_1-\mathbf{x}_4|=0}} = -3,$$

and so on. The s-particle correlation tends to $(-1)^{s-1}(s-1)$ as all the particle separations tend to zero.

Under this ordering in the plasma limit, an explicit solution is obtained for the equlibrium s-particle correlation function to its leading order. The interesting feature of this result is that it is independent of the detailed form of the potential function. Let us consider a large equilibrium system of particles of average density n, interacting through a two-body potential $\phi(|\mathbf{x}_1 - \mathbf{x}_2|)$. Let us also suppose that in this system there are two distance scales a and b such that $a/b \sim \epsilon \ll 1$, $\phi(a)/kT \sim 1$, $\phi(b)/kT \sim \epsilon$, and $na^3 \sim \epsilon^2$. As long as these conditions are satisfied, it is found that the s-particle correlation is, to its leading order, a functional of the two-particle correlation, independent of the detailed form of the potential function. This result is true for a number of other systems as well, such as dilute gases and weakly interacting systems.

It is believed that this new cluster expansion could give a natural way of truncating the hierarchy for dense gases and liquids with short-range forces. If we denote by r the correlation length for particle separations greater than which the two-particle correlation becomes small, then the highest correlation that will be finite in this scheme will be that of the number of molecules that can be enclosed in a sphere of radius r. All correlations higher than this will be small. Thus there is reason to believe that this scheme could yield a consistent theory of dense gases and liquids. Furthermore, by neglecting the three-particle and higher correlations, Kirkwood's superposition approximation can be recovered from the scheme. This suggests the possibility of application of this method to phase transition problems.

It is also believed that this scheme could be useful in treating a number of nonequilibrium problems as well. For example, in the hierarchy, if the three-particle correlation is neglected, the resulting equation for the two-particle function is nonlinear and contains "collisional"-type terms. These are the terms which give rise to the damping of long-wavelength correlations and presumably also the effect of destroying the divergence due to the correlations spreading in space. It would be fruitful to investigate the possibility of applying this method to the treatment of long-wavelength correlations in plasma kinetics and to obtain convergent higher-order corrections to kinetic equations.

1. PHYSICAL BASIS FOR THE DEFINITION OF CORRELATION FUNCTIONS

A heuristic explanation of the definition of correlation functions can be presented in the following manner. The two-particle distribution function can be written as

$$f_{12} = f_1 f_2 e^{\psi_{12}}. \tag{1.1}$$

At equilibrium ψ_{12} may be looked upon as the potential of the average force between two particles. If the two-particle correlation function α_{12} is defined by

$$f_{12} = f_1 f_2 (1 + \alpha_{12}), \tag{1.2}$$

then

$$1 + \alpha_{12} = e^{\psi_{12}}. \tag{1.3}$$

One may try to extend this concept to three and more particles. Then the three-particle distribution function f_{123} is written as

$$f_{123} = f_1 f_2 f_3 e^{\psi_{12}+\psi_{13}+\psi_{23}+\psi_{123}}. \tag{1.4}$$

Here, a new three-particle term ψ_{123} has been introduced. It may be noted that, without this term, Eq. (1.4) corresponds to the well-known superposition approximation of Kirkwood. Defining

$$1 + \alpha_{123} = e^{\psi_{123}}, \tag{1.5}$$

Eq. (1.4) can be written in the form

$$f_{123} = (f_{12} f_{13} f_{23}/f_1 f_2 f_3)(1 + \alpha_{123}) \tag{1.6}$$

or

$$f_{123} = f_1 f_2 f_3 (1 + \alpha_{12})(1 + \alpha_{13}) \times (1 + \alpha_{23})(1 + \alpha_{123}). \tag{1.7}$$

We call α_{123} the three-particle correlation function and take Eqs. (1.2) and (1.6) as definitions of the two-particle and three-particle correlation functions, respectively. Continuing this approach one may write

$$f_{1234} = \frac{f_{123} f_{124} f_{134} f_{234}}{f_{12} f_{13} f_{14} f_{23} f_{24} f_{34}} f_1 f_2 f_3 f_4 (1 + \alpha_{1234}),$$

and so on. A nice argument in favor of calling these α's correlation functions is due to Dawson.[3] It is based on a variational principle for the entropy. The details of this interpretation will be presented in a later paper.

2. NOTATION AND GENERAL DEFINITIONS

We denote by $\{p\}_r^s$ a generic set of p particles chosen from the collection $(r, r+1, \cdots, s)$ of $s-r+1$ particles. $f_{\{p\}_r^s}$ will stand for the p-particle distribution function of these p particles, and $\alpha_{\{p\}_r^s}$ for the corresponding correlation function. By $f_{a_1,a_2,\cdots,a_n,\{p\}_r^s}$ we mean the $(n+p)$-particle distribution function of the n particles $a_1, a_2, \cdots, a_n$ and the p particles chosen from the collection $(r, r+1, \cdots, s)$. The corresponding correlation function is denoted by $\alpha_{a_1,a_2,\cdots,a_n,\{p\}_r^s}$.

Consider a classical system of N identical particles, each of mass m, in a volume V, interacting through a two-body potential ϕ. We assume the normalization

$$\int f_{\{N\}_1^N} \prod_{i=1}^{N} d^3x_i\, d^3v_i = 1. \tag{2.1}$$

Following Bogoliubov's[4] definition of reduced distribution functions, we write

$$f_{\{s\}_1^s} = V^s \int f_{\{N\}_1^N} \prod_{i=s+1}^{N} d^3x_i\, d^3v_i. \tag{2.2}$$

The correlation function $\alpha_{\{s\}_1^s}$ is defined by the expression

$$f_{\{s\}_1^s} = \frac{\Pi f_{\{s-1\}_1^s}}{\Pi f_{\{s-2\}_1^s}} \frac{\Pi f_{\{s-3\}_1^s}}{\Pi f_{\{s-4\}_1^s}} \cdots \times [\Pi f_{\{1\}_1^s}]^{(-1)^s}(1 + \alpha_{\{s\}_1^s}). \tag{2.3}$$

Here, $\Pi f_{\{p\}_1^s}$ stands for the product of all p-particle distribution functions that can be formed from the s particles $1, 2, \cdots, s$. Specifically this product will contain $\binom{s}{p}$ factors.

An alternative way of writing (2.3) would be

$$f_{\{s\}_1^s} = \Pi f_{\{1\}_1^s} \Pi(1 + \alpha_{\{2\}_1^s}) \Pi(1 + \alpha_{\{3\}_1^s}) \cdots \times \Pi(1 + \alpha_{\{s-1\}_1^s})(1 + \alpha_{\{s\}_1^s}). \tag{2.4}$$

By integrating the Liouville equations $N - s$ times one obtains the sth equation of the BBGKY hierarchy[4]:

$$\left[\frac{\partial}{\partial t} + \sum_{i=1}^{s} \mathbf{v}_i \cdot \frac{\partial}{\partial \mathbf{x}_i} - \frac{1}{m} \sum_{i \neq j=1}^{s} \frac{\partial \phi_{ij}}{\partial \mathbf{x}_i} \cdot \frac{\partial}{\partial \mathbf{v}_i}\right] f_{\{s\}_1^s} = \frac{N-s}{mV} \int d^3x_{s+1}\, d^3v_{s+1} \sum_{i=1}^{s} \frac{\partial \phi_{i,s+1}}{\partial \mathbf{x}_i} \cdot \frac{\partial f_{\{s+1\}_1^{s+1}}}{\partial \mathbf{v}_i}.$$

Dividing both sides by $f_{\{s\}_1^s}$ we get

$$\left[\frac{\partial}{\partial t} + \sum_{i=1}^{s} \mathbf{v}_i \cdot \frac{\partial}{\partial \mathbf{x}_i} - \frac{1}{m} \sum_{i \neq j=1}^{s} \frac{\partial \phi_{ij}}{\partial \mathbf{x}_i} \cdot \frac{\partial}{\partial \mathbf{v}_i}\right] \ln f_{\{s\}_1^s} = \frac{N-s}{mV} \int d^3x_{s+1}\, d^3v_{s+1} \times \sum_{i=1}^{s} \frac{1}{f_{\{s\}_1^s}} \frac{\partial \phi_{i,s+1}}{\partial \mathbf{x}_i} \cdot \frac{\partial}{\partial \mathbf{v}_i} f_{\{s+1\}_1^{s+1}}. \tag{2.5}$$

3. THERMAL EQUILIBRIUM

Henceforth the analysis will be restricted to a large, spatially homogeneous system in thermal equilibrium. We let $N \to \infty$ and $V \to \infty$ keeping $n = N/V$ fixed. At equilibrium,

$$(\partial/\partial t) f_{\{s\}_1^s} = 0, \tag{3.1}$$

$$f_1 = (m/2\pi kT)^{\frac{1}{2}} \exp(-mv_1^2/2kT), \tag{3.2}$$

and we can write

$$f_{\{s\}_1^s} = \Pi f_{\{1\}_1^s} F_{\{s\}_1^s}, \tag{3.3}$$

where $F_{\{s\}_1^s}$ is independent of velocities. It can be seen from (2.3) and (2.4) that

$$F_{\{s\}_1^s} = \frac{\Pi F_{\{s-1\}_1^s}}{\Pi F_{\{s-2\}_1^s}} \frac{\Pi F_{\{s-3\}_1^s}}{\Pi F_{\{s-4\}_1^s}} \cdots \times [\Pi F_{\{2\}_1^s}]^{(-1)^{s-1}}(1 + \alpha_{\{s\}_1^s}) = \Pi(1 + \alpha_{\{2\}_1^s}) \Pi(1 + \alpha_{\{3\}_1^s}) \cdots \times \Pi(1 + \alpha_{\{s-1\}_1^s})(1 + \alpha_{\{s\}_1^s}). \tag{3.4}$$

Substituting (3.1), (3.2), and (3.3) in (2.5), replacing $N - s/V$ by n, and noting that the velocities are all independent, we obtain the following equation for $F_{\{s\}_1^s}$:

$$\frac{\partial}{\partial \mathbf{x}_1} \ln F_{\{s\}_1^s} + \frac{1}{kT} \sum_{i=2}^{s} \frac{\partial \phi_{1i}}{\partial \mathbf{x}_1} = -\frac{n}{kT} \int d^3x_{s+1} \frac{F_{\{s+1\}_1^{s+1}}}{F_{\{s\}_1^s}} \frac{\partial \phi_{1,s+1}}{\partial \mathbf{x}_1}. \tag{3.5}$$

The functions $F_{\{s-1\}_1^{s-1}}$, $F_{\{s-s\}_1^{s-s}}$, etc. obey similar equations. To obtain an equation for $\alpha_{\{s\}_1^s}$, (3.4) has to be substituted in (3.5) and the equations for the lower distribution functions have to be

[3] J. M. Dawson (private communication).

[4] N. N. Bogoliubov, *Studies in Statistical Mechanics*, J. de Boer and G. E. Uhlenbeck, Eds. (North-Holland Publishing Company, Amsterdam, 1962), p. 5.

subtracted. By doing this we get

$$\frac{\partial}{\partial \mathbf{x}_1} \ln (1 + \alpha_{\{s\}_1^s}) + \frac{1}{kT} \sum_{i=2}^{s} \frac{\partial \phi_{1i}}{\partial \mathbf{x}_1} - \frac{1}{kT} \sum_{\substack{\text{all} \\ \{s-2\}_2^s}} \sum_{j \in \{s-2\}_2^s} \frac{\partial \phi_{1j}}{\partial \mathbf{x}_1} + \frac{1}{kT} \sum_{\substack{\text{all} \\ \{s-3\}_2^s}} \sum_{j \in \{s-3\}_2^s} \frac{\partial \phi_{1j}}{\partial \mathbf{x}_1} \cdots + (-1)^s \frac{1}{kT} \sum_{\substack{\text{all} \\ \{1\}_2^s}} \sum_{j \in \{1\}_2^s} \frac{\partial \phi_{1j}}{\partial \mathbf{x}_1}$$

$$= -\frac{n}{kT} \int d^3x_{s+1} \frac{\partial \phi_{1,s+1}}{\partial \mathbf{x}_1} \left[\frac{F_{\{s+1\}_1^{s+1}}}{F_{\{s\}_1^s}} - \sum_{\substack{\text{all} \\ \{s-2\}_2^s}} \frac{F_{1,\{s-2\}_2^s,s+1}}{F_{1,\{s-2\}_2^s}} + \sum_{\substack{\text{all} \\ \{s-3\}_2^s}} \frac{F_{1,\{s-3\}_2^s,s+1}}{F_{1,\{s-3\}_2^s}} \cdots + (-1)^s \sum_{\substack{\text{all} \\ \{1\}_2^s}} \frac{F_{1,\{1\}_2^s,s+1}}{F_{1,\{1\}_2^s}} \right]. \tag{3.6}$$

For $s = 2$,

$$\frac{\sigma}{\partial \mathbf{x}_1} \ln (1 + \alpha_{12}) = -\frac{1}{kT} \frac{\partial \phi_{12}}{\partial \mathbf{x}_1} - \frac{n}{kT} \int d^3x_3 \frac{\partial \phi_{13}}{\partial \mathbf{x}_1} \frac{F_{123}}{F_{12}}$$

$$= -\frac{1}{kT} \frac{\partial \phi_{12}}{\partial \mathbf{x}_1} - \frac{n}{kT} \int d^3x_3 \frac{\partial \phi_{13}}{\partial \mathbf{x}_1} \times (1 + \alpha_{13})(1 + \alpha_{23})(1 + \alpha_{123}). \tag{3.7}$$

For $s > 2$, consider a term $\partial \phi_{1m}/\partial \mathbf{x}_1$ occurring on the left side of the Eq. (3.6), where m is a given number between 2 and s. This term occurs once in the first sum. In the second sum it occurs $\binom{s-2}{s-3}$ times, in the third $\binom{s-2}{s-4}$ times and so on. In the last sum it occurs once. It can be checked that there are $(s - 1)$ sums in all. Therefore, for $s > 2$, the coefficient of the given term $\partial \phi_{1m}/\partial \mathbf{x}_1$ is

$$1 - \binom{s-2}{s-3} + \binom{s-2}{s-4} \cdots (-1)^{s-2} = (1 - 1)^{s-2} = 0.$$

Thus, all the terms of the type $\partial \phi_{1m}/\partial \mathbf{x}_1$ on the left side of (3.6) cancel out. This is a natural consequence of the fact that the particles in the system interact only through a two-body potential.

Now, let us consider the right side of the equation (3.6). A typical quotient there is

$$F_{1,\{p\}_2^s,s+1}/F_{1,\{p\}_2^s}.$$

Using (3.4) this can be expressed as

$$(1 + \alpha_{1,s+1}) \prod_{i=1}^{p} \prod_{\substack{\text{all} \\ \{i\}_1^s}} (1 + \alpha_{\{i\}_1^s,s+1}),$$

where all $\{i\}_1^s$'s are different and when $i = 1$, the particle 1 is not a member of $\{i\}_1^s$. Therefore, in the series on the right side of Eq. (3.6), which, for the sake of consistency, can be written in the form

$$\sum_{\substack{\text{all} \\ \{s-1\}_2^s}} \frac{F_{1,\{s-1\}_2^s,s+1}}{F_{1,\{s-1\}_2^s}} - \sum_{\substack{\text{all} \\ \{s-2\}_2^s}} \frac{F_{1,\{s-2\}_2^s,s+1}}{F_{1,\{s-2\}_2^s}} + \cdots + (-1)^s \sum_{\substack{\text{all} \\ \{1\}_2^s}} \frac{F_{1,\{1\}_2^s,s+1}}{F_{1,\{1\}_2^s}}, \tag{3.8}$$

the general term is

$$(1 + \alpha_{1,s+1}) \prod_{i=1}^{m} \alpha_{\{p_i\}_1^s,s+1}, \tag{3.9}$$

where all $\{p_i\}_1^s$'s are different and when $p_i = 1$, the particle $1 \notin \{p_i\}_1^s$. Let

$$\bigcup_{i=1}^{m} \{p_i\}_1^s = \{p\}_1^s, \qquad p \leq s.$$

We have to find the coefficient of the term (3.9). This term can occur once and only once in every quotient in (3.8) containing all the p particles and nowhere else. We distinguish the following cases:

Case 1: $p < s - 1$. A term of the type

$$(1 + \alpha_{1,s+1}) \prod_{i=1}^{m} \alpha_{\{p_i\}_1^s,s+1}$$

occurs once in the first sum in (3.8), $\binom{s-p-1}{s-p-2}$ times in the second sum, $\binom{s-p-1}{s-p-3}$ times in the third sum, and so on. Therefore the coefficient of such a term is

$$1 - \binom{s-p-1}{s-p-2} + \binom{s-p-1}{s-p-3} + \cdots + (-1)^{s-p-1} = (1 - 1)^{s-p-1} = 0.$$

Case 2: $p = s - 1$. (a) Particle $1 \in \{p\}_1^s$. A term of this type occurs once in the first sum, once in the second term, and no further. So its coefficient is $1 - 1 = 0$. (b) Particle $1 \notin \{p\}_1^s$. A term of this type occurs once only in the first term. So its coefficient is 1.

Case 3: $p = s$. This type of term occurs only once in the first term, so it has a coefficient 1.

Finally a term $(1 + \alpha_{1,s+1})$ occurs once in each sum in (3.8). The coefficient of this term is, therefore,

$$1 - \binom{s-1}{s-2} + \binom{s-1}{s-3} + \cdots + (-1)^s\binom{s-1}{1} = (-1)^s.$$

Thus we get the following equation for $\alpha_{\{s\}_1,s}$:

$$\frac{\partial}{\partial \mathbf{x}_1} \ln (1 + \alpha_{\{s\}_1,s}) = -\frac{n}{kT}\int d^3x_{s+1}\,\frac{\partial \phi_{1,s+1}}{\partial \mathbf{x}_1} \times (1 + \alpha_{1,s+1})\left[(-1)^s + \sum \prod_{i=1}^{m} \alpha_{\{p_i\}_1^*,s+1}\right], \quad (3.10)$$

where the summation goes over all products with the following specifications:

(i) All the $\{p_i\}_1^*$'s occurring in any single product are different.

(ii) If $p_i = 1$, the particle $1 \notin \{p_i\}_1^*$.

(iii) Let $\bigcup_{i=1}^{m} \{p_i\}_1^* = \{p\}_1^*$. Then $p = s - 1$ or s.

(iv) When $p = s - 1$, the particle $1 \notin \{p\}_1^*$.

This means that, for example, when $s = 4$, terms like $\alpha_{125}\alpha_{235}$ or $\alpha_{25}\alpha_{235}$ cannot occur in the equation for α_{1234}.

If the potential is spherically symmetric and vanishes at infinity, then $(-1)^s$ term integrates out and (3.10) can be written in the form

$$\frac{\partial}{\partial \mathbf{x}_1} \ln (1 + \alpha_{\{s\}_1,s}) = -\frac{n}{kT}\int d^3x_{s+1}\,\frac{\partial \phi_{1,s+1}}{\partial \mathbf{x}_1} \times (1 + \alpha_{1,s+1}) \sum \prod_{i=1}^{m} \alpha_{\{p_i\}_1^*,s+1}. \quad (3.11)$$

[This equation is stated without deviation in a paper by Nettleton and Green (see Ref. 2).]

4. ORDERING PROCEDURE

So far no assumption has been made concerning the form of the potential. The only approximation used in the derivation of Eq. (3.10) is that the volume of the system is large. It must be noted that (3.10) represents a hierarchy of equations for the correlation functions. To truncate the hierarchy, one generally studies it in various asymptotic limits, such as dilute systems, systems with weak coupling, and systems with long-range forces. We shall illustrate the technique by obtaining explicit leading-order solutions for the correlations in a plasma. The results do not depend on the detailed nature of the potential and therefore, in general, they are applicable to any system, with two distance scales a and b such that $a/b \sim \epsilon \ll 1$, $\phi(a)/kT \sim 1$, $\phi(b)/kT \sim \epsilon$, and $na^3 \sim \epsilon^2$. The ordering arguments and the asymptotic methods outlined here could also be used to study many other systems with slight modifications.

Let us consider a system of electrons in a uniform background of immobile ions. The α's refer to the correlations between electrons. Since the Coulomb potential is spherically symmetric, we use Eq. (3.11) instead of (3.10). In such a system there is a natural small parameter $\epsilon = 1/n\lambda_D^3 \sim (e^2/kT)\lambda_D \ll 1$, where λ_D is the Debye length and e^2/kT is the distance of closest approach. Using this small parameter, the various terms occurring in each equation of the hierarchy can be estimated. The ordering of the correlations will be done by means of a consistency argument. To start with, it is assumed that all the correlations are nonsingular and are of the order unity or smaller. The best ordering will be found in such a manner as to retain the maximum number of terms at every level consistent with the equations (Kruskal's "Principle of Maximum Balance"[5]). After this it can be checked that all the correlations are indeed nonsingular, consistent with the original assumptions.

Let us first consider the equations for α_{12}, α_{123}, and α_{1234}, successively. For $s = 2$,

$$\frac{\partial}{\partial \mathbf{x}_1}\ln(1 + \alpha_{12}) = -\frac{1}{kT}\frac{\partial \phi_{12}}{\partial \mathbf{x}_1} - \frac{n}{\partial T}\int d^3x_3\,\frac{\partial \phi_{13}}{\partial \mathbf{x}_1}(1 + \alpha_{13}) \times [\alpha_{23} + \alpha_{123} + \alpha_{23}\alpha_{123}]. \quad (4.1)$$

Let $\mathbf{R} = (\mathbf{x}_1 - \mathbf{x}_2)$ and $\mathbf{r}_0 = (\mathbf{x}_1 - \mathbf{x}_3)$, the latter being the range of the integration. We shall assume that all correlations are nonsingular and are of the order unity or smaller. The relative orders of the various terms in (4.1) can be written as

$$\ln [1 + \alpha(\mathbf{R})]: e^2/kT/R: (nr_0^3)[(e^2/kT)/r_0]\alpha(\mathbf{r}_0 - \mathbf{R}): (nr_0^3)[(e^2/kT)/r_0]\alpha[\mathbf{r}_0, \mathbf{R}]: (nr_0^3)[(e^2/kT)/r_0]\alpha(\mathbf{r}_0 - \mathbf{R})\alpha[\mathbf{r}_0, \mathbf{R}]. \quad (4.2)$$

Here r_0 and R stand for the scalar distances and $\alpha[\mathbf{r}_0, \mathbf{R}]$ for the three-particle correlation. Consider the quantity $(nr_0^3)[(e^2/kT)/r_0]\alpha(\mathbf{r}_0 - \mathbf{R})$.

(i) When $R \sim \lambda_D$ and $r_0 \sim \epsilon\lambda_D$, this quantity is of the order $\epsilon^2\alpha(\lambda_D)$. When $R \sim \lambda_D$ and $r_0 \sim \lambda_D$, it is of the order $\alpha(\lambda_D)$. Therefore, the dominant contribution to this integral comes from the region $|\mathbf{x}_1 - \mathbf{x}_3| \sim \lambda_D$ and this term is of the order $\alpha(\lambda_D)$. In a similar manner the other terms in Eq. (4.1) can be estimated. When $|\mathbf{x}_1 - \mathbf{x}_2| \sim \lambda_D$, the quantities in (4.2) take the form

$$\ln [1 + \alpha(\lambda_D)]: \epsilon : \alpha(\lambda_D): \alpha[\lambda_D, \lambda_D]: \alpha(\lambda_D)\alpha[\lambda_D, \lambda_D]. \quad (4.3)$$

[5] M. D. Kruskal, *Mathematical Models in Physical Sciences*, S. Drobot and P. A. Viebrock, Eds. (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1963).

If we anticipate that the three-particle correlation is going to be expressed in terms of the two-particle correlation in such a manner that if α_{12} vanishes identically, so would α_{123}, then we can look upon the second term in (4.3) as a forcing term. If we require that α_{12} should be so ordered as to retain the maximum number of terms including the forcing term, then $\alpha(\lambda_D) \sim \epsilon$.

(ii) When $R \sim \epsilon\lambda_D$ and $r_0 \sim \epsilon\lambda_D$,

$$(nr_0^3)[(e^2/kT)/r_0]\alpha(\mathbf{r}_0 - \mathbf{R}) \sim \epsilon^2\alpha(\epsilon\lambda_D).$$

When $R \sim \epsilon\lambda_D$ and $r_0 \sim \lambda_D$, it is of the order of $\alpha(\lambda_D) \sim \epsilon$. Here again the main contribution of the integral term comes from a range $r_0 \sim \lambda_D$. Similar estimates can be made of the other terms to obtain the following:

$$\ln [1 + \alpha(\epsilon\lambda_D)]: 1 : \alpha(\lambda_D):$$

$$\alpha[\lambda_D, \lambda_D]: \alpha(\lambda_D)\alpha[\lambda_D, \lambda_D].$$

Again, to retain the maximum number of terms we have to take $\alpha(\epsilon\lambda_D) \sim 1$.

The equation α_{123} is

$$\frac{\partial}{\partial \mathbf{x}_1} \ln (1 + \alpha_{123}) = -\frac{n}{kT}\int d^3x_4 \frac{\partial\phi_{14}}{\partial x_1}(1 + \alpha_{14}) \times [\alpha_{24}\alpha_{34} + \alpha_{234} + \alpha_{24}\alpha_{234} + \alpha_{34}\alpha_{234} + \alpha_{24}\alpha_{34}\alpha_{234} + \alpha_{1234} + \alpha_{24}\alpha_{1234} + \cdots]. \quad (4.4)$$

The relative orders of magnitude of the various terms can be written as

$$\ln [1 + \alpha(\mathbf{R}_1, \mathbf{R}_2)]: (nr_0^3)\frac{e^2/kT}{r_0}\alpha(\mathbf{R}_1 - \mathbf{r}_0)\alpha(\mathbf{R}_2 - \mathbf{r}_0):$$

$$(nr_0^3)\frac{e^2/kT}{r_0}\alpha[\mathbf{R}_1 - \mathbf{r}_0, \mathbf{R}_2 - \mathbf{r}_0]:$$

$$(nr_0^3)\frac{e^2/kT}{r_0}\alpha(\mathbf{R}_1 - \mathbf{r}_0)\alpha(\mathbf{R}_1 - \mathbf{r}_0, \mathbf{R}_2 - \mathbf{r}_0): \cdots :$$

$$(nr_0^3)\frac{e^2/kT}{r_0}\alpha\{\mathbf{r}_0, \mathbf{R}_1, \mathbf{R}_2\} \cdots .$$

Now, we anticipate that α_{1234} is going to be expressed in terms of α_{12} and α_{123} in such a manner that if α_{12} vanishes identically, so would α_{1234}. Here it must be noted that in (4.4) if α_{12} vanishes identically, then there is no forcing term. If we further anticipate that the contribution from the four-particle term would be much smaller than the contribution from the second term, then the second term would be the main forcing term. By arguments along the same lines as for the two-particle function, it can be shown that, to retain the maximum number of terms including the forcing term in the three-particle equation, we must have

$$\alpha_{123} \sim \frac{n}{kT}\int d^3x_4\, \phi_{14}(1 + \alpha_{14})\alpha_{24}\alpha_{34},$$

and continuing this argument up to the s-particle function, we must have

$$\alpha_{\{s\}_1^s} \sim \frac{n}{kT}\int d^3x_{s+1}\phi_{1,s+1}(1 + \alpha_{1,s+1}) \times \alpha_{2,s+1}\alpha_{3,s+1} \cdots \alpha_{s,s+1}. \quad (4.5)$$

This means that, if every pair of the s-particles is separated by a distance of the order of $\epsilon\lambda_D$, then $\alpha_{\{s\}_1^s} \sim \epsilon^2$, and when every pair is separated by a distance of the order of λ_D, $\alpha_{\{s\}_1^s} \sim \epsilon^{s-1}$. Thus the three-particle and higher correlation functions are infinitesimally small everywhere in phase space.

To sum up, the ordering argument is as follows: (1) We are trying to obtain a solution for every correlation function in terms of the lower correlation functions. (2) We assume *a priori* that all the correlations are nonsingular and are of the order unity or smaller. (3) Then, if in the equation for each correlation the maximum number of terms including the forcing terms are to be retained, the best ordering consistent with the equations is the one given by (4.5). (4) The leading-order solution obtained in the next section provides a verification of the *a priori* assumptions.

5. LEADING-ORDER SOLUTION FOR $\alpha_{\{s\}}$

We now obtain uniformly valid leading-order solutions for all correlations in an equilibrium plasma. By this we mean that the solution for $\alpha_{\{s\}_1^s}$ is given by a single expression valid everywhere in phase space such that at every point the value computed from this expression will be correct to the order of $\alpha_{\{s\}_1^s}$ itself at that point, i.e., the relative error will be small. Thus when all the particle separations are of the order $\epsilon\lambda_D$, the expression will be correct to the order ϵ^2, and when all the particle separations are of the order λ_D, it will be correct to the order ϵ^{s-1}.

Now, we recall (4.5) and retain only terms of the order of

$$n/kT\int d^3x_{s+1}\phi_{1,s+1}(1 + \alpha_{1,s+1})\prod_{i=2}^{s}\alpha_{i,s+1}$$

in Eq. (3.11). Consider a term

$$n/kT\int d^3x_{s+1}\,\partial\phi_{1,s+1}/\partial\mathbf{x}_1(1 + \alpha_{1,s+1})\prod_{i=1}^{m}\alpha_{\{p_i\}_1^s,s+1}$$

occurring on the right side of this equation. If any particle is repeated in the product or if particle $1 \in \{p_i\}_1^s$, then it would imply that

$$\prod_{i=1}^{m} \alpha_{\{p_i\}_1^s, s+1} \ll \prod_{i=2}^{m} \alpha_{i, s+1}.$$

For example, when $s = 4$, $\alpha_{125} \ll \alpha_{25}$, and $\alpha_{25}\alpha_{235} \ll \alpha_{25}\alpha_{35}$. This is due to the fact that three-particle and higher correlations are all of the order ϵ^2 or smaller everywhere. Thus, all the products containing repetitive indices or the index 1 can be neglected in the leading-order equation for $\alpha_{\{s\}_1^s}$. Also, for $s > 2$, we can expand the logarithm on the left side of Eq. (3.11) and we need to keep only the first term. Furthermore, if we retain only leading-order terms in this equation, there will be just one s-particle term on the right side. Transposing this to the left side, we write the equation for $\alpha_{\{s\}_1^s}(s > 2)$ in the form

$$\frac{\partial}{\partial \mathbf{x}_1} \alpha_{\{s\}_1^s} + \frac{n}{kT} \int d^3x_{s+1} \frac{\partial \phi_{1,s+1}}{\partial \mathbf{x}_1} \alpha_{\{s-1\}_2^s, s+1}$$
$$= -\frac{n}{kT} \int d^3x_{s+1} \frac{\partial \phi_{1,s+1}}{\partial \mathbf{x}_1} (1 + \alpha_{1,s+1})$$
$$\times \sum \prod_{i=1}^{m} \alpha_{\{t_i\}_2^s, s+1} + \frac{\partial}{\partial \mathbf{x}_1} O(\epsilon \alpha_{\{s\}_1^s}). \quad (5.1)$$

Now, the summation goes over all products with the following specifications: (i) In any one product all $\{t_i\}_2^s$'s are different; (ii) $t_i < s - 1$ for any i; (iii) $\bigcup_{i=1}^{m} \{t_i\}_2^s = \{s - 1\}_2^s$; and (iv) $\{t_i\}_2^s \bigcap \{t_j\}_2^s = \{0\}_2^s$, the null set for any $i, j, i \neq j$. This means that, for example, when $s = 4$, terms like $\alpha_{235}\alpha_{245}$ cannot occur in the equation for α_{1234}. Here, it must be noticed that to obtain $\alpha_{\{s\}_1^s}$ to its leading order, all the lower correlation functions appearing on the right side of (5.1) need be known only to their leading orders.

The two-particle correlation function $\alpha_{1,s+1}$ satisfies the following equation:

$$\frac{\partial}{\partial \mathbf{x}_1} \ln (1 + \alpha_{1,s+1}) = -\frac{1}{kT} \frac{\partial \phi_{1,s+1}}{\partial \mathbf{x}_1}$$
$$- \frac{n}{kT} \int d^3x_{s+2} \frac{\partial \phi_{1,s+2}}{\partial \mathbf{x}_1} (1 + \alpha_{1,s+2})$$
$$\times (1 + \alpha_{s+1,s+2})(1 + \alpha_{1,s+1,s+2}). \quad (5.2)$$

Differentiating the logarithm and multiplying both sides by $1 + \alpha_{1,s+1}$, this can be written as

$$\frac{\partial}{\partial \mathbf{x}_1} \alpha_{1,s+1} = -\frac{1}{kT} (1 + \alpha_{1,s+1}) \frac{\partial \phi_{1,s+1}}{\partial \mathbf{x}_1} - \frac{n}{kT}$$
$$\times \int d^3x_{s+2} \frac{\partial \phi_{1,s+2}}{\partial \mathbf{x}_1} \alpha_{s+1,s+2} + \frac{\partial}{\partial \mathbf{x}_1} O(\epsilon \alpha_{1,s+1}). \quad (5.3)$$

We have made use of the fact that

$$\frac{n}{kT} \int d^3x_{s+2} \phi_{1,s+2} \alpha_{s+1,s+2} \sim (nr_0^3) \frac{e^2/kT}{r_0} \alpha(\mathbf{r} - \mathbf{R}_0) \sim \epsilon.$$

Substituting (5.3) on the right side of (5.1) and rearranging terms, we obtain to the leading order in $\alpha_{\{s\}_1^s}$,

$$\frac{\partial}{\partial \mathbf{x}_1} \left[\alpha_{\{s\}_1^s} - n \int d^3x_{s+1} \alpha_{1,s+1} \sum \prod_{i=1}^{m} \alpha_{\{t_i\}_2^s, s+1} \right]$$
$$+ \frac{n}{kT} \int d^3x_{s+1} \frac{\partial \phi_{1,s+1}}{\partial \mathbf{x}_1} \left[\alpha_{\{s\}_2^{s+1}} - n \right.$$
$$\left. \times \int d^3x_{s+2} \alpha_{s+1,s+2} \sum \prod_{i=1}^{m} \alpha_{\{t_i\}_1^s, s+2} \right] = 0. \quad (5.4)$$

If the α's are supposed to be nonsingular everywhere, then the solution to this equation is

$$\alpha_{\{s\}_1^s} = n \int d^3x_{s+1} \alpha_{1,s+1}$$
$$\times \sum \prod_{i=1}^{m} \alpha_{\{t_i\}_2^s, s+1} + O(\epsilon \alpha_{\{s\}_1^s}). \quad (5.5)$$

The summation goes over all products with the following specifications:

(i) all the $\{t_i\}_2^s$'s are different;
(ii) $t_i < s - 1$ for any i;
(iii) $\bigcup_{i=1}^{m} \{t_i\}_2^s = \{s - 1\}_2^s$; and
(iv) $\{t_i\}_2^s \bigcap \{t_j\}_2^s = \{0\}_2^s$, the null set for any $i, j, i \neq j$ in any single product. For $s = 3$,

$$\alpha_{123} = n \int d^3x_4 \alpha_{14}\alpha_{24}\alpha_{34} + O(\epsilon \alpha_{123}), \quad (5.6)$$

$$\alpha_{1234} = n \int d^3x_5 \alpha_{15} [\alpha_{25}\alpha_{35}\alpha_{45} + \alpha_{25}\alpha_{345}$$
$$+ \alpha_{35}\alpha_{245} + \alpha_{45}\alpha_{235}] + O(\epsilon \alpha_{1234}). \quad (5.7)$$

The symmetry of the expression (5.7) under the interchange of 1 and 2 can be easily verified by substituting for the three-particle functions from (5.6). With a little bit of algebra, the symmetry of $\alpha_{\{s\}_1^s}$ given by (5.5) can also be checked.

It must be noted that, in obtaining the solution (5.5), no specific form was assumed for the potential function. Thus, in any large system of particles interacting through a two-body potential, if there are two distance scales a and b (e^2/kT and λ_D for a plasma) such that $a/b \sim \epsilon \ll 1$, $\phi(a)/kT \sim \epsilon$, and $na^3 \sim \epsilon^2$, then the result (5.5) must be true for that system.

6. OTHER POSSIBLE APPLICATIONS OF THIS THEORY

It would be fruitful to investigate the possible applications of this theory to the statistical mechanics of dense gases and liquids. The equation (3.11) for the s-particle correlation is quite general and is applicable to any large system of particles interacting through a two-body force. Indeed, this is the sth equation of a hierarchy which has to be truncated in some way to obtain a closed chain of equations. For a dense gas or a liquid the hierarchy represented by (3.11) can presumably be truncated in the following way. Let us first suppose that the successive α's do not grow in order of magnitude. Then, from Eq. (3.10) it can be conjectured that

$$\alpha_{(s)_1^s} \sim n \int d^3x_{s+1}\alpha_{1,s+1}\alpha_{2,s+1} \cdots \alpha_{s,s+1}. \qquad (6.1)$$

Suppose that the two-particle correlation is finite up to distances of the order of r_0 beyond which it becomes small. For simple liquids, r_0 is of the order of 2 or 3 molecular diameters. Clearly, if (6.1) is true, then the highest correlation that will be finite will be that of the number of particles that can be enclosed by a sphere of radius r_0. All correlations higher then this will be successively smaller. Thus, a legitmate method of truncating the hierarchy may be found.

It is also believed that this new scheme may prove to be useful in handling a number of nonequilibrium problems. For instance, the equations obtained for the α's are nonlinear and so enable one to study the collisional damping of long-wavelength correlations in a plasma and to obtain convergent higher-order corrections to kinetic equations.

ACKNOWLEDGMENTS

The author expresses his deep gratitude to Professor M. D. Kruskal and Professor J. M. Dawson for their guidance and help throughout this work. He also thanks Professor E. A. Frieman and Professor A. Lenard for valuable discussions on this topic. The author is grateful to the referee for his comments.

Finally, this work was performed under the auspices of the U. S. Atomic Energy Commission, Contract No. AT(30-1)-1238.

Asymptotic Properties of the Potentials in the Inverse-Scattering Problem at Fixed Energy

PIERRE C. SABATIER
Laboratoire de Physique Théorique et Hautes Energies, Faculté des Sciences, Orsay, France
(Received 17 January 1966)

The problem of finding a nonrelativistic central potential from a knowledge of all the phase shifts at one energy had been previously shown by Newton to reduce to the inversion of a given infinite matrix M. In the framework of Newton's theory, the solution is not unique but depends on one parameter. In the present work, the inverse matrix is explicitly given, together with the vectors annihilated by M. These enable one to construct all the solutions of the problem. The asymptotic behavior of the equivalent potentials is exhibited, and it is shown that one (and only one) of them decreases asymtotically faster than $r^{-2+\epsilon}$, provided that the phase shifts decrease asymptotically faster than $l^{-(3+\epsilon')}$ (for arbitrarily small ϵ, ϵ'). All the other equivalent potentials have an oscillating tail damped by a factor $r^{-3/2}$. The "transparent potentials," which give all phase shifts equal to zero at one energy, are also studied. In subsequent publications, analytic continuation of the potentials in the r plane and of the Jost function in the angular momentum complex plane is studied.

INTRODUCTION

THE problem of finding the nonrelativistic potential from a knowledge of all the phase shifts at one energy is of "obvious physical importance as well as intrinsic interest".[1] It has been treated successively by Wheeler,[2] by Regge,[3] by Martin and Targonski,[4] and by Newton.[1] The last two papers give a complete treatment of the problem, but the method of Martin and Targonski applies only to superpositions of Yukawa potentials. The method of Newton is more general. A consequence of this generality is the lack of uniqueness of the solution. A very important tool in Newton's method is an infinite matrix M, the inversion of which gives the key to the solution. From the study of an auxiliary matrix N, Newton showed that it is possible to build an inverse matrix of M, and also that there necessarily exists at least a column vector $\mathbf{v}$ which is annihilated by M, so that the equation

$$M\mathbf{a} = \mathbf{b}$$

has an infinite number of solutions:

$$\mathbf{b} = M^{-1}\mathbf{a} + \alpha\mathbf{v}.$$

The knowledge of any vector like $\mathbf{v}$ enables one to build a nontrivial central potential which leads to phase shifts equal to zero for any value of l.

It has been shown by Redmond[5] that $\mathbf{v}$ is unique, so that the infinity of potentials equivalent to a given one (at a fixed energy) depends only on one parameter.

In the present paper, we first construct explicitly both M^{-1} and $\mathbf{v}$. These results enable us to study the asymptotic behavior of the potentials leading to a given set of phase shifts. In particular, we show that, if the phase shifts go to zero faster than l^{-3} as l goes to infinity, there exists one potential, and only one, which goes to zero, as $r \to \infty$, faster than $r^{-2+\epsilon}$,[6] and that all the potentials equivalent to this one have an oscillating tail which is damped by a factor $r^{-\frac{1}{2}}$.

In the present paper, we limit ourselves to the asymptotic properties of the potentials. The analytic properties of the potentials in the complex r plane and the analytic properties of the phase shifts in the angular momentum plane will be the subject of a forthcoming communication.[7]

1. PRELIMINARY PHYSICAL AND MATHEMATICAL STUDIES

1.1. A Survey of Newton's Method

It seems necessary to introduce in this survey notations and formulas which are of particular use in Sec. 3.

Newton starts with a given function $f(r, r')$ defined by the infinite series

$$f(r, r') = \sum_{l=0}^{\infty} c_l u_l(r) u_l(r') \qquad (1.1)$$

[1] R. G. Newton, J. Math. Phys. 3, 75 (1962).
[2] J. A. Wheeler, Phys. Rev. 99, 630 (1955).
[3] T. Regge, Nuovo Cimento 14, 951 (1959).
[4] A. Martin and G. Y. Targonski, Nuovo Cimento 20, 1182 (1961).
[5] P. J. Redmond, J. Math. Phys. 5, 1547 (1964).
[6] Throughout this paper, by ϵ we mean a positive number which can be made arbitrarily small, but not equal to zero. It is not meant to have the same value every time it is used, even inside a given formula.
[7] P. Sabatier (to be published).

with real coefficients c_l, where u_l are the regular spherical Ricatti–Bessel functions:

$$u_l(r) = (\tfrac{1}{2}\pi r)^{\frac{1}{2}} J_{l+\frac{1}{2}}(r). \tag{1.2}$$

The radial distance r is measured in units of λ, the reduced wavelength of the relative motion, which is fixed throughout the following. The differential equation satisfied by u_l is written as

$$D_0(r)u_l(r) = l(l+1)u_l(r) \tag{1.3}$$

with the differential operator

$$D_0(r) \equiv r^2(\partial^2/\partial r^2 + 1). \tag{1.4}$$

From $f(r, r')$, Newton defines a function $K(r, r')$ as the solution of the following Fredholm equation:

$$K(r, r') = f(r, r') - \int_0^r dr_1\, r_1^{-2} K(r, r_1) f(r_1, r'). \tag{1.5}$$

He proves that $K(r, r')$ satisfies the following partial differential equation and boundary conditions:

$$\begin{aligned} D(r)K(r, r') &= D_0(r')K(r, r'), \\ K(r, 0) &= K(0, r') = 0, \end{aligned} \tag{1.6}$$

where $D(r)$ is defined according to

$$D(r) \equiv D_0(r) - r^2 V(r), \tag{1.7}$$

$$V(r) = -2r^{-1}(d/dr)[r^{-1}K(r, r)]. \tag{1.8}$$

$K(r, r')$ is used to define the function

$$\phi_l(r) = u_l(r) - \int_0^r dr'\, r'^{-2} K(r, r') u_l(r'). \tag{1.9}$$

Application of the differential operator $D(r)$ to (1.9) together with two integrations by parts and use of (1.8) and (1.2) show that $\phi_l(r)$ satisfies the differential equation

$$D(r)\phi_l(r) = l(l+1)\phi_l(r), \tag{1.10}$$

and it follows from (1.6) that ϕ_l is the regular solution of (1.10); i.e.,

$$\phi_l(0) = 0. \tag{1.11}$$

Using (1.1), (1.5), and (1.9), Newton proves the following result:

$$K(r, r') = \sum_{l=0}^{\infty} c_l \phi_l(r) u_l(r'). \tag{1.12}$$

Substituting this result in (1.9), he obtains the following set of coupled linear algebraic equations, equivalent to the integral equation (1.5):

$$\phi_l(r) = u_l(r) - \sum_{l'} L_{ll'}(r) c_{l'} \phi_{l'}(r), \tag{1.13}$$

where

$$L_{ll'}(r) = \int_0^r dr'\, r'^{-2} u_l(r') u_{l'}(r'). \tag{1.14}$$

For $r \to \infty$, the functions involved in (1.13) have the following asymptotic behavior:

$$\begin{aligned} \phi_l(r) &\sim A_l \sin(r - \tfrac{1}{2}l\pi + \delta_l), \\ u_l(r) &\sim \sin(r - \tfrac{1}{2}l\pi), \\ L_{ll'}(r) &\sim L_{ll'}(\infty). \end{aligned} \tag{1.15}$$

Inserting these relations in (1.13) one obtains the fundamental formulas:

$$\sin \delta_l = \sum_{l'} M_l^{l'} b_{l'} \cos(\delta_{l'} - \delta_l), \tag{1.16}$$

$$A_l = \cos \delta_l - \tfrac{1}{2}\pi b_l/(2l+1) - \sum_{l'} M_l^{l'} b_{l'} \sin(\delta_{l'} - \delta_l), \tag{1.17}$$

where $M_l^{l'}$ and b_l are defined as follows:

$$M_l^{l'} = \begin{cases} i^{l-l'+1} L_l^{l'}(\infty), & \text{for } l \neq l', \\ 0, & \text{for } l = l', \end{cases} \tag{1.18}$$

$$b_l = c_l A_l. \tag{1.19}$$

From (1.16), it is easy to obtain the simpler equation

$$\tan \delta_l = \sum_{l'} M_l^{l'} a_{l'} (1 + \tan \delta_l \tan \delta_{l'}), \tag{1.20}$$

where

$$a_l = b_l \cos \delta_l. \tag{1.21}$$

Inversion of the matrix M yields a_l, and therefore b_l. Insertion of the result in (1.17) then explicitly gives A_l, and (1.19) yields the desired c_l.

1.2. Summability and Bounds[8] for Series of Bessel Functions

In this section we give some properties of the Bessel series or integrals which are of use in the following.

1.2.1. Uniform Convergence of Bessel Series

We deal with series of Ricatti–Bessel functions. These functions are defined as in (1.2), where the index l is now allowed to take any real value. The series may belong to either of the two following types:

$$S(x) = \sum_{0}^{\infty} p_n u_{n+\beta}(x), \tag{1.22}$$

[8] Throughout this paper, we mean by "bound" an upper bound for the absolute value.

$$h(r, r') = \sum_0^\infty a_n u_n(r) u_n(r'). \tag{1.23}$$

In the following we assume that $|p_n|$ and $|a_n|$ have a definite bound independent of n. Owing to the bound (A5), $S(x)$ converges uniformly for any finite real or complex value of x, and defines an entire function of x. In the same way, the series h defines an entire function of x (respectively x') for any finite value of x' (respectively x).

1.2.2. Asymptotic Behavior and Bounds of Series $h(r, r')$

We make a distinction between the following classes of series, according to the asymptotic behavior of the coefficients, which characterizes their own asymptotic behavior,[9] as given in the following table[10]:

Class 0. $a_{2n+1} = 0$, $a_{2n} \sim 1 + O(n^{-5/3-\epsilon})$,

$$h_0(r, r') \sim \tfrac{1}{2}(rr')^{\frac{1}{2}} \times [J_0(r - r') + \mathbf{H}_0(r + r') + O(r^{-\frac{1}{2}} r'^{-\frac{1}{2}})], \tag{1.24}$$

where $\mathbf{H}$ is the Struve function.

Class $\bar{0}$. $a_{2n} = 0$, $a_{2n+1} \sim 1 + O(n^{-5/3-\epsilon})$,

$$h_{\bar{0}}(r, r') \sim \tfrac{1}{2}(rr')^{\frac{1}{2}} \times [J_0(r - r') - \mathbf{H}_0(r + r') + O(r^{-\frac{1}{2}} r'^{-\frac{1}{2}})]. \tag{1.25}$$

Class 1. $a_n \sim n^{-1} + O(n^{-5/3-\epsilon})$,

$$|h_1(r, r')| \lesssim C(\log r)^{\frac{1}{2}}(\log r')^{\frac{1}{2}} + C \quad \text{for } r \text{ and } r' > 1. \tag{1.26}$$

Class 2. $a_n \sim O(n^{-5/3-\epsilon})$,

$$h_2(r, r') \sim C_1 \cos(r - r') + C_2 \cos(r + r') + O(r^{-3/8\epsilon}), \quad \text{for } r \to \infty, \quad r' > r. \tag{1.27}$$

The asymptotic behavior of the class 2 series results readily from the asymptotic behavior of the Bessel functions (A1). The leading term in the remainders comes from the values of p which lie between $\epsilon_0^{\frac{1}{2}} r$ and r, and from the remainders ϵ_0^2 for p lying between 0 and $\epsilon_0^{\frac{1}{2}} r$ [see (A1) and (A2)]. The remainders in (1.27) can then be estimated, by putting $\epsilon_0 = r^{-\frac{1}{2}\epsilon}$.

The series 0, $\bar{0}$, and 1 are defined up to a translation of a series $h_2(r, r')$. Their behavior and bounds follow from the typical series T_0, $T_{\bar{0}}$, and T_1 studied in Appendix B. The series of class 2 can be given a definite bound independent of r and r'. We say of any series with this property that it belongs to class B: it is clear that $2 \subset B$. For series with coefficients decreasing more slowly than the coefficients of class 2, but which do not have the regularity required in class 0 and 1, it is still possible to find bounds with the help of the Bessel functions recurrence[11] relations:

$$2J_\nu(Z) = Z\nu^{-1}[J_{\nu-1}(Z) + J_{\nu+1}(Z)]. \tag{1.28}$$

For instance, the following bound:

$$\sum_0^\infty a_n J_n(Z) J_n(Z') < CZ \tag{1.29}$$

holds for any value of Z', provided that $(n^{\frac{1}{2}+\epsilon} a_n)$ is bounded when $n \to \infty$.

1.2.3. Summability of Double Series

We encounter double series of Bessel functions:

$$\sum_n a_n u_n(x) \sum_p b_n^p u_p(x'). \tag{1.30}$$

The summability of this results from the summability of the double series:

$$\sum_n |a_n| n^{\frac{1}{2}} \sum_p |b_n^p| p^{\frac{1}{2}} \tag{1.31}$$

for which the (weak) reciprocal of the Fubini theorem applies. In the course of these computations, a special type of series will occur frequently; they can be studied on the following general scheme:

$$|S| = \sum_q \frac{|a_q|}{|(r + \alpha)^2 - (q + \gamma)^2|}, \quad r \text{ is an integer}, \quad \alpha \neq \gamma, \quad \alpha < 1, \quad \gamma < 1,$$

and

$$a_q \sim q^{-\beta} \quad \text{as} \quad q \to \infty. \tag{1.32}$$

It is shown in Appendix A that

$$|S| < C_1 r^{-1-\beta+\epsilon} + C_2 r^{-2}. \tag{1.33}$$

1.2.4. Summability of Integrals

We are concerned with two kinds of summability problems corresponding to integrals of Bessel series:

(1) Let us first study the following integral:

$$I = \int_0^\infty u_{\lambda+\beta}(\rho) \rho^{-2} \, d\rho \, S(\rho), \tag{1.34}$$

in which $S(\rho)$ is defined as in (1.22), $\lambda \geq 0$, $\beta > -1$.

[9] In the following, we use C as a general constant. It is not meant to have the same value every time it is used.

[10] The symbol $O(f)$ means $< Cf$.

[11] Bateman Manuscript Project, *Higher Transcendental Functions*, A. Edélryi, Ed. (McGraw-Hill Book Company, Inc., New York, 1953).

We assert that it is possible to perform a term by term integration and obtain

$$I = \sum_0^\infty p_n \int_0^\infty u_{\lambda+\beta}(\rho)u_{n+\beta}(\rho)\rho^{-2}\,d\rho, \qquad (1.35)$$

provided that $|p_n| < n^{1-\epsilon}$ and I converge in the form (1.34).

Proof: We divide the path of integration into two parts $(0, r)$ and (r, ∞). On the first path the series is absolutely convergent, and it is possible to perform a term by term integration, so that the problem is to show that the contribution coming from the path (r, ∞) goes to zero as $r \to \infty$ no matter how I is evaluated. This is certainly true for the form (1.34), due to the convergence of the integral. In the form (1.35) we can write

$$I_r = \sum_0 p_n \int_r^\infty u_{\lambda+\beta}(\rho)u_{n+\beta}(\rho)\rho^{-2}\,d\rho$$

$$= \sum_0^\infty p_n \times \frac{\sin(\lambda-n)\frac{1}{2}\pi - [u_{(n+\beta)}(r)u'_{(\lambda+\beta)}(r) - u_{(\lambda+\beta)}(r)u'_{(n+\beta)}(r)]}{(\lambda+\beta)(\lambda+\beta+1) - (n+\beta)(n+\beta+1)}. \qquad (1.36)$$

The series converge and, for large r, tend to the value obtained by replacing all terms by their asymptotic behavior, according to the discussion of Sec. 1.2.2.

(2) We now study the term by term integrability of a product of two series:

$$\int_0^\infty h_\alpha(x_1, x)h_\beta(x, x_2)x^{-2}\,dx. \qquad (1.37)$$

If the two series are of class 2, the problem is trivial. Furthermore, it is easy in that case to obtain a bound for the integral by choosing $\theta = \tan^{-1} x$ as a new variable, and

$$\bar{u}_\nu(\theta) = u_\nu(x)x^{-1}[dx/d\theta]^{\frac{1}{2}} \qquad (1.38)$$

as new functions inside the series. The integral is thus reduced to a finite integral of two bounded functions, or of two absolutely convergent series. For more general series, it is not easy to get bounds valid for any value of x_1 and x_2, but it is easy to show that, for any couple of *finite* values of x_1 and x_2, the term-by-term integration is possible and leads to the same result as the integral (1.3), provided this integral be convergent and the coefficients in the series be bounded by a power of their index. The proof is the same as above. The contribution of the path (r, ∞) leads to the following series:

$$\sum a_n^{(\alpha)}u_n(x_1) \sum a_\nu^{(\beta)}u_\nu(x_2) \times \frac{\sin(\nu-n)\frac{1}{2}\pi - [u_n(r)u'_\nu(r) - u_\nu(r)u'_n(r)]}{\nu(\nu+1) - n(n+1)}. \qquad (1.39)$$

These series can be bounded with the help of the formula (A5), and for $r \gg x_1$ and x_2, say $r > \max(x_1^2, x_2^2)$, the result tends to zero.

1.2.5. Bounds for Integrals

In order to denote integrals of the form (1.37), we use the condensed notation

$$h_\alpha\begin{bmatrix}x_1\\x\end{bmatrix}\cdot h_\beta\begin{bmatrix}x\\x_2\end{bmatrix}. \qquad (1.40)$$

A term-by-term integration and a majorization of the series with the help of (1.35) show readily the following properties:

$$h_i\begin{bmatrix}x_1\\x\end{bmatrix}\cdot h_2\begin{bmatrix}x\\x_2\end{bmatrix} \in B, \qquad \forall i = 0, \bar{0}, 1, 2, \qquad (1.41)$$

$$h_1\begin{bmatrix}x_1\\x\end{bmatrix}\cdot h_1\begin{bmatrix}x\\x_2\end{bmatrix} \in B, \qquad (1.42)$$

$$h_0\begin{pmatrix}x_1\\x\end{pmatrix}\cdot h_1\begin{pmatrix}x\\x_2\end{pmatrix} \in B, \qquad h_{\bar{0}}\begin{pmatrix}x_1\\x\end{pmatrix}\cdot h_1\begin{pmatrix}x\\x_2\end{pmatrix} \in B, \qquad (1.43)$$

$$h_{\bar{0}}\begin{bmatrix}x_1\\x\end{bmatrix}\cdot h_{\bar{0}}\begin{bmatrix}x\\x_2\end{bmatrix} = h_1\begin{bmatrix}x_1\\x_2\end{bmatrix}; \quad h_0\begin{bmatrix}x_1\\x\end{bmatrix}\cdot h_0\begin{bmatrix}x\\x_2\end{bmatrix} = h_1\begin{bmatrix}x_1\\x_2\end{bmatrix}. \qquad (1.44)$$

Schwarz's inequality[12] enables us to get an upper bound for the dis-symetric double integral:

$$\left|h_0\begin{bmatrix}x_1\\x\end{bmatrix}\cdot h_{\bar{0}}\begin{bmatrix}x\\x_2\end{bmatrix}\right| < \left\{\left[h_0\begin{bmatrix}x\\x\end{bmatrix}\cdot h_0\begin{bmatrix}x\\x_1\end{bmatrix}\right]\left[h_{\bar{0}}\begin{bmatrix}x_2\\x\end{bmatrix}\cdot h_{\bar{0}}\begin{bmatrix}x\\x_2\end{bmatrix}\right]\right\}^{\frac{1}{2}}$$

$$< C(\log x_1)^{\frac{1}{2}}(\log x_2)^{\frac{1}{2}}. \qquad (1.45)$$

It is easy to see from the above tables that any triple integral is bounded independently of x_1 and x_2.

1.2.6. Bounds for Derived Series and Integrals

With the help of the well-known formula[7]

$$2J'_\nu(Z) = J_{\nu-1}(Z) - J_{\nu+1}(Z), \qquad (1.46)$$

it is easy to obtain upper bounds for the derivative of series of the class 2—and to show, in particular, that

$$r(d/dr)r^{-1}h_2(r, r) \sim C\sin 2r, \qquad r \to \infty. \qquad (1.47)$$

[12] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities* (Cambridge University Press, New York, 1959).

The following other formulas have been derived from the results of Appendix B (Sec. A5):

$$\frac{\partial}{\partial r} T_0(r, r') \sim -\frac{\partial}{\partial r'} T_{\bar{0}}(r, r')$$

$$\sim \tfrac{1}{2}(rr')^{\frac{1}{2}}[-J_1(|r - r'|) \operatorname{sgn}(r - r') + O(rr')^{-\frac{1}{4}} + O(r^{-1}) + O(r'^{-1})], \quad (1.48a)$$

$$\frac{d}{dr} r^{-1} T_0(r, r) = J_0(2r), \quad (1.48b)$$

$$\frac{d}{dr} r^{-1} T_{\bar{0}}(r, r) = J_2(2r),$$

$$\frac{d}{dr} r^{-1} T_1(r, r) < Cr^{-2} \log r. \quad (1.49)$$

It is easy, with the help of these formulas, to get bounds for the derivatives of integrals, in particular:

$$r \frac{d}{dr} r^{-1}\left[h_i\begin{bmatrix} r \\ x \end{bmatrix} \cdot h_2\begin{bmatrix} x \\ r \end{bmatrix}\right] \in B, \quad (1.50)$$

$$r \frac{d}{dr} r^{-1}\left[h_i\begin{bmatrix} r \\ x \end{bmatrix} \cdot h_1\begin{bmatrix} x \\ r \end{bmatrix}\right] \in B. \quad (1.51)$$

For the dis-symetric integral, we do an exact calculation for the typical example:

$$T_{0,\bar{0}}(r, r') = \int_0^\infty T_0(r, r_1) T_{\bar{0}}(r_1, r') r_1^{-2}\, dr_1. \quad (1.52)$$

Using exact expressions of T_0 and $T_{\bar{0}}$ (Appendix B), and a well-known Weber–Schafheitlein integral,[13] we get the following result:

$$T_{\bar{0},0}(r, r') = rr' \int_0^{\frac{1}{2}\pi} J_1(r \sin\alpha)\, d\alpha \times \int_0^\alpha J_0(r' \sin\beta) \sin\beta\, d\beta \quad (1.53)$$

and

$$r \frac{d}{dr} r^{-1} T_{\bar{0}0}(r, r) = r \frac{d}{dr} r^{-1} T_{0,\bar{0}}(r, r) \quad (1.54)$$

$$= \frac{r^2}{2}\left[\int_0^{\frac{1}{2}\pi} J_0(r \sin\alpha) \sin\alpha\, d\alpha\right]^2 = \tfrac{1}{2} \sin^2 r. \quad (1.55)$$

2. INVERSION OF M

We want to construct a two-sided inverse of the matrix M, the elements of which are given by

$$M_l^{l'} = [(l' + \tfrac{1}{2})^2 - (l + \tfrac{1}{2})^2]^{-1}, \text{ for } l' - l \text{ odd},$$
$$= 0, \text{ for } l' - l \text{ even}. \quad (2.1)$$

Since no additional labor is involved by doing so, we construct an inverse of the more general matrix:

$$M_l^{(\alpha) l'} = [(l' + \alpha)^2 - (l + \alpha)^2]^{-1}, \text{ for } l' - l \text{ odd}.$$
$$= 0, \text{ for } l' - l \text{ even}, \quad \alpha > -\tfrac{1}{2}. \quad (2.2)$$

With the help of the following well-known[14] formula

$$\int_0^\infty J_\mu(\rho) J_\nu(\rho) \rho^{-1}\, d\rho = \frac{\sin(\frac{1}{2}\pi(\nu - \mu))}{\frac{1}{2}\pi(\nu^2 - \mu^2)}, \quad (2.3)$$

we write $M_l^{(\alpha) l'}$ in the following form which is equivalent to (2.2):

$$M_l^{(\alpha) l'} = \tfrac{1}{2}\pi \sin\tfrac{1}{2}\pi(l' - l) \times \int_0^\infty J_{l+\alpha}(\rho) J_{l'+\alpha}(\rho) \rho^{-1}\, d\rho \quad (\alpha > -\tfrac{1}{2}).$$

We first give a formal procedure. All necessary convergence proofs are given below.

Formal procedure: Let $\gamma_{l'}^q$ be an element of the right inverse Γ of $M^{(\alpha)}$. We assume *a priori* that $\gamma_{l'}^q = 0$ for $|l' - q|$ even. We denote by $S_0^{(p)}$ and $S_1^{(p)}$ the following series:

$$S_1^{(p)} = \sum_{n=0}^\infty J_{2n+1+\alpha}(\rho)(-1)^n \gamma_{2n+1}^{2p},$$
$$S_0^{(p)} = \sum_{n=0}^\infty J_{2n+\alpha}(\rho)(-1)^n \gamma_{2n}^{2p+1}. \quad (2.4)$$

The fundamental equation

$$M^{(\alpha)} \Gamma = I \quad (2.5)$$

is equivalent to the following set of equations which should hold for any integer value of r and p:

$$(-1)^r \int_0^\infty \frac{d\rho}{\rho} J_{2r+\alpha}(\rho) S_1^{(p)}(\rho) = \frac{2}{\pi} \delta_r^p, \quad (2.6)$$

$$(-1)^{r+1} \int_0^\infty \frac{d\rho}{\rho} J_{2r+\alpha+1}(\rho) S_0^{(p)}(\rho) = \frac{2}{\pi} \delta_r^p. \quad (2.7)$$

The problem is therefore to obtain two sets of functions $S_1^{(p)}(\rho)$ and $S_0^{(p)}(\rho)$ which are respectively orthogonal to any function $J_{2r+\alpha}(\rho)$ and to any function $J_{2r+1+\alpha}(\rho)$, except for $r = p$.

In order to do this, let us introduce two auxiliary functions

$$S_1(\rho) \equiv (\tfrac{1}{2}\rho)^{\frac{1}{2}} J_{\frac{1}{2}+\alpha}(\rho) = \sum_{n=0}^\infty \alpha_n^1 J_{2n+1+\alpha}(\rho), \quad (2.8)$$

$$S_0(\rho) \equiv (\tfrac{1}{2}\rho)^{\frac{1}{2}} J_{-\frac{1}{2}+\alpha}(\rho) = \sum_{n=0}^\infty \alpha_n^0 J_{2n+\alpha}(\rho), \quad (2.9)$$

[13] Reference 11, 7.7(29).

[14] Reference 11, 7.14(32).

where, according to well-known formulas,[15]

$$\alpha_n^1 = \frac{\Gamma(n+1+\alpha)\Gamma(\frac{1}{2})(2n+1+\alpha)}{n!\,\Gamma(\frac{1}{2}-n)\Gamma(n+\alpha+\frac{3}{2})}, \tag{2.10}$$

$$\alpha_n^0 = \frac{\Gamma(n+\alpha)\Gamma(\frac{1}{2})(2n+\alpha)}{n!\,\Gamma(\frac{1}{2}-n)\Gamma(n+\alpha+\frac{1}{2})}. \tag{2.11}$$

The scalar product of these functions with Bessel functions is equal to[16]

$$\int_0^\infty J_{\lambda+\alpha}(\rho)S_1(\rho)\frac{d\rho}{\rho} = \frac{\Gamma(\frac{1}{2})\Gamma(\alpha+\frac{1}{2}+\frac{1}{2}\lambda)}{2\Gamma(1-\frac{1}{2}\lambda)\Gamma(1+\alpha+\frac{1}{2}\lambda)\Gamma(\frac{1}{2}+\frac{1}{2}\lambda)}, \tag{2.12}$$

$$\int_0^\infty J_{\lambda+\alpha}(\rho)S_0(\rho)\frac{d\rho}{\rho} = \frac{\Gamma(\frac{1}{2})\Gamma(\alpha+\frac{1}{2}\lambda)}{2\Gamma(\frac{1}{2}-\frac{1}{2}\lambda)\Gamma(\alpha+\frac{1}{2}\lambda+\frac{1}{2})\Gamma(1+\frac{1}{2}\lambda)}. \tag{2.13}$$

The formula (2.12) shows that $S_1(\rho)$ is orthogonal to any function $J_{2r+\alpha}$ except for $r = 0$, and, according to (2.13), $S_0(\rho)$ is orthogonal to any function $J_{2r+1+\alpha}$. With S_0 and S_1, we build the following functions:

$$\tilde{S}_0^{(p)} = \sum_{n=0}^\infty \alpha_n^0 \frac{1}{(2n+\alpha)^2-(2p+1+\alpha)^2} \times J_{2n+\alpha}(\rho). \tag{2.14}$$

$$\tilde{S}_1^{(p)} = \sum_{n=0}^\infty \alpha_n^1 \frac{1}{(2n+1+\alpha)^2-(2p+\alpha)^2} \times J_{2n+1+\alpha}(\rho). \tag{2.15}$$

Let us compute the scalar product of such a function $\tilde{S}$ with a function $J_{\lambda+\alpha}(\rho)$:

$$A(\tilde{S}) = \int_0^\infty J_{\lambda+\alpha}(\rho)\frac{d\rho}{\rho}\tilde{S}(\rho)\,d\rho. \tag{2.16}$$

Using the Bessel differential equation and making two integrations by parts, we obtain the following result:

$$(\lambda+\alpha)^2 A(\tilde{S}) = \left[\tilde{S}(\rho)\rho\frac{d}{d\rho}J_{\lambda+\alpha}(\rho) - J_{\lambda+\alpha}(\rho)\rho\frac{d}{d\rho}\tilde{S}(\rho)\right]^\infty + \int_0^\infty J_{\lambda+\alpha}(\rho)\left(\frac{d}{d\rho}\rho\frac{d}{d\rho}\tilde{S}(\rho)\right)d\rho. \tag{2.17}$$

The results of Sec. 1.2.2 show that the terms in the square brackets in (2.17) are bounded when $\rho \to \infty$, and that we can use the asymptotic behavior of the functions[17] involved to compute them: we find for the bracket $B_0^{(p)}$ and $B_1^{(p)}$ associated respectively with $\tilde{S}_0^{(p)}$ and $\tilde{S}_1^{(p)}$

$$B_0^{(p)}(\rho) = \sin\lambda\frac{\pi}{2}\sum_{n=0}^\infty \alpha_n^0 \times \frac{(-1)^n}{(2n+\alpha)^2-(2p+1+\alpha)^2} + O\left(\frac{1}{\rho^{\frac{1}{2}}}\right), \tag{2.18}$$

$$B_1^{(p)}(\rho) = \sin(\lambda-1)\frac{\pi}{2}\sum_{n=0}^\infty \alpha_n^1 \times \frac{(-1)^n}{(2n+1+\alpha)^2-(2p+\alpha)^2} + O\left(\frac{1}{\rho^{\frac{1}{2}}}\right). \tag{2.19}$$

Using (2.8), (2.9), and well-known formulas, we can write (2.18) and (2.19) in the following equivalent forms[11]:

$$B_0^{(p)}(\infty) = C\sin\left(\lambda\frac{\pi}{2}\right) \times \int_0^\infty J_{2p+1+\alpha}(\rho)S_0(\rho)\frac{d\rho}{\rho}, \tag{2.20}$$

$$B_1^{(p)}(\infty) = C\sin\left[(\lambda-1)\frac{\pi}{2}\right] \times \int_0^\infty J_{2p+\alpha}(\rho)S_1(\rho)\frac{d\rho}{\rho}. \tag{2.21}$$

It results from (2.12) and (2.13) that $B_0^{(p)}(\infty)$ is equal to zero for any p, whereas $B_1^{(p)}(\infty)$ is equal to zero for any p except $p = 0$, so that the following relations are derived, using (2.17) and Bessel differential equations:

$$[(\lambda+\alpha)^2-(2p+1+\alpha)^2]A(\tilde{S}_0^{(p)}) = \int_0^\infty J_{\lambda+\alpha}(\rho)S_0(\rho)\frac{d\rho}{\rho}, \tag{2.22}$$

$$[(\lambda+\alpha)^2-(2p+\alpha)^2]A(\tilde{S}_1^{(p)}) = \int_0^\infty J_{\lambda+\alpha}(\rho)S_1(\rho)\frac{d\rho}{\rho}, \tag{2.23}$$

with the help of (2.13) and (2.12), and a glance at (2.16) ($p \neq 0$), it is easy to deduce from (2.22) and (2.23) the two formulas:

$$\int_0^\infty J_{2r+1+\alpha}(\rho)\frac{d\rho}{\rho}[\tilde{S}_0^{(p)}(\rho) - C(p)S_0(\rho)] = \frac{(-1)^{p+1}}{8} \times \frac{\Gamma(\frac{1}{2})\Gamma(p+1)\Gamma(p+\alpha+\frac{1}{2})\,\delta_r^p}{(2p+1+\alpha)\Gamma(p+\frac{3}{2})\Gamma(p+\alpha+1)}, \tag{2.24}$$

where the arbitrary constant $C(p)$ may depend on p;

[15] Reference 11, 7.15(2).
[16] Reference 11, 7.7(30).
[17] Reference 11, 7.13(3).

$$\int_0^\infty J_{2r+\alpha}(\rho)\frac{d\rho}{\rho}\left[\tilde{S}_1^{(p)}(\rho)-\frac{S_1(\rho)}{\alpha^2-(2p+\alpha)^2}\right]$$
$$=\frac{(-1)^p}{8}\frac{\Gamma(\frac{1}{2})}{2p+\alpha}\frac{\Gamma(p)\Gamma(p+\alpha+\frac{1}{2})}{\Gamma(p+\frac{1}{2})\Gamma(p+\alpha+1)}\delta_r^p \qquad (p\neq 0). \qquad (2.25)$$

Comparing now (2.24) with (2.7), we get, with the help of (2.4), the formula

$$\gamma_{2n}^{2p+1}=\frac{16(2p+1+\alpha)}{\pi^2}\frac{\Gamma(p+1+\alpha)\Gamma(p+\frac{3}{2})}{\Gamma(p+\frac{1}{2}+\alpha)\Gamma(p+1)}$$
$$\times\left[\frac{1}{(2n+\alpha)^2-(2p+1+\alpha)^2}-C(p)\right]$$
$$\times\frac{\Gamma(n+\alpha)\Gamma(n+\frac{1}{2})}{\Gamma(n+\alpha+\frac{1}{2})\Gamma(n+1)}(2n+\alpha). \qquad (2.26)$$

In the same way, from (2.6), (2.24), and (2.3), we obtain the value of γ_{2n+1}^{2p} for $p\neq 0$:

$$\gamma_{2n+1}^{2p}=\frac{16(2p+\alpha)}{\pi^2}\frac{\Gamma(p+\alpha)\Gamma(p+\frac{1}{2})}{\Gamma(p+\alpha+\frac{1}{2})\Gamma(p+1)}$$
$$\times\left[\frac{1}{(2n+1+\alpha)^2-(2p+\alpha)^2}\right]$$
$$\times\frac{\Gamma(n+1+\alpha)\Gamma(n+\frac{3}{2})}{\Gamma(n+\frac{1}{2}+\alpha)\Gamma(n+1)}(2n+1+\alpha). \qquad (2.27)$$

Now, since $[M^{(\alpha)}]$ is antisymmetric, and if Γ is also antisymmetric, it is simultaneously a right inverse and a left inverse of $[M^{(\alpha)}]$. Comparison of (2.26) and (2.27) shows that this condition is fulfilled if $C(p)$ is equal to zero and yields the last unknown coefficients, γ_{2n+1}^0.

We write the final result in the following form, valid for any value of n and p, including 0.

$$\gamma_{2n}^{2p+1}(\alpha)=\mu_{2n}(\alpha)$$
$$\times\frac{1}{(2n+\alpha)^2-(2p+1+\alpha)^2}\mu_{2p+1}(\alpha), \qquad (2.28)$$

$$\gamma_{2n+1}^{2p}(\alpha)=\mu_{2n+1}(\alpha)$$
$$\times\frac{1}{(2n+1+\alpha)^2-(2p+\alpha)^2}\mu_{2p}(\alpha), \qquad (2.29)$$

where

$$\mu_{2n}(\alpha)=\frac{4}{\pi}\frac{\Gamma(n+\alpha)\Gamma(n+\frac{1}{2})}{\Gamma(n+\alpha+\frac{1}{2})\Gamma(n+1)}(2n+\alpha), \qquad (2.30)$$

$$\mu_{2n+1}(\alpha)$$
$$=\frac{4}{\pi}\frac{\Gamma(n+1+\alpha)\Gamma(n+\frac{3}{2})}{\Gamma(n+\alpha+\frac{1}{2})\Gamma(n+1)}(2n+1+\alpha). \qquad (2.31)$$

For $\alpha=0$, it is easy to find again Newton's results (with Redmond's correction of sign). The results for $\alpha=\frac{1}{2}$ are written in Sec. 3.

Proofs of convergence. Two kinds of convergence problems have been encountered throughout the formal derivation. The first one concerns the convergence of definite integrals involving Bessel functions. It is easy to verify that the condition $\alpha>-\frac{1}{2}$ ensures this convergence in all cases. The second one concerns the possibility of interchanging the order of summation and integration on the series $S_0(\rho)$, $S_1(\rho)$, $S_0^{(p)}(\rho)$, $S_1^{(p)}(\rho)$, in order to integrate these series term by term. It is easy to deduce from (2.10), (2.11), (2.28), and (2.29), with the help of Stirling formula,[18] the following equivalence, valid for large values of n, with p being a fixed quantity:

$$\gamma_{2n+1}^{2p}\sim C\alpha_n^1+O(n^{-2})\sim C+O(n^{-2}), \quad \text{for } n\to\infty,$$
$$\gamma_{2n}^{2p+1}\sim C\alpha_n^0+O(n^{-2})\sim C+O(n^{-2}) \quad \text{for } n\to\infty.$$

Referring to the analysis of Sec. 1.2.4, we see that the problem is solved if the integrals $S_0(\rho)$ and $S_1(\rho)$ are convergent. This is true for $\alpha>-\frac{1}{2}$.

Vector annihilated by M. From (2.9), (2.11), and (2.13), it is obvious that the vector $\hat{v}$ which has the following components:

$$v_{2n}=2(-1)^n\alpha_n^0/\Gamma(\tfrac{1}{2}),$$
$$v_{2n+1}=0 \qquad (2.32)$$

is annihilated by $M^{(\alpha)}$. In the case of $M(=M^{(\frac{1}{2})})$, the components are

$$v_{2n}=\frac{2}{\pi}\frac{\Gamma(n+\frac{1}{2})\Gamma(n+\frac{1}{2})}{\Gamma(n+1)\Gamma(n+1)}(2n+\tfrac{1}{2}). \qquad (2.33)$$

It results from Redmond's analysis and Newton's construction of $\hat{v}_{(\alpha=\frac{1}{2})}$ from $\hat{v}_{(\alpha=0)}$ that $\hat{v}_{(\alpha=\frac{1}{2})}$ is unique. We have not tried to show the uniqueness here, and to extend it for any other α: actually, we think that Newton's construction may also be extended, so that the uniqueness holds for any α.

3. CONSTRUCTION AND ASYMPTOTIC BEHAVIOR OF THE POTENTIALS

In this section we are interested in the behavior of the potentials constructed from a given set of phase shifts. The special case of zero phase shifts and the corresponding class of potentials (which we call "transparent" potentials) are first considered. This illustrates the importance of the asymptotic behavior of the coefficients a_l. This behavior can be obtained, modulo some very weak conditions

[18] Reference 11, 1.18(4).

satisfied by the phase shifts. The asymptotic behavior of the potentials is then derived in the general case.

3.1. Transparent Potentials

The solution of (1.20) can be written, in matrix notations, as follows:

$$\hat{a} = [1 + [R]]^{-1}[\alpha\hat{v} + [M^{-1}][\tan\Delta]\hat{e}], \tag{3.1}$$

where $\hat{a}$ and $\hat{e}$ are the column vectors a_l and 1, respectively.

$$[\tan\Delta]_l^{l'} \equiv \delta_l^{l'} \tan\delta_l, \tag{3.2}$$

$$[R] = [M^{-1}][\tan\Delta][M][\tan\Delta], \tag{3.3}$$

$\hat{v}$ being defined by (2.32), and α an arbitrary parameter.

If all the phase shifts are equal to zero, (3.1) reduces to

$$\hat{a}_0 = \alpha\hat{v}. \tag{3.4}$$

From (1.21), (1.19), and (1.17), we obtain easily the values of $c_l^{(0)}$, which define, through (1.1), a function $f_0(r, r')$:

$$\begin{aligned} c_{2n+1}^{(0)} &= 0, \\ c_{2n}^{(0)} &= \alpha v_{2n}\left[1 - \frac{\pi}{2}\alpha\frac{v_{2n}}{(4n+1)}\right]^{-1}. \end{aligned} \tag{3.5}$$

Let the function $g_0(r, r')$ be the solution of the following integral equation:

$$g_0(r,r') = f_0(r,r') - \int_0^\infty f_0(r,r_1)g_0(r_1,r')r_1^{-2}\,dr_1. \tag{3.6}$$

Using (3.5), (1.2), and (2.3), it is easy to derive the following expansion for $g_0(r, r')$:

$$g_0(r,r') = \alpha\sum_0^\infty v_{2n}u_{2n}(r)u_{2n}(r'). \tag{3.7}$$

We assert now that $(d/dr)r^{-1}K_0(r, r)$ and $(d/dr)r^{-1}g_0(r, r)$ have the same asymptotic behavior when r tends to infinity.

Proof: We deal with Fredholm equations in which the domain of integration extends to $+\infty$. However, this is not a difficulty as long as all the integrals involved are convergent. The infinite limits may be easily removed by a suitable change of variable and a trivial change of functions, as we see below (Sec. 3.3). The new kernel belongs to L_2, so that the method of Neuman series applies without any difficulty. We first write down a formal derivation.

We recall now the Fredholm equation (1.5) for $K_0(r, r')$ and, replacing in its right-hand side $f_0(r_1, r')$ by its expression deduced from (3.6), we get

$$\begin{aligned} \int_0^r r_1^{-2}\,dr_1\,K_0(r,r_1)f_0(r_1,r') &= \int_0^r r_1^{-2}\,dr_1\,K_0(r,r_1)g_0(r_1,r') \\ &+ \int_0^r K_0(r,r_1)r_1^{-2}\,dr_1 \\ &\times \int_0^\infty f_0(r_1,r_2)g_0(r_2,r')r_2^{-2}\,dr_2. \end{aligned} \tag{3.8}$$

If we interchange the order of integrations and use again (1.5), we see that the last term in (3.8) is equal to

$$\begin{aligned} &-\int_0^\infty r_2^{-2}\,dr_2\,g_0(r_2,r')K_0(r,r_2) \\ &\quad + \int_0^\infty r_2^{-2}\,dr_2\,g_0(r_2,r')f_0(r,r_2). \end{aligned} \tag{3.9}$$

The last term in (3.9) is replaced by its expression (3.5). Insertion of the result in (3.8) and subsequently of (3.8) in (1.5) yields

$$\begin{aligned} K_0(r,r') &= g_0(r,r') \\ &+ \int_r^\infty r_1^{-2}\,dr_1\,K_0(r,r_1)g_0(r_1,r'). \end{aligned} \tag{3.10}$$

For a given r, (3.10) is a Fredholm equation. Let us introduce also the Fredholm equation[19] for the resolvent:

$$\begin{aligned} S_r(\rho,r') &= g_0(\rho,r') \\ &+ \int_r^\infty r_1^{-2}\,dr_1\,S_r(\rho,r_1)g_0(r_1,r'). \end{aligned} \tag{3.11}$$

It is clear that $S_r(\rho, r')$ is a symmetric function of ρ and r' and that

$$K_0(r,r') = S_r(r,r') = S_r(r',r). \tag{3.12}$$

In order to obtain $V(r)$, we need to know $K_0(r, r')$ and its two partial derivatives at $r = r'$. They can be obtained as follows: From (3.10), we get easily the following equations,

$$\begin{aligned} \frac{\partial}{\partial r'}K_0(r,r') &= \frac{\partial}{\partial r'}g_0(r,r') \\ &+ \int_r^\infty r_1^{-2}\,dr_1\,K_0(r,r_1)\frac{\partial}{\partial r'}g_0(r_1,r'), \end{aligned} \tag{3.13}$$

$$\begin{aligned} \frac{\partial}{\partial r}K_0(r,r') &= \frac{\partial}{\partial r}g_0(r,r') + r^{-2}K_0(r,r)g_0(r,r') \\ &+ \int_r^\infty r_1^{-2}\,dr_1\left[\frac{\partial}{\partial r}K_0(r,r_1)\right]g_0(r_1,r'), \end{aligned} \tag{3.14}$$

[19] F. G. Tricomi, *Integral Equations* (Interscience Publishers, Inc., New York, 1957).

or, in obvious notations,

$$\frac{\partial}{\partial r}[K_0(r, r')] = s_r(r') + \int_r^\infty r_1^{-2}\, dr_1 \frac{\partial}{\partial r} K_0(r, r_1) g_0(r_1, r'). \quad (3.15)$$

The well-known properties of Fredholm equations[19] enable us to write down the solution of (3.15), using (3.11),

$$\frac{\partial}{\partial r} K_0(r, r') = s_r(r') + \int_r^\infty r_1^{-2}\, dr_1\, s_r(r_1) S_r(r_1, r'). \quad (3.16)$$

Keeping in mind the symmetric properties of the function g_0, we derive easily from (3.13), (3.16), (3.12), and (3.10) the following result:

$$\frac{d}{dr} K_0(r, r) = \frac{d}{dr} g_0(r, r) + r^{-2}[K_0(r, r)]^2 + 2\int_r^\infty r_1^{-2}\, dr_1\, K_0(r, r_1) \frac{\partial}{\partial r} g_0(r_1, r). \quad (3.17)$$

The problem reduces therefore to the determination of $K_0(r, r')$. From the formula (1.24) it follows that, for large values of r and r', we can take for $g_0(r, r')$ the following bounds:

$$|g_0(r, r')| < C(rr')^{\frac{1}{4}}(1 + |r - r'|)^{-\frac{1}{2}}. \quad (3.18)$$

Using this bound, it is easy to derive the following bounds for the norm of the kernel of (3.10), and related quantities:

$$\left[\int_r^\infty g_0^2(x, y) x^{-2}\, dx\right]^{\frac{1}{2}} < C[\log y]^{\frac{1}{2}}. \quad (3.19)$$

$$N^2 = \int_r^\infty y^{-2}\, dy \int_r^\infty x^{-2} g_0^2(x, y)\, dx < Cr^{-1} \log r. \quad (3.20)$$

The Neuman series for (3.10) converges provided that r is large enough, and the classical[19] study of those series yields the following behavior for $K_0(r, r')$:

$$|K_0(r, r') - g_0(r, r')| < C(\log r)^{\frac{1}{2}}(\log r')^{\frac{1}{2}}. \quad (3.21)$$

On the other hand, we know from formula (1.48a) that, for large enough values of r and r', the following bound holds:

$$\left|\frac{\partial}{\partial r} g_0(r, r')\right| < C(rr')^{\frac{1}{4}}(1 + |r - r'|)^{-\frac{1}{2}}. \quad (3.22)$$

According to well-known theorems[20] we infer from the bounds (3.18), (3.21), (3.22), and the obvious continuity of the functions, the convergence of all the integrals written and the legitimacy of differentiating under the integral sign. Furthermore, it follows from (1.25), (1.26), and a calculation analogous to (3.20) that $f_0(r, r')$ is a Hilbert–Schmidt kernel, so that $g_0(r, r')$ is the unique solution of (3.6) in $\mathfrak{L}_2$. For r sufficiently large, the Neuman series defines unambiguously $K_0(r, r')$ from $g_0(r, r')$ or conversely, so that $K_0(r, r')$ exists and is unique for large values of r.

Now, we can derive from (1.48a) the following formula:

$$(\partial/\partial r) g_0(r, r') + (\partial/\partial r') g_0(r, r') = O(r^{\frac{1}{4}} r'^{\frac{1}{4}}). \quad (3.23)$$

If we substitute (3.23) and (3.21) in the integral in (3.17), we find, after one integration by parts and an evaluation of the remainder:

$$(d/dr) K_0(r, r) = (d/dr) g_0(r, r) + O(r^{-\frac{1}{2}} \log r). \quad (3.24)$$

From (3.21) and (3.24) the asymptotic behavior of $V(r)$ follows readily:

$$V(r) = -2r^{-1}(d/dr) r^{-1} g_0(r, r) + O(r^{-5/2} \log r) \quad (3.25)$$

or, according to (1.48b),

$$V(r) \sim -2\alpha\pi^{-\frac{1}{2}} r^{-\frac{1}{2}} \cos(2r - \tfrac{1}{4}\pi). \quad (3.26)$$

The analytic continuation of $V(r)$ will be studied in a following paper.[7]

3.2. Asymptotic Behavior of the Coefficients in the General Case

Fundamental Assumption. We assume in the following that the phase shifts go to zero when $l \to \infty$ faster than $l^{-3-\epsilon}$. Such an assumption is not very restrictive. It has been proved by several authors that this assumption certainly holds for potentials going to zero faster than r^{-4} when $r \to \infty$. Furthermore, in special cases as, say, superposition of Yukawa potentials, or potentials bounded by a function $e^{-\mu r}$, it has been proved that the phase shifts go exponentially to zero when $l \to \infty$. With this assumption, it is possible to obtain the asymptotic behavior of the coefficients a_l.

In order to obtain the coefficients a_l from the phase shifts, we have to solve (3.1), in which the elements $\gamma_l^{l'}$ of $[M^{-1}]$ are obtained by putting $\alpha = \frac{1}{2}$ in (2.28)–(2.31). We write them in the following form:

$$\gamma_{2n}^{2p+1} = \bar{\mu}_{2n}\bar{\mu}_{2p+1}(2p + 1)(2p + 2) \times [2n(2n + 1) - (2p + 1)(2p + 2)]^{-1},$$
$$\gamma_{2n+1}^{2p} = \bar{\mu}_{2n+1}\bar{\mu}_{2p} + \rho_{2n+1}^{2p}, \quad (3.27)$$

[20] L. Schwartz, *Méthodes mathématiques pour les sciences physiques* (Hermann & Cie, Paris, 1961), pp. 62, 33.

where

$$\rho_{2n+1}^{2p} = \bar{\mu}_{2n+1}\bar{\mu}_{2p}2p(2p+1) \times [(2n+1)(2n+2) - 2p(2p+1)]^{-1},$$

$$\bar{\mu}_{2n} = 2\pi^{-1}(2n+\tfrac{1}{2})\Gamma(n+\tfrac{1}{2})\Gamma(n+\tfrac{1}{2})[n!\,n!]^{-1}, \quad (3.28)$$

$$\bar{\mu}_{2n+1} = 2\pi^{-1}(2n+\tfrac{3}{2}) \times \Gamma(n+\tfrac{1}{2})\Gamma(n+\tfrac{3}{2})[n!\,(n+1)!]^{-1}.$$

Observe that $\bar{\mu}_q$ is defined differently according to the parity of q. It is clear that, when $n \to \infty$,

$$\bar{\mu}_{2n} \sim \bar{\mu}_{2n+1} \sim 4\pi^{-1} + O(n^{-2}). \quad (3.29)$$

We denote by $\hat{w}$ the vector defined as follows:

$$\hat{w} = \alpha\hat{v} + [M^{-1}][\tan\Delta]\hat{e}. \quad (3.30)$$

Its components are therefore given by the formulas

$$w_{2n} = \alpha v_{2n} + \sum_p \gamma_{2n}^{2p+1} \tan\delta_{2p+1}, \quad (3.31)$$

$$w_{2n+1} = \beta_0\bar{\mu}_{2n+1} + \sum_p \rho_{2n+1}^{2p} \tan\delta_{2p}, \quad (3.32)$$

where

$$\beta_0 = \sum_p \bar{\mu}_{2p} \tan\delta_{2p}. \quad (3.33)$$

It follows from (3.27)–(3.29) and the remark given in Sec. 1.2.4 that w_{2n} and w_{2n+1} have, for large values of n (if our fundamental assumption is fulfilled), the following behavior:

$$w_{2n} \sim 4\pi^{-1}\alpha + O(n^{-2}), \qquad w_{2n+1} \sim 4\pi^{-1}\beta_0 + O(n^{-2}). \quad (3.34)$$

In order to relate the asymptotic behavior of a_l to that of w_l, we now have to solve the following equations:

$$a_{2r} = w_{2r} - \sum_p R_{2r}^{2p} a_{2p}, \quad (3.35)$$

$$a_{2r+1} = w_{2r+1} - \sum_p R_{2r+1}^{2p+1} a_{2p+1}. \quad (3.36)$$

Resolution of (3.35). Owing to (3.2) and (2.1), R_{2r}^{2p} is equal to

$$R_{2r}^{2p} = -\sum_q \gamma_{2r}^{2q+1} \tan\delta_{2q+1}(-1)^q \times \int_0^{\frac{1}{2}\pi} \bar{u}_{2q+1}(\theta)(-1)^p u_{2p}(\theta)\,d\theta \tan\delta_{2p}, \quad (3.37)$$

in which the functions $\bar{u}_\nu(\theta)$ are defined as in Sec. 1.2.4.

After inserting this expression in (3.35), multiplying both sides of (3.35) by $(-1)^r \tan\delta_{2r}u_{2r}(\theta)$, and summing over r, we obtain the equation

$$\alpha_0(\theta) = \omega_0(\theta) + \int_0^{\frac{1}{2}\pi} \gamma_0(\theta, \theta')\alpha_0(\theta')\,d\theta', \quad (3.38)$$

where we put

$$\alpha_0(\theta) = \sum_r a_{2r} \tan\delta_{2r}(-1)^r\bar{u}_{2r}(\theta), \quad (3.39)$$

$$\omega_0(\theta) = \sum_r w_{2r} \tan\delta_{2r}(-1)^r\bar{u}_{2r}(\theta), \quad (3.40)$$

$$\gamma_0(\theta, \theta') = \sum_r \tan\delta_{2r}\bar{u}_{2r}(\theta)(-1)^r \times \sum_q \gamma_{2r}^{2q+1} \tan\delta_{2q+1}(-1)^q\bar{u}_{2q+1}(\theta). \quad (3.41)$$

The convergence of all these simple or double series is shown in Sec. 1.2.3. We see also that $\gamma_0(\theta, \theta')$ is bounded for any positive value of θ or θ'. As a result, the Fredholm equation (3.38) has a solution, and this solution is unique, except for exceptional sets of phase shifts for which 1 is an eigenvalue of Eq. (2.38). No efforts have been made to pin down these exceptional sets of values, or to show that they do not exist at all. We assume in the following that the given set of phase shifts does not happen to belong to one of these cases. This assumption is certainly valid, for instance, if the phase shifts are so small that the norm of the kernel $\gamma_0(\theta, \theta')$ is smaller than 1 (this case will be studied with more details in a later paper).

Let us now write the solution of (3.38) in the Fredholm form[19]:

$$\alpha_0(\theta) = \omega_0(\theta) + \mathfrak{D}^{-1} \int_0^{\frac{1}{2}\pi} \mathfrak{D}(\theta, \theta')\omega_0(\theta')\,d\theta', \quad (3.42)$$

where

$$\mathfrak{D} = 1 + \sum_{m=1}^{\infty} \frac{(-1)^m}{m!} \iint_0^{\frac{1}{2}\pi} \cdots \times \int_0^{\frac{1}{2}\pi} K\begin{pmatrix}\theta_1 & \cdots & \theta_m \\ \theta_1 & \cdots & \theta_m\end{pmatrix} d\theta_1 \cdots d\theta_m, \quad (3.43)$$

$$\mathfrak{D}(\theta, \theta') = \sum_{m=0} \frac{(-1)^m}{m!} \iint_0^{\frac{1}{2}\pi} \cdots \times \int_0^{\frac{1}{2}\pi} K\begin{pmatrix}\theta & \theta_1 & \cdots & \theta_m \\ \theta' & \theta_1 & \cdots & \theta_m\end{pmatrix} d\theta_1 \cdots d\theta_m. \quad (3.44)$$

We recall the expressions of the Fredholm determinant in function of the kernel

$$K\begin{pmatrix}x_1 & \cdots & x_m \\ y_1 & \cdots & y_m\end{pmatrix} = \begin{vmatrix} \gamma_0(x_1, y_1) & \gamma(x_1, y_2) & \cdots & \gamma(x_1, y_m) \\ \gamma_0(x_2, y_1) & & & \\ \vdots & & & \\ \gamma_0(x_m, y_1) & & \cdots & \gamma(x_m, y_m) \end{vmatrix}. \quad (3.45)$$

Any determinant in (3.44) may be expanded along the elements of its first line, which is the only line where lies the variable θ. Since any of these elements may be expanded in terms of the functions $\bar{u}_{2r}(\theta)$, we see that it is easy to expand the right-hand side of (3.42) in terms of the functions $\bar{u}_{2r}(\theta)$. Identification[21] of the coefficient of $\bar{u}_{2r}(\theta)$ in (3.42) yields the value of a_{2r}:

$$a_{2r} = w_{2r} + \mathfrak{D}^{-1}\mathfrak{D}_r(\omega_0), \tag{3.46}$$

where

$$\mathfrak{D}_r(\omega_0) = \sum_{m=0}^{\infty} \frac{(-1)^m}{m!} \int \cdots \int \begin{bmatrix} \int \gamma_0^r(\theta')\omega_0(\theta')\,d\theta', & \gamma_0^r(\theta_1) \cdots \gamma_0^r(\theta_m) \\ \int \gamma_0(\theta_1, \theta')\omega_0(\theta')\,d\theta', & \gamma_0(\theta_1, \theta_1) \cdots \gamma_0(\theta_1, \theta_m) \\ \vdots & \\ \int \gamma_0(\theta_m, \theta')\omega_0(\theta')\,d\theta', & \gamma_0(\theta_m, \theta_1) \cdots \gamma_0(\theta_m, \theta_m) \end{bmatrix} d\theta_1 \cdots d\theta_m, \tag{3.47}$$

$$\gamma_0^r(\theta') = \sum_q \gamma_{2r}^{2q+1} \tan \delta_{2q+1}(-1)^q \bar{u}_{2q+1}(\theta'). \tag{3.48}$$

All the elements of the determinant (3.47) are bounded. According to Sec. 1.2.3, if our fundamental assumption is valid, $|\gamma_0^r(\theta)|$ is bounded, for any value of θ, by $Cr^{-5/3-\epsilon}$. The same bound holds for

$$\left| \int_0^{\frac{1}{2}\pi} \gamma_0^r(\theta')\omega_0(\theta')\,d\theta' \right|,$$

since $\omega_0(\theta')$ is bounded. From Hadamard's theorem[12,19] and (3.47), it follows that

$$|\mathfrak{D}_r(\omega_0)| \leq Cr^{-5/3-\epsilon} \sum_m (m!)^{-1} C^m (m+1)^{\frac{1}{2}(m+1)}. \tag{3.49}$$

The series on the right is convergent. Provided that the phase shifts are not one of those exceptional sets for which $\mathfrak{D}$ is equal to zero, we get the asymptotic behavior of the a_{2r}:

$$a_{2r} = w_{2r} + O(r^{-5/3-\epsilon}) = 4\pi^{-1}\alpha + O(r^{-5/3-\epsilon}). \tag{3.50}$$

Method of Solution of (3.36). From (3.2) and (3.27), we derive easily the following formula:

$$R_{2r+1}^{2p+1} = \bar{R}_{2r+1}^{2p+1} + \bar{\mu}_{2r+1}\sigma_{2p+1}, \tag{3.51}$$

where

$$\bar{R}_{2r+1}^{2p+1} = \sum_q \rho_{2r+1}^{2q} \tan \delta_{2q} M_{2q}^{2p+1} \tan \delta_{2p+1} \tag{3.52}$$

and

$$\sigma_{2p+1} = \sum_q \bar{\mu}_{2q} \tan \delta_{2q} M_{2q}^{2p+1} \tan \delta_{2p+1}. \tag{3.53}$$

We also introduce the following notations:

$$\beta_1 = \sum_p a_{2p+1}\sigma_{2p+1}, \tag{3.54}$$

$$t_{2p+1} = w_{2p+1} - \beta_1 \bar{\mu}_{2p+1}. \tag{3.55}$$

Insertion of (3.51) in (3.36) yields

$$a_{2r+1} = t_{2r+1} - \sum_p \bar{R}_{2r+1}^{2p+1} a_{2p+1}. \tag{3.56}$$

[21] Observe that the functions $\bar{u}_{2r}(\theta)$ are orthogonal.

Suppose now we know how to solve this equation: let us write the solution in the general form

$$a_{2r+1} = \sum_p U_{2r+1}^{2p+1} t_{2p+1}. \tag{3.57}$$

If we replace t_{2p+1} by (3.55) in (3.57), and replace this expression of a_{2r+1} in (3.54), we easily get

$$\beta_1 = \frac{\sum_r \sigma_{2r+1} \sum_p U_{2r+1}^{2p+1} w_{2p+1}}{1 + \sum_r \sigma_{2r+1} \sum_p U_{2r+1}^{2p+1} \bar{\mu}_{2p+1}}. \tag{3.58}$$

Insertion of β_1 in (3.55) and of (3.55) in (3.57) yields in turn the value of a_{2r+1}. The problem reduces therefore to Eq. (3.56). We can use, to solve it, the same method as we did for (3.35). Let us introduce the following functions:

$$\tau_1(\theta) = \sum_{p=0}^{\infty} t_{2p+1} \tan \delta_{2p+1}(-1)^p \bar{u}_{2p+1}(\theta), \tag{3.59}$$

$$\gamma_1(\theta, \theta') = -\sum_{r=0}^{\infty} \tan \delta_{2r+1} \bar{u}_{2r+1}(\theta)(-1)^r \times \sum_q \rho_{2r+1}^{2q} \tan \delta_{2q}(-1)^q \bar{u}_{2q}(\theta'), \tag{3.60}$$

$$\gamma_1^r(\theta') = -\sum_q \rho_{2r+1}^{2q} \tan \delta_{2q}(-1)^q \bar{u}_{2q}(\theta'). \tag{3.61}$$

We find that a_{2r+1} is expressed in terms of $\tau_1(\theta)$ by the same formula as (3.46), with the following replacements:

$$w_{2r} \to t_{2r+1}, \quad \omega_0(\theta) \to \tau_1(\theta), \quad \gamma_0(\theta, \theta') \to \gamma_1(\theta, \theta'), \quad \gamma_0^{(r)} \to \gamma_1^{(r)}. \tag{3.62}$$

Quantities like

$$\sum_p U_{2r+1}^{2p+1} w_{2p+1} \quad \text{or} \quad \sum_p U_{2r+1}^{2p+1} \bar{\mu}_{2p+1}$$

may easily be obtained by replacing t_{2p+1} by w_{2p+1} or $\bar{\mu}_{2p+1}$ in $\tau_1(\theta)$. Proceeding exactly as above, we

find the asymptotic behavior of the coefficients a_{2r+1}:

$$a_{2r+1} = t_{2r+1} + O(r^{-5/3-\epsilon}) = 4\pi^{-1}\beta + O(r^{-5/3-\epsilon}), \quad (3.63)$$

where

$$\beta = \beta_0 - \beta_1 \quad (3.64)$$

and use is made of (3.55), (3.34), and (3.29).

β does not reduce in general to a trivial value since it is easy to see, for instance, that, if all the odd phase shifts are equal to zero, β_1 is zero, according to (3.53) and (3.54), whereas β_0 is not, according to (3.33).

It may be of interest to observe that the above analysis enables one to deduce the asymptotic behavior of the a_l, even if weaker conditions are fulfilled by the phase shifts: It is easy to see that if

$$\tan \delta_{2p} \sim \tan \delta_{2p+1} \sim O(p^{-4/3-\gamma-\epsilon}), \quad (3.65)$$

where

$$0 \leq \gamma \leq 2,$$

then the following formulas hold:

$$a_{2r} = \alpha v_{2r} + O(r^{-\gamma}), \quad (3.66)$$

$$a_{2r+1} = \beta \bar{u}_{2r+1} + O(r^{-\gamma}). \quad (3.67)$$

If $\tan \delta_{2p}$ and $\tan \delta_{2p+1}$ are smaller than $Cp^{-10/3}$, the remainders are of the order of n^{-2}. The Bessel series constructed with the leading term of (3.50) and (3.63) as coefficients are the series T_0 and $T_{\bar{0}}$ studied in Appendix B. For practical computations of the coefficients a_l, the simplest case will be that of small phase shifts. The norms of the operators γ_0 and γ_1 are

$$N_0^2 = \int_0^{\frac{1}{2}\pi} \int_0^{\frac{1}{2}\pi} \gamma_0^2(\theta, \theta')\, d\theta\, d\theta' = \frac{\pi^2}{4} \sum_r \frac{\tan^2 \delta_{2r}}{4r+1} \sum_q \frac{[\gamma_{2r}^{2q+1} \tan \delta_{2q+1}]^2}{4q+3}, \quad (3.68)$$

$$N_1^2 = \int_0^{\frac{1}{2}\pi} \int_0^{\frac{1}{2}\pi} \gamma_1^2(\theta, \theta')\, d\theta\, d\theta' = \frac{\pi^2}{4} \sum_r \frac{\tan^2 \delta_{2r+1}}{4r+3} \sum_q \frac{[\rho_{2r+1}^{2q} \tan \delta_{2q}]^2}{4q+1}. \quad (3.69)$$

If all the phase shifts are small enough, N_0^2 and N_1^2 are smaller than 1; for this, it is necessary, (but not sufficient), that $|\frac{1}{4}\pi \tan \delta_0 \tan \delta_1| < 1$. Then the Neuman series converge and give the solution of (3.42) and its analog with γ_1. The first-order approximation for (3.35) and (3.36) is particularly simple:

$$\hat{a} = \hat{w}. \quad (3.70)$$

We finally observe that our derivations and proofs are valid for any finite value of the $\tan \delta_l$. The consideration of a finite number of infinite values introduces only nonessential difficulties.

3.3. Asymptotic Behavior of the Potentials

From (1.17), (1.19), and (1.21), we easily derive the following formula for c_l:

$$c_l = d_l \left(1 - \frac{\pi}{2} \frac{d_l}{2l+1}\right)^{-1}, \quad (3.71)$$

where

$$d_l = a_l [\cos^2 \delta_l]^{-1} \times [1 - \sum_{l'} M_{ll'} a_{l'} (\tan \delta_{l'} - \tan \delta_l)]^{-1}. \quad (3.72)$$

The asymptotic behavior of d_l follows readily from the analysis of Sec. 1.2.3.:

$$d_l \sim a_l \sim 4\pi^{-1}[1 + O(l^{-5/3-\epsilon})]. \quad (3.73)$$

We now define, from the function $f(r, r')$, a function $g(r, r')$ in the same way as we did for $g_0(r, r')$ from $f_0(r, r')$, that is to say, in the condensed notations of Sec. 1.2.5:

$$g\begin{pmatrix} r \\ r' \end{pmatrix} = f\begin{pmatrix} r \\ r' \end{pmatrix} - f\begin{pmatrix} r \\ r_1 \end{pmatrix} \cdot g\begin{pmatrix} r_1 \\ r' \end{pmatrix}. \quad (3.74)$$

In order to study this equation, we decompose the kernel as follows:

$$f\begin{pmatrix} r \\ r' \end{pmatrix} = f_0\begin{pmatrix} r \\ r' \end{pmatrix} + f_2\begin{pmatrix} r \\ r' \end{pmatrix}, \quad (3.75)$$

where

$$f_0\begin{pmatrix} r \\ r' \end{pmatrix} = (\alpha + \beta) \sum_n v_{2n} \times [1 - \tfrac{1}{2}\pi(\alpha+\beta) v_{2n}/(4n+1)]^{-1} u_{2n}(r) u_{2n}(r'). \quad (3.76)$$

The equation corresponding to the kernel f_0 is similar to the Eq. (3.6), and its solution is g_0, except for the trivial replacement $\alpha \to (\alpha + \beta)$. With the help of this resolvent kernel, we know that a classical method[22] enables us to transform Eq. (3.74) into

$$g\begin{pmatrix} r \\ r' \end{pmatrix} + S_1\begin{pmatrix} r \\ r_1 \end{pmatrix} \cdot g\begin{pmatrix} r_1 \\ r' \end{pmatrix} = \bar{g}_0\begin{pmatrix} r \\ r' \end{pmatrix} + S_1\begin{pmatrix} r \\ r' \end{pmatrix}, \quad (3.77)$$

[22] Reference 19, p. 65.

where

$$S_1\begin{pmatrix} r \\ r' \end{pmatrix} = f_2\begin{pmatrix} r \\ r' \end{pmatrix} - \bar{g}_0\begin{pmatrix} r \\ r_1 \end{pmatrix} \cdot f_2\begin{pmatrix} r_1 \\ r' \end{pmatrix} \tag{3.78}$$

and

$$\bar{g}_0\begin{pmatrix} r \\ r_1 \end{pmatrix} = (\alpha + \beta) T_0\begin{pmatrix} r \\ r_1 \end{pmatrix}. \tag{3.79}$$

Let us now introduce the difference g_1 between g and $\bar{g}_0$. Equation (3.71) takes the form

$$g_1\begin{pmatrix} r \\ r' \end{pmatrix} + S_1\begin{pmatrix} r \\ r_1 \end{pmatrix} \cdot g_1\begin{pmatrix} r_1 \\ r' \end{pmatrix} = S_1\begin{pmatrix} r \\ r' \end{pmatrix} - S_1\begin{pmatrix} r \\ r_1 \end{pmatrix} \cdot g_0\begin{pmatrix} r_1 \\ r' \end{pmatrix}. \tag{3.80}$$

According to the formulas (1.24) and (1.25), together with (3.50) and (3.63), the function $f_2(r, r')$ can be bounded by $C(rr')^{\frac{1}{2}}$, and, according to Sec. 1.2.5, the same bound holds for $S_1(r, r')$. Let us now introduce the following new variables and functions:

$$r = (\tan \theta)^2; \qquad r' = (\tan \theta')^2; \tag{3.81a}$$

$$(rr')^{-1}(dr\, dr')^{\frac{1}{2}} \varphi(r, r') = (d\theta\, d\theta')^{\frac{1}{2}} \tilde{\varphi}(\theta, \theta'), \tag{3.81b}$$

in which φ stands for S_1, g_0, or g_1. Equation (3.80) is transformed into an integral equation whose domain of integration and kernel are bounded. Any solution of such an equation is bounded by a constant for all values of θ and θ', so that

$$g_1\begin{pmatrix} r \\ r' \end{pmatrix}$$

is bounded by $C(rr')^{\frac{1}{2}}$ for large values of r and r'. To ensure the validity of this result, we have to assume that the Fredholm determinant of (3.80) does not happen to be zero. This can happen only for particular sets of phase shifts, which we shall discard from our study. With these assumptions, the solution of (3.80) is unique, and the analysis of section (3.1) applies, which yields the asymptotic behavior of the potential:

$$V(r) = -2r^{-1}(d/dr)r^{-1} g(r, r) + O(r^{-5/2} \log r). \tag{3.82}$$

In order to derive the asymptotic behavior of g_1, let us apply two times the operator S_1 on the left of (3.80), and subtract the first result from the second one. We get, using obvious notations,

$$-S_1 \cdot g_1 + S_1 \cdot S_1 \cdot S_1 \cdot g_1 = [S_1 \cdot S_1 - S_1] \cdot [S_1 - S_1 \cdot g_0]. \tag{3.83}$$

The uniqueness of the solution of (3.38) follows from the uniqueness of the solution of (3.80). Since $S_1 \cdot S_1$, according to Sec. 1.2.5, is bounded by $C(\log r)^{\frac{1}{2}}(\log r')^{\frac{1}{2}}$, and since the right-hand side of (3.83) has this same bound, a change of variables and function defined by (3.81b) and

$$r = (\tan \theta)^{1+\epsilon}, \qquad r' = (\tan \theta')^{1+\epsilon} \tag{3.84}$$

enables us to get a Fredholm equation with finite domain and finite kernel. $S_1 \cdot g_1$ is therefore bounded by $Cr^{\epsilon}r'^{\epsilon}$. From this remark and a careful estimate of all the quantities appearing in (3.80), with the help of Secs. 1.2.4 and 1.2.5, we easily derive formula

$$g_1(r, r') \sim f_2(r, r') + O(r^{\epsilon}r'^{\epsilon}) \sim \beta[T_{\bar{0}}(r, r') - T_0(r, r')] + O(r^{\epsilon}r'^{\epsilon}). \tag{3.85}$$

Through $(\partial/\partial r')g_1(r, r')$ and $(\partial/\partial r)g_1(r, r')$, a similar analysis leads to

$$(d/dr)g_1(r, r) \sim \beta(d/dr)[T_{\bar{0}}(r, r) - T_0(r, r)] + O(r^{\epsilon}). \tag{3.86}$$

Combination of (3.85), (3.86), and (3.79), yields the value of $V(r)$:

$$\begin{aligned} V(r) &= -2r^{-1}(d/dr)r^{-1}[\alpha T_0(r, r) + \beta T_{\bar{0}}(r,r)] + O(r^{\epsilon-2}) \\ &= -2r^{-1}[(\alpha - \beta)J_1'(2r) \\ &\quad + (\alpha + \beta)(2r)^{-1}J_1(2r)] + O(r^{\epsilon-2}) \end{aligned}$$

or

$$-2\pi^{-\frac{1}{2}}(\alpha - \beta)r^{-\frac{1}{2}} \cos(2r - \tfrac{1}{4}\pi) + O(r^{\epsilon-2}), \tag{3.87}$$

where we used some results of Appendix B and the well-known recurrence relations[11] of Bessel functions. The formula (3.87) shows that only one potential exists which goes to zero faster than $r^{-\frac{1}{2}}$ as $r \to \infty$. Furthermore, this potential is the only potential, in the set of equivalent potentials, which may have a non-oscillating tail. We propose to call this potential the "special" inverse potential, and we shall use this denomination in a later paper, where we shall study the analytic properties of these potentials and of their Jost functions.

ACKNOWLEDGMENT

The author would like to thank Dr. K. Chadan for interesting discussions.

APPENDIX A

A.1. Bounds for Ricatti–Bessel Functions

The following bounds hold for these functions, defined as in Sec. 1.2:

$$p < \epsilon_0 x^{\frac{1}{2}}, \qquad u_p(x) = \sin(x - \tfrac{1}{2}p\pi) + O(\epsilon_0^2), \tag{A1}$$

$$\epsilon_0 x^{\frac{1}{3}} < p < x, \quad |u_p(x)| < C\epsilon_0^{-\frac{1}{3}} p^{\frac{1}{3}} + O(p^{-\frac{1}{3}}), \tag{A2}$$

$$x < p < x[1 + O(x^{-\frac{2}{3}})], \quad |u_p(x)| < Cp^{\frac{1}{3}} + O(p^{-5/6}), \tag{A3}$$

$$x[1 + O(x^{-\frac{2}{3}})] < p < x^2, \quad |u_p(x)| < O(p^{-5/6}), \tag{A4}$$

$$|Z|^2 < p, \quad |u_p(Z)| < C(\tfrac{1}{2}e)^p p^{-\frac{1}{3}p}, \tag{A5}$$

valid for Z real or complex.

Proof of (A1):

$$p < \epsilon_0 x^{\frac{1}{3}}.$$

Recall the asymptotic expansion[23] of a Bessel function for large argument:

$$u_p(x) = \sin(x - \tfrac{1}{2}p\pi)P_p(x) + \cos(x - \tfrac{1}{2}p\pi)Q_p(x),$$

in which $P_p(x)$ and $Q_p(x)$ can be estimated as follows[24]:

$$\left|P_p(x) - \sum_{m=0}^{M=E(p/2+1)} \frac{(-1)^m \Gamma(p+1+2m)}{(2m)!\,\Gamma(p+1-2m)} (2x)^{-2m}\right| \leq \left|\frac{\Gamma(p+3+2M)(2x)^{-2M-2}}{(2M+2)!\,\Gamma(p-1-2M)}\right|,$$

$$\left|Q_p(x) - \sum_{m=0}^{M'=E(p+1/2)} \frac{(-1)^m \Gamma(p+2+2m)}{(2m+1)!\,\Gamma(p-2m)} (2x)^{-2m-1}\right| \leq \left|\frac{\Gamma(p+4+2M)(2x)^{-2M-3}}{(2M+3)!\,\Gamma(p-2-2M)}\right|.$$

The problem of obtaining estimations and bounds is easily solved with the help of the Stirling's formula, and leads to (A1). Q.E.D.

Proof of (A2):

$$\epsilon_0 x^{\frac{1}{3}} < p < (x - \tfrac{1}{2}).$$

We start from Langer's formula[25]

$$J_p(x) = w^{-\frac{1}{3}}(w - \tan^{-1} w)^{\frac{1}{3}} \times [J_{\frac{1}{3}}(Z)\cos\tfrac{1}{6}\pi - Y_{\frac{1}{3}}(Z)\sin\tfrac{1}{6}\pi] + O(p^{-4/3}),$$

where

$$w = (x^2 p^{-2} - 1)^{\frac{1}{2}} \quad \text{and} \quad Z = p(w - \tan^{-1} w).$$

Let us now replace the function inside the brackets by an upper bound, say, $CZ^{-\frac{1}{3}}$, and express x in terms of w:

$$|x^{\frac{1}{3}} J_p(x)| < C\{\max_{0<w<p\epsilon^{-2}} [w^{-\frac{1}{3}}(1+w^2)^{\frac{1}{3}}(w - \tan^{-1} w)^{\frac{1}{3}}]p^{\frac{1}{3}} + O(p^{-\frac{1}{3}})\} < C\epsilon_0^{-\frac{1}{3}} p^{\frac{1}{3}} + O(p^{-\frac{1}{3}}), \quad \text{Q.E.D.}$$

[23] Reference 11, 7.13(3).

[24] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge University Press, New York, 1962), p. 206.

[25] Reference 11, 7.13 (32).

Proof of (A3):

$$x^{-\frac{1}{3}} < p < x(1 + \epsilon_1), \quad \epsilon_1 > p^{-\frac{1}{3}}.$$

We start again with a Langer's formula[26]

$$\pi J_p(x) = w^{-\frac{1}{3}}(\tanh^{-1} w - w)^{\frac{1}{3}} K_{\frac{1}{3}}(Z) + O(p^{-4/3}),$$

where

$$w = (1 - p^{-2}x^2)^{\frac{1}{2}}, \quad Z = p(\tanh^{-1} w - w).$$

Replacing now $K_{\frac{1}{3}}(Z)$ by the upper bound[11] $CZ^{-\frac{1}{3}}$, we achieve the evaluation as in the above proof. Q.E.D.

Proof of (A4):

$$p > x(1 + \epsilon_1), \quad \epsilon_1 \gg p^{-\frac{1}{3}}.$$

In this domain, Z is large: we can replace $K_{\frac{1}{3}}(Z)$ by[11] $CZ^{-\frac{1}{2}}e^{-Z}$, and the bound is $p^{\frac{1}{2}}O(p^{-4/3})$ or $O(p^{-5/6})$. Q.E.D.

Proof of (A5):

From the hypergeometric expansion[11]

$$J_p(Z) = \frac{(\frac{1}{2}Z)^p}{p!} \sum_{n=0}^{\infty} \frac{\Gamma(p+1)}{\Gamma(p+1+n)} \frac{1}{n!} (-\tfrac{1}{4}Z^2)^n,$$

we easily derive the following bound:

$$|J_p(Z)| < \frac{|\frac{1}{2}Z|^p}{p!} \sum_{n=0}^{\infty} \frac{1}{n!} \left|\frac{Z^2}{4p}\right|^n,$$

which straightforwardly yields (A5). Q.E.D

Absolute bound independent of p: From the formulas (A1)–(A5), it is clear that the following bound holds for any value of x:

$$|u_p(x)| < Cp^{\frac{1}{3}}. \tag{A6}$$

A.2. Proof of (1.33)

For the sake of simplicity, we assume $\alpha < \gamma$, and study the series S of the form

$$|S| = \sum_{0}^{r-2} \frac{|a_q|}{(r+\alpha)^2 - (q+\gamma)^2} + \sum_{r+1}^{\infty} \frac{|a_q|}{(q+\gamma)^2 - (r+\alpha)^2} + O(r^{-\beta-1}). \tag{A7}$$

The two parts of (1.36) can be bounded respectively by the two following integrals:

$$I_1 = C\int_1^{r-1} \frac{q^{-\beta}\,dq}{r^2 - q^2}, \tag{A8}$$

$$I_2 = C\int_{r+1}^{\infty} \frac{q^{-\beta}\,dq}{r^2 - q^2}. \tag{A9}$$

[26] Reference 11, 7.13 (34).

We perform the following transformations on (A8):

$$I_1 = r^{-\beta}\left[\int_1^{r-1} \frac{dq}{r^2-q^2} + \int_1^{r-1} \frac{dq}{q^\beta}\frac{r^\beta - q^\beta}{r^2-q^2}\right]$$
$$= r^{-\beta-1}\left\{\left[\log\frac{r+q}{r-q}\right]_1^{r-1} + \int_{r^{-1}}^{1-r^{-1}} \frac{1-x^\beta}{1-x^2}\,x^{-\beta}\,dx\right\}. \quad \text{(A10)}$$

If we observe that $(1-x^\beta)/(1-x)$ is bounded for $0 < x < 1$, we see that the second term of (A10) is bounded by $(C + Cr^{\beta-1})$ for $\beta > 0$, by C for $\beta \leq 0$, so that, for large values of r:

$$|I_1| < r^{-\beta-1}[C\log r + C] + Cr^{-2}. \quad \text{(A11)}$$

We can show in a very similar way that

$$|I_2| < r^{-\beta-1}[C\log r + C]. \quad \text{(A12)}$$

These results are written in slightly different forms in (1.33).

APPENDIX B

In this Appendix we give in closed forms some special series of Bessel functions and their evaluations.

B.1. Series $T_0(r, r')$

We start with the well-known formula[27]

$$Z^{\frac{1}{2}}J_0(Z\sin\alpha\sin\beta)e^{iZ\cos\alpha\cos\beta}$$
$$= (2\pi)^{\frac{1}{2}}\sum_{n=0}^{\infty} i^n(n+\tfrac{1}{2})J_{n+\frac{1}{2}}(Z)P_n(\cos\alpha)P_n(\cos\beta), \quad \text{(B1)}$$

which reduces for $\cos\beta = 0$ to

$$Z^{\frac{1}{2}}J_0(Z\sin\alpha) = \sqrt{2}\sum_{p=0}^{\infty}(2p+\tfrac{1}{2})J_{2p+\frac{1}{2}}(Z)$$
$$\times \Gamma(p+\tfrac{1}{2})(p!)^{-1}P_{2p}(\cos\alpha). \quad \text{(B2)}$$

If we put successively $Z = r$ and $Z = r'$ in (B2), and perform the scalar product of the two Legendre polynomials expansions thus obtained, we get

$$T_0(r, r') \equiv (rr')^{\frac{1}{2}}\sum_0^{\infty}(2p+\tfrac{1}{2})$$
$$\times [\Gamma(p+\tfrac{1}{2})]^2(p!)^{-2}J_{2p+\frac{1}{2}}(r)J_{2p+\frac{1}{2}}(r') \quad \text{(B3)}$$
$$= rr'\int_0^{\frac{1}{2}\pi} J_0(r\sin\alpha)$$
$$\times J_0(r'\sin\alpha)\sin\alpha\,d\alpha.$$

[27] Reference 11, 7.15 (43).

B.2. Series $T_{\bar{0}}(r, r')$

If we differentiate (B1) with respect to $\cos\alpha$, and put $\cos\alpha = 0$, we get

$$2\sqrt{2}\sum_{p=0}^{\infty}(2p+\tfrac{3}{2})\Gamma(p+\tfrac{3}{2})(p!)^{-1}J_{2p+\frac{3}{2}}(Z)P_{2p+1}(\cos\beta)$$
$$= Z^{\frac{1}{2}}\cos\beta J_0(Z\sin\beta). \quad \text{(B4)}$$

Recall now the following well-known formula[11]:

$$P_{2p+1}(x) = (2p+1)(2p+2)$$
$$\times \int_0^x d\tau\,(1-\tau^2)^{-1}\int_\tau^1 P_{2p+1}(\sigma)\,d\sigma. \quad \text{(B5)}$$

If we apply to (B4) the integral operator involved in (B5), we obtain, after one integration by parts:

$$\sum_{p=0}^{\infty}(2p+\tfrac{3}{2})\Gamma(p+\tfrac{1}{2})[(p+1)!]^{-1}J_{2p+\frac{3}{2}}(Z)P_{2p+1}(\cos\beta)$$
$$= (2Z)^{\frac{1}{2}}\int_0^{\cos\beta} d\tau(1-\tau^2)^{-\frac{1}{4}}J_1[Z(1-\tau^2)^{\frac{1}{2}}]. \quad \text{(B6)}$$

If now we put $Z = r$ in (B4) and $Z = r'$ in (B6), and perform the scalar product of these two functions, we get

$$T_{\bar{0}}(r, r') \equiv (rr')^{\frac{1}{2}}\sum_{p=0}^{\infty}(2p+\tfrac{3}{2})\frac{\Gamma(p+\tfrac{1}{2})\Gamma(p+\tfrac{3}{2})}{\Gamma(p+1)\Gamma(p+2)}$$
$$\times J_{2p+\frac{3}{2}}(r)J_{2p+\frac{3}{2}}(r') \quad \text{(B7)}$$
$$= rr'\int_0^{\frac{1}{2}\pi} J_1(r\sin\beta)J_1(r'\sin\beta)\sin\beta\,d\beta.$$

B.3. Closed forms for $T_0(r, r)$ and $T_{\bar{0}}(r, r)$

According to well-known formulas[28,29] T_0 and T_1 reduce for $r = r'$ to

$$T_0(r, r) = r\sum_{m=0}^{\infty}J_{2m+1}(2r) = \tfrac{1}{2}r\int_0^{2r} J_0(t)\,dt, \quad \text{(B8)}$$

$$T_{\bar{0}}(r, r) = r\sum_{m=0}^{\infty}J_{2m+3}(2r) = T_0(r, r) - rJ_1(2r). \quad \text{(B9)}$$

B.4. Asymptotic Expansions of T_0 and $T_{\bar{0}}$

In order to get them, we transform[30] formulas (B3) and (B7) with the help of the Parseval theorem:

$$T_0(r, r') = rr'\int_0^{\infty} w\chi(w)\psi(w),$$

[28] Reference 11, 7.14 (21), and see erratum of Ref. 11.

[29] *Handbook of Mathematical Functions* (National Bureau of Standards, AMS 55. Washington, D. C., 1964), formula 11.1.2.

[30] For an application of this method to similar integrals, see V. M. Kisler, Prikl. Matem. i Mekh. **24** (3), 496 (1960).

where

$$\chi(w) = \int_0^\infty J_0(r\xi)J_0(r'\xi)J_0(w\xi)\xi\, d\xi,$$

$$\psi(w) = \int_0^\infty J_0(w\xi)\xi(1-\xi^2)^{-\frac{1}{2}}Y(1-\xi)\, d\xi.$$

A similar transformation of $T_{\bar{0}}(r, r')$ and use of well-known formulas give

$$T_0(r, r') = 2\pi^{-1}rr' \times \int_a^b (\sin w)(b^2-w^2)^{-\frac{1}{2}}(w^2-a^2)^{-\frac{1}{2}}\, dw, \quad \text{(B10)}$$

$$T_{\bar{0}}(r, r') = -(4\pi)^{-1} \times \int_a^b (b^2-w^2)^{\frac{1}{2}}(w^2-a^2)^{\frac{1}{2}}\frac{d}{dw}(w^{-1}\sin w)\, dw, \quad \text{(B11)}$$

in which we put

$$a = |r-r'|; \qquad b = r+r'.$$

Integration by parts in (B11) and subtraction of the result from T_0 yield:

$$T_0 - T_{\bar{0}} = \pi^{-1}\int_a^b (\sin w)(w^2-a^2)^{\frac{1}{2}} \times (b^2-w^2)^{-\frac{1}{2}}\, dw \quad \text{(B12)}$$

$$= \pi^{-1}\left\{\int_a^b \sin w\, dw \times \frac{(b^2-a^2)^{\frac{1}{2}}}{(b^2-w^2)^{\frac{1}{2}}} - \int_a^b \sin w\, dw \times \frac{(b^2-a^2)^{\frac{1}{2}} - (w^2-a^2)^{\frac{1}{2}}}{(b^2-w^2)^{\frac{1}{2}}}\right\}. \quad \text{(B13)}$$

With the help of one integration by parts, the second term in (B13) can be transformed into

$$-\cos w + \int_a^b \frac{(\cos w)(b^2-a^2)^{\frac{1}{2}}}{(b^2-a^2)^{\frac{1}{2}} + (w^2-a^2)^{\frac{1}{2}}} \times \frac{w\, dw}{[(w^2-a^2)(b^2-w^2)]^{\frac{1}{2}}}. \quad \text{(B14)}$$

Replacing the first factor under the integral sign in (B14) by 1, we can show that (B14) is bounded by numbers independent of a and b. We separate now the first term of (B13) into two parts:

$$\pi^{-1}\int_0^b \sin w\, dw \frac{(b^2-a^2)^{\frac{1}{2}}}{(b^2-w^2)^{\frac{1}{2}}} - \pi^{-1}\int_0^a \sin w\, dw \frac{(b^2-a^2)^{\frac{1}{2}}}{(b^2-w^2)^{\frac{1}{2}}}. \quad \text{(B15)}$$

One integration by parts and a majorization of the results show that the second term can be bounded by numbers independent of a and b, whereas the first term can be given a closed form, involving the Struve function of $(r + r')$:

$$T_0 - T_{\bar{0}} = (rr')^{\frac{1}{2}}[\mathbf{H}_0(r+r') + O(r^{-\frac{1}{2}}r'^{-\frac{1}{2}})]. \quad \text{(B16)}$$

We find in the same way from (B10) and (B11):

$$T_0 + T_{\bar{0}} = \pi^{-1}\int_a^b \sin w\, dw \times (b^2-w^2)^{\frac{1}{2}}(w^2-a^2)^{-\frac{1}{2}} \quad \text{(B17)}$$

$$= \pi^{-1}\left\{\int_a^\infty \sin w\, dw \times \frac{(b^2-a^2)^{\frac{1}{2}}}{(w^2-a^2)^{\frac{1}{2}}} - \int_b^\infty \sin w\, dw \times \frac{(b^2-a^2)^{\frac{1}{2}}}{(w^2-a^2)^{\frac{1}{2}}} - \int_a^b \sin w\, dw \times \frac{(b^2-a^2)^{\frac{1}{2}} - (b^2-w^2)^{\frac{1}{2}}}{(w^2-a^2)^{\frac{1}{2}}}\right\}. \quad \text{(B18)}$$

Following exactly the same method as above, we can show easily that all the terms in (B18) are bounded independently of a and b, except the first one, which yields a Bessel function of $(r - r')$:

$$T_0 + T_{\bar{0}} = (rr')^{\frac{1}{2}}[J_0(r-r') + O(r^{-\frac{1}{2}}r'^{-\frac{1}{2}})]. \quad \text{(B19)}$$

B.5. Derived Series

We now define a function T' which is typical of all the series obtained from T_0 and $T_{\bar{0}}$ by differentiation:

$$T' = r\frac{\partial}{\partial r}r^{-1}T_0(r, r') = -\frac{\partial}{\partial r'}T_{\bar{0}}(r, r')$$

$$= -rr'\int_0^{\frac{1}{2}\pi} J_1(r\sin\theta)J_0(r'\sin\theta)\sin^2\theta\, d\theta. \quad \text{(B20)}$$

It is easy to evaluate this function as a sum of two terms:

$$-2\pi^{-1}(rr')^{\frac{1}{2}}\int_0^{\frac{1}{2}\pi} \sin(r\sin\theta - \tfrac{1}{4}\pi) \times \cos(r'\sin\theta - \tfrac{1}{4}\pi)\sin\theta\, d\theta + (rr')^{\frac{1}{2}}\int_0^{\frac{1}{2}\pi} [2\pi^{-1}\sin(r\sin\theta - \tfrac{1}{4}\pi) \times \cos(r'\sin\mathbf{x} - \tfrac{1}{4}\pi) - (rr')^{\frac{1}{2}}(\sin\theta) \times J_1(r\sin\theta)J_0(r'\sin\theta)]\sin\theta\, d\theta. \quad \text{(B21)}$$

With the help of the well-known[11] asymptotic behavior of Bessel functions, the second term of the right-hand side of (B21) is easily shown to be of

the order of $(C/r) + (C/r')$, whereas the first term can be given an exact closed form:

$$T' = \tfrac{1}{2}(rr')^{\frac{1}{2}}\{-J_1(|r - r'|)\ \mathrm{sgn}\ (r - r') + [2\pi^{-1} - \mathbf{H}_1(r + r')] + O(r^{-1}) + O(r'^{-1})\}. \quad \text{(B22)}$$

B.6. Series T_1

We now define a series typical of the class 1:

$$T_1(r, r') = \sum_0^\infty (2n + 1)^{-1} u_n(r) u_n(r'). \quad \text{(B23)}$$

Let us first study it for $(r = r')$; recall the two following well-known[11] formulas:

$$J_{n+\frac{1}{2}}(r) J_{n+\frac{1}{2}}(r) = \frac{2}{\pi} \int_0^{\frac{1}{2}\pi} J_{2n+1}[2r \sin \theta]\, d\theta,$$
$$\sum_0^\infty (2n + 1)^{-1} J_{2n+1}(r) = \pi/4\mathbf{H}_0(r). \quad \text{(B24)}$$

They give for $T_1(r, r)$ the following formula:

$$T_1(r, r) = \tfrac{1}{4}\pi r \int_0^{\frac{1}{2}\pi} \mathbf{H}_0(2r \sin \theta)\, d\theta. \quad \text{(B25)}$$

In order to evaluate the integral in (B25), we consider it as the sum of four terms:

$$\int_0^{\frac{1}{2}\pi} Y_0(2r \sin \theta)\, d\theta + \int_0^{r^{-\frac{1}{2}}} [\mathbf{H}_0(2r \sin \theta) - Y_0(2r \sin \theta)] \cos \theta\, d\theta + \int_0^{r^{-\frac{1}{2}}} [\mathbf{H}_0(2r \sin \theta) - Y_0(2r \sin \theta)](1 - \cos \theta)\, d\theta + \int_{r^{-\frac{1}{2}}}^{\frac{1}{2}\pi} [\mathbf{H}_0(2r \sin \theta) - Y_0(2r \sin \theta)]\, d\theta.$$

The first term can be computed exactly[31] and shown to be of the order of r^{-1}. The second term can be evaluated with well-known formulas[32] and shown to be of the order of $(\pi r)^{-1}[\log r^{\frac{1}{2}}] + O(r^{-1})$. The third term can be bounded by $Cr^{-\frac{1}{2}} \log r$. The fourth term can be evaluated with the help of the asymptotic expansions of the functions involved[33] and shown to be of the order of $(\pi r)^{-1} \log (r^{\frac{1}{2}}) + O(r^{-1})$.

As a result, we can write

$$T_1(r, r) \sim \tfrac{1}{4} \log r + O(Cte). \quad \text{(B26)}$$

In the same way, with a slightly more complicated algebra, it is possible, by differentiating $T_1(r, r)$ under the integral sign, to show that

$$\left|\frac{\partial}{\partial r} T_1(r, r)\right| < \frac{1}{4r} \log r. \quad \text{(B27)}$$

The formula (A26) can also be used to get a higher bound for $|T_1(r, r')|$: with the help of Schwarz's inequality, it is easy to show that

$$|T_1(r, r')| < C(\log r)^{\frac{1}{2}}(\log r')^{\frac{1}{2}}. \quad \text{(B28)}$$

[31] Reference 29, formula 11.4.9.
[32] Reference 29, formula 12.1.32.
[33] Reference 29, formula 12.1.29.

Separation of the Interaction Potential into Two Parts in Statistical Mechanics. II. Graph Theory for Lattice Gases and Spin Systems with Application to Systems with Long-Range Potentials*

G. STELL,† J. L. LEBOWITZ, S. BAER,‡ AND W. THEUMANN
Belfer Graduate School of Science, Yeshiva University, New York, New York
(Received 22 December 1965)

The methods developed in a previous paper for treating systems with a pair potential of the form $v(\mathbf{r}) = q(\mathbf{r}) + w(\mathbf{r})$ are here applied to lattice gases (isomorphic to Ising systems). We chose $q(\mathbf{r})$ to be the "hard-core" potential preventing the multiple occupancy of a lattice site and $w(\mathbf{r})$ the interaction between two particles (or parallel spins) separated by $\mathbf{r}$. The resulting graphical formalism is similar to that obtained by other authors exclusively for spin systems. We are thus able to connect their work with the general Mayer theory as it was originally applied to fluids and also to find new interpretations for some of the quantities appearing in the spin-system expansion. The formalism is then used in the case where $w(\mathbf{r})$ is a "Kac potential" of the form $w(\mathbf{r}, \gamma) \sim \gamma^{\nu} \varphi(\gamma \mathbf{r})$, where ν is the dimensionality of the space considered and γ^{-1} is the range or w, assumed very large. We then obtain systematic expansions in γ for the correlation functions and thermodynamic properties of the system. These expansions are, however, invalid inside the two-phase region and near the critical point of the "van der Waals" system; i.e., a system with $\gamma \to 0$. To remedy this we introduce a new self-consistent type of approximation which is suggested by our graphical analysis of the γ expansion but is applicable also to systems with general interactions $w(\mathbf{r})$, not necessarily parametrized by γ. The spatially asymptotic behavior of the two-body correlation function at the critical point is then discussed using these graphical methods. From the expansion procedures it seems possible to find specific subsets of graphs which will give any desired asymptotic behavior of the two-body correlation function including known exact ones. However, we could find no *a priori* reason for the retention of these subsets of graphs to the exclusion of all others.

I. INTRODUCTION

In Part I of this series[1] of papers, a systematic method was developed for investigating the correlation functions and the thermodynamic properties of a classical system of particles interacting via a pair potential $v(\mathbf{r})$, which may be usefully considered as composed of two distinct parts: a "short-range" part $q(\mathbf{r})$ and a "long-range" part $w(\mathbf{r})$,

$$v(\mathbf{r}) = q(\mathbf{r}) + w(\mathbf{r}). \tag{1.1}$$

It is the purpose of this paper to apply these methods, with new extensions, to a system of particles whose positions are confined to a regular lattice, i.e., a lattice gas.

The motivation for dividing $v(\mathbf{r})$ into two parts is to take advantage of the fact that in many cases the properties of the reference system, i.e., a system for which the interparticle potential is $q(\mathbf{r})$, are better known than those of the actual system with interaction $v(\mathbf{r})$. As shown in I, we can express the properties of the actual system in terms of $w(\mathbf{r})$ and the properties of the reference system by noting that the decomposition of $v(\mathbf{r})$ into the form (1.1) induces a corresponding decomposition of the correlation functions (i.e., the modified l-particle Ursell functions, $\hat{F}_l$) into short-ranged and long-ranged parts,

$$\hat{F}_l = \hat{F}_l^{\mathrm{s}} + \hat{F}_l^{\mathrm{L}}, \qquad l = 2, 3, \cdots .$$

The full $\hat{F}_l$ can be represented as a sum of graphs composed of "long-range" potential bonds that represent $\Phi(\mathbf{r}) \equiv -\beta w(\mathbf{r})$, $[\beta = 1/kT]$, and "hypervertices" that represent the $\hat{F}_l^{\mathrm{s}}$. The latter functions can in turn be expressed in terms of Φ and the correlation functions of the reference system.

Applying to lattice gases the formalism developed in I for continuum fluids, we identify (unless otherwise stated) the short-range part of the interparticle potential $q(\mathbf{r})$ with the "hard-core" repulsion, which excludes the multiple occupation of a lattice site,

$$q(\mathbf{r}) = \begin{cases} \infty, & \mathbf{r} = 0, \\ 0, & \mathbf{r} \neq 0. \end{cases} \tag{1.2}$$

$w(\mathbf{r})$ then represents the total *finite* interaction between two particles at different lattice sites. This identification of $q(\mathbf{r})$ and $w(\mathbf{r})$ greatly simplifies the structure of the "hypervertices," enabling us to

* This work was supported by the U. S. Air Force Office of Scientific Research under Grant 508-66.
† Present address: Polytechnic Institute of Brooklyn, New York.
‡ Permanent address: Hebrew University, Jerusalem, Israel.
[1] J. L. Lebowitz, G. Stell, and S. Baer, J. Math. Phys. **6**, 1282 (1965); referred to as I; e.g., Eq. (I-3.1) refers to Eq. (3.1) in I.

develop new methods for their evaluation. It also makes this lattice gas, confined to a volume Ω (i.e., Ω lattice sites) at fugacity z isomorphic to a system of Ising spins[2] interacting with a pair potential $w(\mathbf{R}-\mathbf{R}')\sigma(\mathbf{R})\sigma(\mathbf{R}')$ and subject to a uniform external magnetic field H with

$$z \equiv \exp(\beta H + \tfrac{1}{2}\beta\alpha'). \tag{1.3}$$

Here, $\mathbf{R}$ is a vector characterizing a particular lattice site, $\sigma(\mathbf{R})$ is a spin variable, $\sigma(\mathbf{R}) = \pm\tfrac{1}{2}$, and

$$\alpha' \equiv \sum_{\mathbf{R}'} w(\mathbf{R}-\mathbf{R}'). \tag{1.4}$$

Calling $\Xi(\beta, z, \Omega)$ the grand partition function of the lattice gas and $Q(\beta, H, \Omega)$ the canonical partition function of the spin system, we have

$$\Omega^{-1}\ln\Xi = \tfrac{1}{2}\ln z + \Omega^{-1}\ln Q - \tfrac{1}{8}\alpha'\beta. \tag{1.5}$$

In the thermodynamic limit $\Omega \to \infty$, (1.5) assumes the form

$$\beta p = \tfrac{1}{2}\ln z - \beta\Psi - \tfrac{1}{8}\alpha'\beta, \tag{1.6}$$

where p is the pressure of the lattice gas, and Ψ is the Helmholtz free energy per spin of the spin system.

The equality (1.5) is based on the relation between the microscopic density operator of the lattice gas $\rho(\mathbf{R})$, which can only assume the value zero or unity and the spin variable at the $\mathbf{R}$th site $\sigma(\mathbf{R})$,

$$\rho(\mathbf{R}) = \sum_i \delta(\mathbf{R}-\mathbf{r}_i) = \sigma(\mathbf{R}) + \tfrac{1}{2} = (0, 1), \tag{1.7}$$

where $\mathbf{r}_i$ is the position of the ith particle, and $\delta(\mathbf{r})$ is the Kroenecker delta function. It follows immediately from (1.7) that the one-particle distribution function of the lattice gas[3] is

$$n_1(\mathbf{R}) = \langle\rho(\mathbf{R})\rangle = \langle\sigma(\mathbf{R})\rangle + \tfrac{1}{2},$$
$$\langle N\rangle \equiv \sum_R n_1(\mathbf{R}) \equiv \rho\Omega = M + \tfrac{1}{2}\Omega, \tag{1.8}$$

where $\langle\sigma(\mathbf{R})\rangle$ is the average magnetization at the site $\mathbf{R}$, $\langle N\rangle$ is the average number of particles in the system, and M is the total magnetization. Similarly,[3]

$$\hat{F}_2(\mathbf{R}, \mathbf{R}') \equiv \langle\rho(\mathbf{R})\rho(\mathbf{R}')\rangle - \langle\rho(\mathbf{R})\rangle\langle\rho(\mathbf{R}')\rangle$$
$$= \langle\sigma(\mathbf{R})\sigma(\mathbf{R}')\rangle - \langle\sigma(\mathbf{R})\rangle\langle\sigma(\mathbf{R}')\rangle, \tag{1.9}$$

and generally, $\hat{F}_l(\mathbf{R}_1, \cdots, \mathbf{R}_l)$ is equal to the lth spin semi-invariant for $l \geq 2$. This isomorphism makes the entire formalism developed in I, specialized to lattice gases, immediately applicable to spin systems and directly comparable with other formalisms developed specifically for the Ising problem. It is one of the aims of the present work to show the way in which our formalism is related to and generalizes a number of specific Ising spin (lattice-gas) results previously derived by other authors.[4]

Our work on lattice systems will be presented in two parts to be referred to as II (this paper) and III (a paper to follow), with II devoted to formally rigorous results, and III to approximation methods.

In Sec. II of this paper, we obtain the general graphical expressions for the correlation functions and thermodynamic properties of a lattice system in terms of graphs with Φ bonds and $\hat{F}_l^s$ hypervertices. It is also shown there how to express the $\hat{F}_l^s$ as functionals of Φ and of the density derivatives of $\hat{F}_l^0$, the modified Ursell function of the reference system. For the situation considered here, with $q(\mathbf{r})$ given by (1.2), the reference system is an ideal lattice gas, making the $\hat{F}_l^0$ polynomials in the density ρ. Comparison with other work is also made here; the $\hat{F}_l^s$ coinciding with the cumulants $\bar{M}_l$ of Horwitz and Callen and of Englert.

While our expansions do not depend upon the introduction of any particular parametrization, they are especially well suited for use in the case in which $w(\mathbf{r})$ is a "Kac potential" containing a parameter γ

$$w(\mathbf{r}, \gamma) = \gamma^\nu\lambda\varphi(\gamma\mathbf{r}), \tag{1.10}$$

where ν is the dimensionality of the space considered. The value of γ thus corresponds to the inverse range of $w(\mathbf{r}, \gamma)$. We are able to identify the terms in our expansions that contribute to any given order in γ, and explicitly give the expansion of the free energy in terms of Φ and ρ through terms of order $(\gamma^\nu)^2$ (as well as the prescription for finding $\ln\Xi$ to any order). The result to order $(\gamma^\nu)^2$ agrees with that of Coopersmith and Brout[5]; our general result cannot be directly compared with theirs. At $\rho = \tfrac{1}{2}$ we can also compare our explicit result with that of Siegert[6] (who uses spin-system language and considers the case $H = 0$) and we find agreement.

In the limit $\gamma \to 0$, it was shown by Lebowitz and Penrose[7] [for a wide class of potentials $q(\mathbf{r})$ and $\varphi(\mathbf{y})$] that the Helmholtz free energy per unit volume $a(\beta, \rho, 0+) = \lim_{\gamma\to 0} a(\beta, \rho, \gamma)$, from which the other thermodynamic properties of the system

[2] C. N. Yang and T. D. Lee, Phys. Rev. 87, 404 (1952).
[3] J. L. Lebowitz and J. K. Percus, J. Math. Phys. 4, 1495 (1963); cf. also Sec. II.
[4] See Refs. 12, 13, 17–19.
[5] M. Coopersmith and R. Brout, Phys. Rev. 130, 2539 (1963).
[6] A. J. F. Siegert, "On the Ising Model with Long-Range Interaction," Northwestern University preprint (1962).
[7] J. L. Lebowitz and O. Penrose, J. Math. Phys. 7, 98 (1966).

may be obtained, is given rigorously by applying the Gibbs double-tangent construction to the function

$$a_0(\beta, \rho) = a^0(\beta, \rho) + \tfrac{1}{2}\alpha\rho^2. \tag{1.11}$$

Here, $a^0(\beta, \rho)$ is the free energy per unit volume of the reference system corresponding to $w(\mathbf{r}, \gamma) = 0$, and

$$\alpha = \lim_{\gamma\to 0} \sum_{\mathbf{R}'} w(\mathbf{R} - \mathbf{R}')$$

$$= \lim_{\gamma\to 0} \alpha'(\gamma) = \int \varphi(\mathbf{y})\, d\mathbf{y}. \tag{1.12}$$

For the lattice gas considered here, we have

$$a^0(\beta, \rho) = \beta^{-1}[\rho \ln \rho + (1 - \rho) \ln (1 - \rho)]. \tag{1.13}$$

When these results are translated into spin language, we obtain the Weiss self-consistent theory of magnetism, which is thus proven rigorously for a spin-interaction potential of the form (1.10) in the limit $\gamma \to 0$. This generalizes the results of Baker,[8] and Kac and Helfand,[9] who proved the Weiss theory for a one-dimensional spin system with a special type of Kac potential $w(\mathbf{r}, \gamma) = \frac{1}{2}\alpha\gamma e^{-\gamma|\mathbf{r}|}$; $\alpha < 0$. (Baker also considered similar potentials in three dimensions, cf. Sec. III.)

When we carry through our expansion in γ, we obtain

$$a(\beta, \rho, \gamma) = a_0(\beta, \rho) + \sum_{n=\nu}^{\infty} a_n(\beta, \rho, \gamma), \quad a_n \sim 0(\gamma^n). \tag{1.14}$$

For $\alpha < 0$, ferromagnetic interaction, the $a_n(\beta, \rho, \gamma)$ diverge for $n > \nu$ when β and ρ approach values corresponding to $(dp_0/d\rho)(\beta, \rho) = 0$, i.e., the boundary of the meta-stable region in the van der Waals–Maxwell (or Bragg–Williams) equation of state (cf. Fig. 1),

$$p_0(\beta, \rho) = \rho^2[d(a_0/\rho)/d\rho] = -\beta^{-1} \ln (1 - \rho) + \tfrac{1}{2}\alpha\rho^2. \tag{1.15}$$

The failure of the γ expansion inside the Bragg–Williams (BW) two-phase region, as well as in the vicinity of the BW critical point, leads us to propose in Sec. IV a nonperturbative self-consistent iterative approximation method for the evaluation of the $\hat{F}_l^*$ occurring in the graphical expansion of $\hat{F}_l$. To lowest order, this new method yields an $\hat{F}_2$ identical to that obtained from the (mean) spherical model of Lewis and Wannier[10] (this is a modification of the Berlin and Kac[11] spherical model). Explicit calculations with this approximation are carried out in III.

Higher-order approximations can be obtained in several ways, one of which involves an auxiliary function $\hat{W}$ (Eq. 2.22) that has independently been considered by Stillinger[12] and by Abe.[13] The latter used it in discussing the spacially asymptotic behavior of $\hat{F}_2$ at the critical point. Here we give a somewhat more general discussion than Abe's, pointing out the way in which various assumptions concerning the relationships between $\hat{W}$ and $\hat{F}_2$ are related to the spatially asymptotic behavior of $\hat{F}(\mathbf{r}_{12})$ at the critical point.

II. GENERAL GRAPHICAL FORMALISM FOR THE CORRELATION FUNCTIONS AND THERMODYNAMIC POTENTIALS

In this part we summarize, for lattice systems, the graphical description given in I for the modified m-particle Ursell function $\hat{F}_m(\mathbf{r}_1, \cdots, \mathbf{r}_m)$. The $\hat{F}_m$ are defined[3] in terms of the k-particle distribution functions $\hat{n}_k(\mathbf{r}_1, \cdots, \mathbf{r}_k)$ in the same way as the ordinary Ursell functions $F_m(\mathbf{r}_1, \cdots, \mathbf{r}_m)$ are defined in terms of the ordinary distribution functions $n_k(\mathbf{r}_1, \cdots, \mathbf{r}_k)$. The $\hat{n}_k(\mathbf{r}_1, \cdots, \mathbf{r}_k)$ differ from the $n_k(\mathbf{r}_1, \cdots, \mathbf{r}_k)$ by being the probability densities of finding k particles, *not necessarily distinct*, at positions $\mathbf{r}_1, \cdots, \mathbf{r}_k$ on the lattice. Thus

$$\hat{n}_1(\mathbf{r}_1) = n_1(\mathbf{r}_1) = \langle\rho(\mathbf{r}_1)\rangle,$$

$$\hat{n}_2(\mathbf{r}_1, \mathbf{r}_2) = \langle\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)\rangle$$

$$= n_2(\mathbf{r}_1, \mathbf{r}_2) + n_1(\mathbf{r}_1)\, \delta(\mathbf{r}_1 - \mathbf{r}_2), \cdots, \tag{2.1}$$

with $\delta(\mathbf{r})$ denoting (for lattice systems) the Kronecker delta function. Correspondingly,

$$\hat{F}_1(\mathbf{r}_1) = \hat{n}_1(\mathbf{r}_1) = F_1(\mathbf{r}_1), \tag{2.2}$$

$$\hat{F}_2(\mathbf{r}_1, \mathbf{r}_2) = \hat{n}_2(\mathbf{r}_1, \mathbf{r}_2) - \hat{n}_1(\mathbf{r}_1)\hat{n}_1(\mathbf{r}_2)$$

$$= F_2(\mathbf{r}_1, \mathbf{r}_2) + F_1(\mathbf{r}_1)\, \delta(\mathbf{r}_1 - \mathbf{r}_2), \quad \text{etc.}$$

Using the relationships (1.7) and (1.8) shows that

[8] G. A. Baker, Jr., Phys. Rev. **126**, 2071 (1962).
[9] M. Kac and E. Helfand, J. Math. Phys. **4**, 1078 (1963).
[10] H. W. Lewis and G. H. Wannier, Phys. Rev. **88**, 682 (1952). An extension of this model to lattice gases for which $q(\mathbf{r})$, the infinitely repulsive potential, is not confined to $\mathbf{r} = 0$ (and are thus not isomorphic to spin systems) has been made recently by Lebowitz and Percus (Ref. 21). This exact model leads to an integral equation for the radial distribution function which remains valid also in the continuum limit and coincides when $w(\mathbf{r}) = 0$ with the Percus–Yevick integral equation.
[11] T. Berlin and M. Kac, Phys. Rev. **86**, 821 (1952).
[12] F. H. Stillinger, Phys. Rev. **135**, A1646 (1964).
[13] R. Abe, Progr. Theoret. Phys. (Kyoto) **33**, 600 (1965).

$\hat{F}_l$ is equal to the lth semi-invariant of the spin system for $l \geq 2$. Thus, [cf. (1.9)],

$$\begin{aligned}\hat{F}_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) &= \hat{n}_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) - \hat{n}_1(\mathbf{r}_1)\hat{n}_2(\mathbf{r}_2, \mathbf{r}_3) \\ &- \hat{n}_1(\mathbf{r}_2)\hat{n}_2(\mathbf{r}_1, \mathbf{r}_3) - \hat{n}_1(\mathbf{r}_3)\hat{n}_2(\mathbf{r}_1, \mathbf{r}_2) \\ &+ 2\hat{n}_1(\mathbf{r}_1)\hat{n}_1(\mathbf{r}_2)\hat{n}_1(\mathbf{r}_3) = \langle\sigma(\mathbf{r}_1)\sigma(\mathbf{r}_2)\sigma(\mathbf{r}_3)\rangle \\ &- \langle\sigma(\mathbf{r}_1)\rangle\langle\sigma(\mathbf{r}_2)\sigma(\mathbf{r}_3)\rangle - \langle\sigma(\mathbf{r}_2)\rangle\langle\sigma(\mathbf{r}_1)\sigma(\mathbf{r}_3)\rangle \\ &- \langle\sigma(\mathbf{r}_3)\rangle\langle\sigma(\mathbf{r}_1)\sigma(\mathbf{r}_2)\rangle + 2\langle\sigma(\mathbf{r}_1)\rangle\langle\sigma(\mathbf{r}_2)\rangle\langle\sigma(\mathbf{r}_3)\rangle. \qquad (2.3)\end{aligned}$$

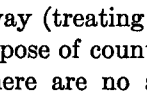
Starting with the usual representation of the $\hat{F}_l(\mathbf{r}_1, \cdots, \mathbf{r}_l)$ in terms of composite graphs with density ρ, or fugacity z vertices, and two type of bonds: "short-range" K-bonds, $K(\mathbf{x}_{12}) \equiv \{\exp[-\beta q(\mathbf{x}_{12})] - 1\}$ and "long-range" Φ-bonds, $\Phi(\mathbf{x}_{12}) \equiv -\beta w(\mathbf{x}_{12})$, cf. (I-2.8), we now divide each $\hat{F}_l$ into two parts; $\hat{F}_l^s$ ($\hat{F}_l$ short-range) and $\hat{F}_l^L$ ($\hat{F}_l$ long-range),

$$\hat{F}_l(\mathbf{r}_1, \cdots, \mathbf{r}_l) = \hat{F}_l^s(\mathbf{r}_1, \cdots, \mathbf{r}_l) + \hat{F}_l^L(\mathbf{r}_1, \cdots, \mathbf{r}_l). \qquad (2.4)$$

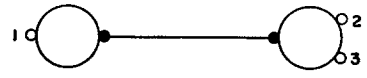
Here, $\hat{F}_l^s$ is the subset of all composite graphs in $\hat{F}_l$ in which there is a path, *consisting of K-bonds alone*, connecting the labeled points $\mathbf{r}_1, \cdots, \mathbf{r}_l$. The central graphical result in I, (I-2.15), states

> $\hat{F}_l(\mathbf{r}_1, \cdots, \mathbf{r}_l)$ = the sum of all irreducible graphs, with Φ bonds and $\hat{F}_k^s(\mathbf{x}_1, \cdots, \mathbf{x}_k)$ *hypervertices*, having l white circles labeled by $1, 2, \cdots, l$, respectively. (2.5)

A hypervertex of order k, which represents a function $w_k(\mathbf{x}_1, \cdots, \mathbf{x}_k)$, can be pictured as a large circle, along the circumference of which are attached k vertices (or points). The small vertices can be either black or white and correspond, respectively, to *field points* over which summations are performed, and to labeled points (root points). Each field point has *one* and only one Φ bond coming out of it going to another vertex. A graph is associated with its corresponding sum (over the field points) in the usual way (treating each hypervertex as a point for the purpose of counting). In (2.5), *irreducible* means that there are no articulation hypervertices. Also $\hat{F}_l^s(\mathbf{r}_1, \cdots, \mathbf{r}_l)$ is represented by a single hypervertex, e.g.,



$= \hat{F}_4^s(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4)$ and



$$= \sum_{\mathbf{x}_1} \sum_{\mathbf{x}_2} \hat{F}_2^s(\mathbf{r}_1, \mathbf{x}_1)\Phi(\mathbf{x}_{12})\hat{F}_3^s(\mathbf{x}_2, \mathbf{r}_2, \mathbf{r}_3).$$

For lattice gases with a short-range potential $q(\mathbf{r})$, defined in (1.2), $K(\mathbf{r}) \equiv -\delta(\mathbf{r})$, and hence the $\hat{F}_k^s(\mathbf{x}_1, \cdots, \mathbf{x}_k)$ can differ from zero only when all the k points coincide. We have $\hat{F}_1 = \hat{F}_1^s = \nu_1 = \rho$ and

$$\hat{F}_k^s(\mathbf{x}_1, \cdots, \mathbf{x}_k) = \nu_k\, \delta(\mathbf{x}_1 - \mathbf{x}_2) \times \delta(\mathbf{x}_1 - \mathbf{x}_3) \cdots \delta(\mathbf{x}_1 - \mathbf{x}_k), \qquad k \geq 2. \qquad (2.6)$$

This greatly simplifies the graphical description (2.5), since the hypervertices now become ordinary point vertices albeit with values ν_k depending on their order. (We therefore represent them as simple points when there is no danger of confusion.) It should be noted also that the ν_k depend on the interaction potential $w(\mathbf{r})$, since only the labeled (root) points in a hypervertex have to be connected by a path consisting of K bonds. In the reference system $w(\mathbf{r}) = 0$ the $\hat{F}_k$ and the ν_k then assume simple values ν_k^0,

$$\hat{F}_k^0(\mathbf{r}_1, \cdots, \mathbf{r}_k) = \nu_k^0\, \delta(\mathbf{r}_1 - \mathbf{r}_2)\, \delta(\mathbf{r}_1 - \mathbf{r}_3) \cdots \delta(\mathbf{r}_1 - \mathbf{r}_k). \qquad (2.7)$$

The values of ν_k^0 can be obtained easily from the definition of the $\hat{n}_k$ as expectation values of products of the microscopic density variable $\rho(\mathbf{x}_l)$ by noting from the definition of $\rho(\mathbf{x}_l)$, (1.7), that, *independent* of w, $\hat{n}_k(\mathbf{r}_1, \cdots, \mathbf{r}_k)$ for $\mathbf{r}_1 = \cdots = \mathbf{r}_k$,

$$\hat{n}_k(\mathbf{r}_1, \cdots, \mathbf{r}_1) \equiv \langle[\rho(\mathbf{r}_1)]^k\rangle = \langle\rho(\mathbf{r}_1)\rangle = \rho, \qquad (2.8)$$

with the last equality holding for a uniform system. Hence, considered as functions of ρ,

$$\begin{aligned}\nu_1^0 &= \hat{F}_1(\mathbf{r}_1) = \rho = \langle\sigma\rangle + \tfrac{1}{2}, \\ \nu_2^0 &= \hat{F}_2(\mathbf{r}_1, \mathbf{r}_1) = \nu_2 + \hat{F}_2^L(\mathbf{r}_1, \mathbf{r}_1) \\ &= \rho(1 - \rho) = \tfrac{1}{4} - \langle\sigma\rangle^2, \qquad (2.9) \\ \nu_3^0 &= \hat{F}_3(\mathbf{r}_1, \mathbf{r}_1, \mathbf{r}_1) = \nu_3 + \hat{F}_3^L(\mathbf{r}_1, \mathbf{r}_1, \mathbf{r}_1) \\ &= \rho(1 - \rho)(1 - 2\rho) = [\langle\sigma\rangle^2 - \tfrac{1}{4}]\langle\sigma\rangle, \quad \text{etc.}\end{aligned}$$

It follows from the definition of the $\hat{F}_k$'s, $k \geq 2$, that they (and thus also the ν_k^0) are even/odd functions of $\rho - \frac{1}{2}$, (i.e., $\langle\sigma\rangle$) for k even/odd. The same symmetry properties hold also for the ν_m. This can be proved as follows: The $\hat{F}_m(\mathbf{r}_1, \cdots, \mathbf{r}_m)$, when expressed graphically by (2.5), are functions of the ν_k's and Φ. Hence, we can write from (2.9),

$$\nu_m^0(\rho) = f_m(\nu_2, \nu_3, \nu_4 \cdots), \qquad m = 2, 3, \cdots. \qquad (2.10)$$

Equations (2.10) then provide solutions for the ν_k in terms of ρ. We now note that every graph in $\hat{F}_m$ which has l Φ-bonds and s_k vertices ν_k ($k = 2, 3, \cdots$), satisfies the relation $\sum_k k s_k = 2l + m$. Therefore

$$(-1)^m = (-1)^{\sum k s_k} = (-1)^{s_2 + s_3 + \cdots}. \qquad (2.11)$$

Here $s_3 + s_5 + \cdots$ is the total number of odd-order factors ν_k in the graph. If we rewrite our set of Eqs. (2.10) as

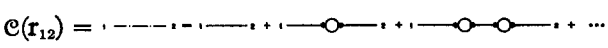
$$\nu_m^0(\rho) \equiv \nu_m^0(\langle\sigma\rangle) = f_m(\nu_3, \nu_5, \cdots \mid \nu_2, \nu_4, \cdots) \quad (2.12)$$

and multiply each by $(-1)^m$, we obtain with the help of (2.11)

$$\nu_m^0(-\langle\sigma\rangle) = f_m(-\nu_3, -\nu_5, \cdots \mid \nu_2, \nu_4, \cdots). \quad (2.13)$$

Solving the two sets of Eqs. (2.12) and (2.13) for the ν_k, we obtain the required result: the ν_m are even or odd functions of $\langle\sigma\rangle = \rho - \frac{1}{2}$ when m is even or odd, respectively, for $m \geq 2$.

It is convenient sometimes, in the graphical representation of $\hat{F}_l$, not to use Φ-bonds but $\mathfrak{C}$-bonds which are the sum of all chains of Φ-bonds and ν_2-hypervertices. Representing such bonds by dot–dash lines we have,



$$= \Phi(\mathbf{r}_{12}) + \sum_{\mathbf{r}_3} \sum_{\mathbf{r}_4} \Phi(\mathbf{r}_{13})\nu_2\, \delta(\mathbf{r}_3 - \mathbf{r}_4)\Phi(\mathbf{r}_{42}) + \cdots$$

$$= \Omega^{-1} \sum_{\mathbf{k}} \exp(i\mathbf{k}\cdot\mathbf{r}_{12})\tilde{\Phi}(\mathbf{k})/[1 - \nu_2\tilde{\Phi}(\mathbf{k})]$$

$$\underset{\Omega\to\infty}{\to} \left(\frac{1}{2\pi}\right)^{\nu} \int \exp(i\mathbf{k}\cdot\mathbf{r}_{12})\tilde{\Phi}(\mathbf{k})/[1 - \nu_2\tilde{\Phi}(\mathbf{k})]\, d\mathbf{k}, \quad (2.14)$$

where $\tilde{\Phi}(\mathbf{k})$ is the Fourier transform of[14] $\Phi(\mathbf{r})$,

$$\tilde{\Phi}(\mathbf{k}) = \sum_{\mathbf{r}_{12}} \exp(-i\mathbf{k}\cdot\mathbf{r}_{12})\Phi(\mathbf{r}_{12}). \quad (2.15)$$

with $\mathbf{k}$ confined to the first Brillouin zone of the reciprocal lattice, $\mathbf{k} = (2\pi/L)\mathbf{m}$, with $L = \Omega^{1/\nu}$, and the components of $\mathbf{m}$ taking on integer values between $-\frac{1}{2}L$ and $\frac{1}{2}L$. In terms of graphs with $\mathfrak{C}$-bonds, the equation analogous to (2.5) has the form

$\hat{F}_l(\mathbf{r}_1, \cdots, \mathbf{r}_l)$ = the sum of all irreducible graphs, with $\mathfrak{C}$ bonds and ν_k hypervertices having l white circles labeled by $1, 2, \cdots, l$, respectively, such that each hypervertex of second order ν_2 must contain at least one *labeled* point. (2.16)

A. Auxiliary Functions

We also introduce here, for lattice gases, the direct correlation function $C(\mathbf{r}_1, \mathbf{r}_2)$ of Ornstein and Zernike,[15] defined for uniform systems by the relation

$$F_2(\mathbf{r}_1 - \mathbf{r}_2) = \rho^2 C(\mathbf{r}_1 - \mathbf{r}_2) + \rho \sum_{\mathbf{r}_3} C(\mathbf{r}_1 - \mathbf{r}_3)F_2(\mathbf{r}_3 - \mathbf{r}_2). \quad (2.17)$$

The function

$$\hat{C}(\mathbf{r}_1, \mathbf{r}_2) = \rho^{-1}\, \delta(\mathbf{r}_1 - \mathbf{r}_2) - C(\mathbf{r}_1, \mathbf{r}_2) \quad (2.18)$$

is the matrix inverse of $\hat{F}_2(\mathbf{r}_1, \mathbf{r}_2)$,

$$\bar{C}(\mathbf{k}) = [\bar{F}_2(\mathbf{k})]^{-1}, \quad (2.19)$$

where $\bar{C}(\mathbf{k})$ and $\bar{F}_2(\mathbf{k})$ are the Fourier transforms of $\hat{C}(\mathbf{r})$ and $\hat{F}_2(\mathbf{r})$. $C(\mathbf{r}_1, \mathbf{r}_2)$ [or $\hat{C}(\mathbf{r}_1, \mathbf{r}_2)$] has a simple graphical representation in terms of graphs with ρ vertices and Mayer f bonds,[16] and can also be divided, in analogy with $\hat{F}_2(\mathbf{r}_1, \mathbf{r}_2)$, into two parts (cf. I, Sec. V)

$$\hat{C}(\mathbf{r}_1, \mathbf{r}_2) = \hat{C}^s(\mathbf{r}_1, \mathbf{r}_2) + \hat{C}^L(\mathbf{r}_1, \mathbf{r}_2), \quad (2.20)$$

and we have, for lattice systems with $q(\mathbf{r})$ given by (1.2),

$$\hat{C}^s(\mathbf{r}_1, \mathbf{r}_2) = \nu_2^{-1}\, \delta(\mathbf{r}_1 - \mathbf{r}_2). \quad (2.21)$$

Unfortunately, however, the long-range part of C, $C^L = -\hat{C}^L$, cannot be given a simple representation in terms of graphs with Φ bonds and ν_k hypervertices. For this purpose we introduced in I, Eq. (I-5.23), a different function, $W(\mathbf{r})$, closely related to $C(\mathbf{r})$, which does have such an expansion. Calling $\hat{W}(\mathbf{r}_{12}) = \rho^2 W(\mathbf{r}_{12}) + \rho\delta(\mathbf{r}_{12})$, we have for its Fourier transform $\bar{W}(\mathbf{k})$,

$$\bar{W}(\mathbf{k}) = [\bar{C}(\mathbf{k}) + \tilde{\Phi}(\mathbf{k})]^{-1} = \bar{F}_2(\mathbf{k})/[1 + \tilde{\Phi}(\mathbf{k})\bar{F}_2(\mathbf{k})] \quad (2.22)$$

or

$$\bar{F}_2(\mathbf{k}) = \bar{W}(\mathbf{k})/[1 - \bar{W}(\mathbf{k})\tilde{\Phi}(\mathbf{k})]. \quad (2.23)$$

[14] We are dealing here with simple cubic lattices in ν dimensions of unit spacing. The lattice is assumed wrapped on a torus of sides with length L, $\Omega = L^\nu$, L an even integer. The components of $\mathbf{R}$ then take all (integer) values from $-\frac{1}{2}L$ to $\frac{1}{2}L$; the two end points coinciding and $w(\mathbf{R} - \mathbf{R}') = w(\mathbf{r})$, an even functions of $\mathbf{r}$, has to be defined for the components of $\mathbf{r}$ assuming all (integer) values from $-L$ to L, and then periodic with periodicity L. This is readily done by having first $w(\mathbf{r})$ defined over the infinite lattice, e.g., $w(\mathbf{r}) \sim e^{-\gamma|\mathbf{r}|}$ then, for finite Ω, setting the interaction $w(\mathbf{r}; \Omega) = \Omega^{-1} \Sigma_{\mathbf{k}}\, e^{i\mathbf{k}\cdot\mathbf{r}}\, \tilde{w}(k)$ where $\tilde{w}(k) \equiv \Sigma_{\mathbf{r}}\, e^{-i\mathbf{k}\cdot\mathbf{r}}\, w(\mathbf{r})$, the summation over $\mathbf{r}$ being over an infinite lattice and $\mathbf{k}$ restricted to the first reciprocal Brillouin zone, $\mathbf{k} = (2\pi/L)\,\mathbf{n}$, the components of $\mathbf{n}$ being integers going from $-\frac{1}{2}L$ to $\frac{1}{2}L$. Since this dependence of w on Ω produces no effect in the thermodynamic limit, we do not write it out explicitly (cf. also Ref. 6). We also ignore, for this reason, graphs which "wind around the torus" and vanish when $L \to \infty$ (cf. also Ref. 10).

[15] L. S. Ornstein and F. Zernike, Proc. Akad. Sci. Amsterdam 17, 793 (1914). Reprinted in *The Equilibrium Theory of Classical Fluids*, H. L. Frisch and J. L. Lebowitz, Eds. (W. A. Benjamin, Inc., New York, 1964).

[16] See, e.g., G. Stell in Ref. 15.

Now, $\hat{W}(\mathbf{r}) = \hat{W}^s(\mathbf{r}) + \hat{W}^L(\mathbf{r})$ (with $\hat{W}^s = \hat{F}_2^s$) has the same graphical expansion as $\hat{F}_2(\mathbf{r})$, Eq. (2.5) or Eq. (2.16), with the *additional* restrictions that no graph contain any cutting bonds, i.e., bonds whose removal separates the graph into two parts, each of which contains a white vertex.

In the case of a field-free ($H = 0$) lattice system, the function $\hat{W}(\mathbf{r})$ reduces (when $r \neq 0$) to the function $\frac{1}{4}[w(\mathbf{r})]$ introduced by Stillinger,[12] and coincides with the function $I(\mathbf{r})$ introduced by Abe.[13] (Both authors consider only the field-free case.)

B. Thermodynamic Properties

The thermodynamic properties of our system may be obtained from $\hat{F}_2$ in several ways (cf. Sec. VI of I), chief among these being the fluctuation–compressibility relation[15]

$$\rho/\beta(dp/d\rho) = \sum \hat{F}_2(\mathbf{r}) = \hat{F}_2(0), \tag{2.24}$$

and the energy relation

$$u(\beta, \rho) = \tfrac{1}{2}\rho^2\alpha' + \tfrac{1}{2}\sum w(\mathbf{r})\hat{F}_2(\mathbf{r}) = \partial\beta a(\beta, \rho)/\partial\beta, \tag{2.25}$$

where $u(\beta, \rho)$ and $a(\beta, \rho)$ are the internal and Helmholtz free energies per unit volume, and use has been made of the fact that $w(0) = 0$. All the thermodynamic properties of the system follow from $a(\beta, \rho)$, with

$$a(\beta, \rho) = a_0'(\beta, \rho) + \tfrac{1}{2}\beta^{-1}\sum w(\mathbf{r}) \times \int_0^\beta \hat{F}_2(\mathbf{r}; \beta', \rho)\, d\beta', \tag{2.26}$$

where

$$a_0'(\beta, \rho) = a^0(\beta, \rho) + \tfrac{1}{2}\alpha'\rho^2$$

with a^0 given in (1.13) and α' defined in (1.4).

An explicit graphical representation for $\beta p = \Omega^{-1} \ln \Xi$ was derived in Appendix C of I, which yields, for lattice systems,

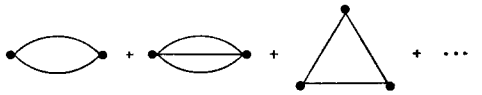
$$\ln \Xi = \Omega\beta p = \Omega\beta p_0 + S - \sum_{n\geq 2} \nu_n G_n(\mathbf{r}_1, \cdots, \mathbf{r}_n)|_{\mathbf{r}_1=\cdots=\mathbf{r}_n}, \tag{2.27}$$

where

$$G_n(\mathbf{r}_1, \cdots, \mathbf{r}_n) = \frac{1}{n!}\sum_{x_i}\prod_i \Phi(\mathbf{r}_i, \mathbf{x}_i)\hat{F}_n(\mathbf{x}_1, \cdots, \mathbf{x}_n) \tag{2.28}$$

and

S = the sum of all irreducible unrooted graphs, consisting of Φ-bonds and at least two ν_k-hypervertices and two Φ-bonds =



and $\beta p_0 = -\ln(1-\rho) + \frac{1}{2}\beta\alpha'\rho^2$, as given by (1.15). (2.29)

C. Evaluation of the ν_k

In order to be able to use Eqs. (2.5) and (2.27) to express the $\hat{F}_n$ and $\ln \Xi$ in terms of Φ and quantities that refer only to the reference system, it is necessary to have a prescription that enables us to express the ν_k in such terms. The Eq. (I-2.18), derived and discussed in I, yields this prescription when combined with (2.5). For our lattice system (I-2.18) can be written as

$$\nu_l = \mathfrak{N} \exp\left\{\sum_{k\geq 2}\left(\frac{1}{k!}\right)[\hat{F}_k^{vL}(0)]\frac{\partial^k}{\partial\rho^k}\right\}\nu_l^0, \tag{2.30}$$

where $\mathfrak{N}$ indicates a normal order in which all derivatives go to the right before evaluation, and $\hat{F}_k^{vL}(0) = \hat{F}_k^{vL}(\mathbf{r}_1, \cdots, \mathbf{r}_k)$ for $\mathbf{r}_1 = \cdots = \mathbf{r}_k$ with $\hat{F}_k^{vL}$ the "very long"-range part of $\hat{F}_k$ defined as

> $\hat{F}_k^{vL}$ is given by the same graphs, Eq. (2.5), that we use for $\hat{F}_k$ itself *except* that we exclude all graphs in which any two labeled circles are shared by a single hypervertex. (2.31)

(For $k = 2$, $\hat{F}_k^{vL}$ and $\hat{F}_k^L$ coincide.) Actually, for our lattice system the right-hand side of (2.30) contains only a finite number of terms for every l, since ν_k^0 is a polynomial in ρ of order k, so that $\partial^l\nu_k^0/\partial\rho^l = 0$ for $l > k$. As a result we have

$$\begin{aligned}\nu_2 &= \nu_2^0 + \tfrac{1}{2}(\partial^2\nu_2^0/\partial\rho^2)\hat{F}_2^L(0) = \nu_2^0 - \hat{F}_2^L(0),\\ \nu_3 &= \nu_3^0 + \tfrac{1}{2}(\partial^2\nu_3^0/\partial\rho^2)\hat{F}_2^L(0) + \tfrac{1}{6}(\partial^3\nu_3^0/\partial\rho^3)\hat{F}_3^{vL}(0)\\ &= \nu_3^0 - 3(1-2\rho)\hat{F}_2^L(0) + 2\hat{F}_3^{vL}(0).\end{aligned} \tag{2.32}$$

We can also get the above equations for ν_k by using (2.9) and (2.10) to solve for the ν_k in terms of the ν_m^0, but (2.30) is more direct and comes from the general relation Eq. (I-2.18).

The equations considered in this section are not based on any particular parameterization or ordering scheme. Nevertheless, the detailed results of eliminating the ν_k in favor of ν_k^0 and $\partial^l\nu_k^0/\partial\rho^l$ through the use of (2.30) are only of use if we have some means of estimating the relative importance of the terms in the series that we finally obtain. One such

means involves the use of a range parameter γ, which we discuss in the next section.

D. Use of Fugacity, (or H), as the Independent Variable

The functions ν_k and ν_k^0 are natural functions to use when the properties of the system are to be expressed in terms of ρ and Φ or, in spin-system language, $\langle\sigma\rangle$ and Φ. However, H or z rather than $\langle\sigma\rangle$ is more likely to be the independent variable of interest when the system is being used as a model for a magnet. To obtain expansions in terms of z and Φ, we use the expansions in I that contain $\hat{F}_k'$ hypervertices instead of $\hat{F}_k^*$ hypervertices. Applying the results of I to a lattice system then yields, instead of (2.5),

$$\hat{F}_l = \text{the sum of all connected graphs with } \Phi \text{ bonds and } \hat{F}_k' \text{ hypervertices, having } l \text{ white circles labeled by } 1, 2, \cdots, l, \text{ respectively.} \tag{2.33}$$

Here, $\hat{F}_k'(\mathbf{r}_1, \cdots, \mathbf{r}_k)$ is the subset of all composite graphs with z-vertices and K and Φ-bonds in $\hat{F}_k$, in which there is a path consisting of K-bonds alone, connecting *every* pair of points in the graph. The last requirement implies that, for the lattice gases considered here, $K(\mathbf{r}) = -\delta(\mathbf{r})$, all the z-vertices in $\hat{F}_k'$ must coincide. Hence, since we have assumed that $\Phi(0) = 0$, $\hat{F}_k'$, considered as a function of the fugacity z, must be equal to its value in the reference system

$$\hat{F}_k'(\mathbf{r}_1, \cdots, \mathbf{r}_k; z) = \hat{F}_k^0(\mathbf{r}_1, \cdots, \mathbf{r}_k; z) = \mu_k(z)\, \delta(\mathbf{r}_2 - \mathbf{r}_1) \cdots \delta(\mathbf{r}_k - \mathbf{r}_1). \tag{2.34}$$

The μ_k may now be found directly from the properties of the reference system (ideal lattice gas). Introducing the variable $x = \ln z = \beta H + \frac{1}{2}\beta\alpha'$, we have

$$\mu_k = \partial^k[\ln(1+z)]/\partial(\ln z)^k = (\partial^k/\partial x^k)[\ln(2\cosh x/2) + x/2], \tag{2.35}$$

so

$$\partial^l \mu_k/\partial x^l = \mu_{k+l} \tag{2.36}$$

and

$$\mu_1 = z/1+z = \tfrac{1}{2}\tanh(\tfrac{1}{2}x) + \tfrac{1}{2},$$
$$\mu_2 = z/(1+z)^2 = \tfrac{1}{4}\operatorname{sech}^2(\tfrac{1}{2}x),$$
$$\mu_3 = z(1-z)/(1+z)^3 = -2(\mu_1 - \tfrac{1}{2})\mu_2, \cdots. \tag{2.37}$$

Instead of (2.30) we have [from (I-C-6)]

$$\nu_l = \mathfrak{N}\{\exp[\sum_{k\geq 1} G_k(0)(\partial^k/\partial x^k)]\}\mu_l, \tag{2.38}$$

where G_k is defined in (2.28) and

$$G_k(0) = G_k(\mathbf{r}_1, \cdots, \mathbf{r}_k) \quad \text{for} \quad \mathbf{r}_1 = \cdots = \mathbf{r}_k. \tag{2.39}$$

Finally, instead of (2.27), we have [from (I-C-3)]

$$\ln \Xi = (\ln \Xi)^0 + \text{the sum of all connected graphs with } \Phi \text{ bonds and at least two unlabeled } \mu_k \text{ hypervertices,} \tag{2.40}$$

where

$$\Omega^{-1}(\ln \Xi)^0 = \ln(1+z), \tag{2.41}$$

so that, from (1.2), we have the expression that gives the free energy of the spin system:

$$\ln Q = \ln(1+z) - \tfrac{1}{2}\ln z + \text{the sum of Eq. (2.40).} \tag{2.42}$$

E. Comparison with Other Work

Having obtained the μ_k-hypervertex and ν_k-hypervertex expressions from the general results of I, we can make contact with the spin-system expansions of others.

Although Eq. (2.40) and (2.33) have the same graphical structure as the "unrenormalized" linked-cluster expansions that have been derived by others[4] especially for spin systems, they are not identical, graph by graph, to the latter owing to a difference in the functions represented by the hypervertices. The semi-invariants used most often in discussing the spin systems—for example, Englert's[17] $M_k^0(x')$ or Stillinger's D_k—are somewhat different functions [e.g., after adjusting the units, $M_1^0(x') = \frac{1}{2}\tanh(x'/2)$ instead of $\mu_1 = \frac{1}{2}\tanh(x/2) + \frac{1}{2}$] of somewhat different arguments ($x' = \beta H$ instead of $x = \beta H + \frac{1}{2}\beta\alpha'$), the general relation being

$$\mu_k(x) = M_k^0(x' + \tfrac{1}{2}\beta\alpha') + \tfrac{1}{2}\,\delta_{k,1}. \tag{2.43}$$

These two differences just compensate one another to lead to the same final graphical prescription whether the hypervertices represent the μ_k's or the M_k^0's.

On the other hand, in comparing the ν_k's with the "renormalized" semi-invariants—for example, Englert's M_k—these differences are no longer found. There is, at most, only the trivial difference of notation; in a spin system, it is natural to express the ν_k as a polynomial in the long-range order $R = 1 - 2\rho$ instead of ρ itself.[5] Whether ν_k is expressed as the function $\nu_k(\rho)$ or as $M_k(R)$, each graph in the expansion (2.16) represents the same function of ρ and Φ.

[17] F. Englert, Phys. Rev. 129, 567 (1963).

The $\bar{M}_k$ of Horwitz and Callen[18] would also coincide with ν_k if their renormalization procedure to obtain $\bar{M}_k$ to all orders were explicitly carried through (which is indeed what Englert did) and in the field-free case ($H = 0$) our ν_k also appear to reduce to Stillinger's $\tilde{D}_k$.

Our ν_k^0 are related to the vertex functions appearing in the expansions of Brout,[19] and of Coopersmith and Brout,[5] in the sense that they depend explicitly on ρ or R rather than on z or H, but there is a technical difference resulting from our use of the $\partial^k \nu_l^0/\partial \rho^k$ instead of the combinations of Kronecker δ's and ν_k^0's that appear in the expansions of those authors.

We have thus established the connection between the general expansions that were derived in I for an arbitrary decomposition of a potential into two parts and the Ising spin-system expansions heretofore derived by means of procedures that are immediately applicable only to those systems. The connection is made via lattice systems for which the reference potential is identified with the exclusion of multiple occupancy of a single site (i.e., the association of a single spin to each site). Our method has the advantage of giving a simple direct interpretation to the hypervertices ν_k in terms of the short-range part of $\hat{F}_k$, as well as suggesting some new generalizations.

One natural generalization that our expansions suggest is the consideration of a wider class of reference systems. One might, for example, use the exactly solvable 2-dimensional Ising model with nearest-neighbor interaction as a reference system, and introduce a further interaction as a perturbation. Alternatively, the nearest-neighbor potential itself could be considered as part reference potential and part perturbing potential to facilitate the development of new approximation schemes. This is the starting point of an approximation scheme developed by G. Horwitz.[20] For such reference systems, the $\hat{F}_k^s$ no longer have the simple form (2.6), but all our formal results immediately apply.

III. LONG-RANGE KAC POTENTIALS

Following the analysis developed in I, we now consider the case where $w(\mathbf{r})$, which was arbitrary so far, contains an inverse range parameter γ which can approach zero (*after* the size of the system has become infinitely large). Following Baker,[8] we shall use, for lattice systems, a slightly modified "Kac potential" of form (1.10) for $w(\mathbf{r}, \gamma)$,

$$w(\mathbf{r}, \gamma) = \begin{cases} \gamma^{\nu}\lambda(\gamma)\varphi(\gamma \mathbf{r}), & r \neq 0, \\ 0, & r = 0, \end{cases} \tag{3.1}$$

where ν is the dimensionality of the space considered,[14] and $\lambda(\gamma)$ is so chosen that

$$\lim_{\gamma \to 0} \lambda(\gamma) = 1 \tag{3.2}$$

and the "integrated strength" of the long-range potential

$$\sum w(\mathbf{r}, \gamma) \equiv \alpha' = \lambda(\gamma) \sum{}' \varphi(\gamma \mathbf{r})\gamma^{\nu} = \alpha \tag{3.3}$$

is independent of γ. The last summation is over an infinite lattice with the $\mathbf{r} = 0$ term omitted. We assume for simplicity[14] that we are dealing with simple cubic lattices of unit spacing, so that in the limit $\gamma \to 0$, (3.3) becomes the integral of $\varphi(\mathbf{x})$ over all space, coinciding with (1.12). [The advantage of introducing the multiplying factor $\lambda(\gamma)$ which has only trivial effects for $\gamma \to 0$ is that it permits simultaneous consideration also of very large γ, in which case $w(\mathbf{r}, \gamma)$ becomes a nearest-neighbor potential with the "integrated strength" of the potential remaining fixed.] We generally leave λ as a parameter and consider its explicit dependence on γ only at the end. It was shown in I how to obtain an expansion of $\hat{F}_2$ (and other $\hat{F}_l$) in powers of γ [for well-behaved $\varphi(\mathbf{y})$'s, cf. (I-3.1) and (I-3.3)]. For this purpose, it is necessary to treat $\hat{F}_2^s(\mathbf{r}, \gamma)$ and $\hat{F}_2^L(\mathbf{y}, \gamma)$ differently, considering them, respectively, as functions of $\mathbf{r}$ and γ, and as functions of $\mathbf{y}$ and γ; $\mathbf{y} \equiv \gamma \mathbf{r}$. This difference is completely obvious for the lattice systems considered here, where $\hat{F}_2^s(\mathbf{r}, \gamma) = \nu_2(\gamma)\delta(\mathbf{r})$, and need not, therefore, always be kept track of explicitly, as long as we use the convention that $\delta(\mathbf{y}) = \delta(\mathbf{r})$.

To facilitate the γ expansion of $\hat{F}_l$, two kinds of ordering, γ and Γ ordering, of the graphs entering into $\hat{F}_l$, were introduced in I. In the γ-ordering, we classify all composite graphs with *density vertices*, and Φ- and K-bonds according to the difference between the number t of Φ-bonds and the number of free integrations m occurring in the graph. The number of free integrations in a graph is found by deleting all the Φ-bonds in a graph for $\hat{F}_l$ and counting the number of separate pieces not connected to any root point by a K-bond. All such graphs are then of $O(\gamma^{\nu})^{t-m}$. We then have

$$\hat{F}_l = \hat{F}_{l[0]} + \hat{F}_{l[1]} + \cdots \tag{3.4}$$

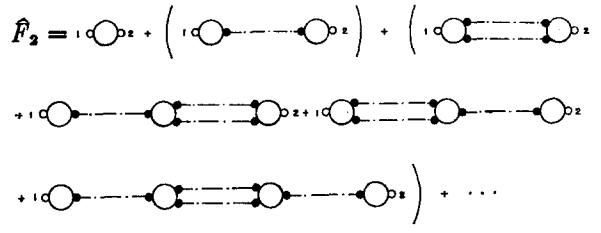
[18] G. Horwitz and H. B. Callen, Phys. Rev. 124, 1757 (1961).
[19] R. Brout, Phys. Rev. 115, 824 (1959); *ibid.* 118, 1009 (1960).
[20] G. Horwitz (to be published).

with

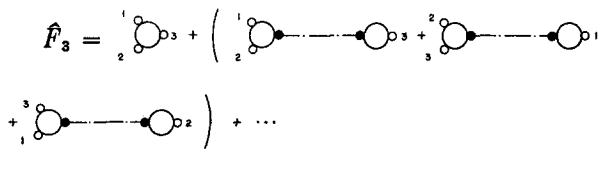
$$\hat{F}_{l[k]}, \qquad k \equiv t - m \text{ of } O(\gamma^r)^k,$$

$$\hat{F}_{l[k]} = \sum_{j=\nu k}^{\infty} \gamma^j \hat{F}_{l[k]j}. \tag{3.5}$$

The Γ ordering applies directly to graphs with Φ-bonds (or $\mathcal{C}$-bonds) and $\hat{F}_k^s$ hypervertices entering $\hat{F}_l$ in (2.5). These are ordered according to the difference j between the number of Φ bonds (or $\mathcal{C}$ bonds) and the number of hypervertices containing no labeled vertices in the graph. Calling such a subset $\Gamma_j^{(l)}$, we have



$$= \Gamma_0(\mathbf{r}_{12}, \gamma) + \Gamma_1(\mathbf{y}_{12}, \gamma; \nu_2) + \Gamma_2(\mathbf{y}_{12}, \gamma; \nu_2, \nu_3) + \cdots, \tag{3.6}$$



$$= \Gamma_0^{(3)}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \gamma) + \Gamma_1^{(3)}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \gamma; \nu_2, \nu_3) + \cdots, \tag{3.7}$$

where we have set $\Gamma_j^{(2)} = \Gamma_j$. Writing out the first few terms more explicitly yields

$$\hat{F}_2^s(\mathbf{r}) = \Gamma_0(\mathbf{r}, \gamma) = \nu_2(\gamma)\, \delta(\mathbf{r}), \tag{3.8}$$

$$\Gamma_1(\mathbf{y}, \gamma) = \nu_2^2 \mathcal{C}(\mathbf{y}, \gamma) = \left(\frac{\gamma}{2\pi}\right)^r \int_{-\pi/\gamma}^{\pi/\gamma} \cdots \int e^{i\boldsymbol{\kappa}\cdot\mathbf{y}} \frac{\nu_2^2 \tilde{\Phi}(\boldsymbol{\kappa}, \gamma)}{1 - \nu_2 \tilde{\Phi}(\boldsymbol{\kappa}, \gamma)}\, d\boldsymbol{\kappa}, \tag{3.9}$$

with

$$\tilde{\Phi}(\boldsymbol{\kappa}, \gamma) = \sum_{\mathbf{y}}{}' e^{-i\boldsymbol{\kappa}\cdot\mathbf{y}} \Phi(\mathbf{y}) = -\beta\lambda(\gamma) \sum_{\mathbf{y}}{}' e^{-i\boldsymbol{\kappa}\cdot\mathbf{y}} \varphi(\mathbf{y}) \gamma^r \underset{\gamma\to 0}{\to} -\beta \int e^{-i\boldsymbol{\kappa}\cdot\mathbf{y}} \varphi(\mathbf{y})\, d\mathbf{y} = \tilde{\Phi}(\boldsymbol{\kappa}), \tag{3.10}$$

the summation over $\mathbf{y} = \gamma\mathbf{r}$ being over a lattice with spacing γ, and the integration over $\boldsymbol{\kappa} \equiv \gamma^{-1}\mathbf{k}$ having a range $-\pi/\gamma$ to π/γ (after taking the thermodynamic limit $\Omega \to \infty$). $\tilde{\Phi}(\boldsymbol{\kappa}, \gamma)$ is equal to $\tilde{\Phi}(\mathbf{k}) = -\beta\tilde{w}(\mathbf{k})$ defined in (2.15), which we shall also write sometimes as $\tilde{\Phi}(\mathbf{k}, \gamma) = -\beta\tilde{w}(\mathbf{k}, \gamma)$; the range of $\mathbf{k} = \gamma\boldsymbol{\kappa}$ always being $-\pi$ to π. When we go to the limit $\gamma \to 0$, $\tilde{\Phi}(\boldsymbol{\kappa}, \gamma) \to \tilde{\Phi}(\boldsymbol{\kappa})$ which coincides with the continuum Fourier transform of $-\beta\varphi(\mathbf{y})$ used in I. Similarly,

$$\Gamma_2(\mathbf{y}, \gamma) = \frac{1}{2!}\left(\frac{\gamma}{2\pi}\right)^r \times \int_{-\pi/\gamma}^{\pi/\gamma} \cdots \int d\boldsymbol{\kappa}\, \frac{e^{i\boldsymbol{\kappa}\cdot\mathbf{y}} \nu_3^2 \tilde{\mathcal{S}}(\boldsymbol{\kappa}, \gamma)}{[1 - \nu_2 \tilde{\Phi}(\boldsymbol{\kappa}, \gamma)]^2}, \tag{3.11}$$

where $\tilde{\mathcal{S}}$, the Fourier transform of $\mathcal{C}^2(\mathbf{y}, \gamma)$, is given by

$$\tilde{\mathcal{S}}(\kappa, \gamma) = \left(\frac{\gamma}{2\pi}\right) \int d\boldsymbol{\kappa}'\, \tilde{\mathcal{C}}(\boldsymbol{\kappa} - \boldsymbol{\kappa}', \gamma) \tilde{\mathcal{C}}(\boldsymbol{\kappa}', \gamma) \tag{3.12}$$

with

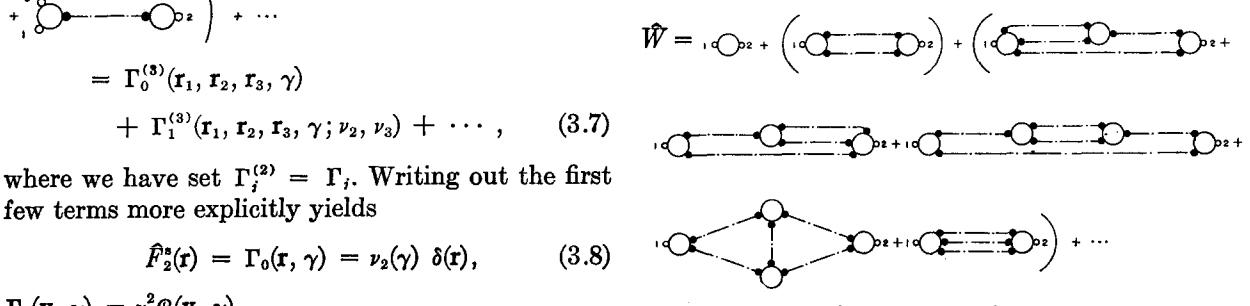
$$\tilde{\mathcal{C}}(\boldsymbol{\kappa}, \gamma) = \tilde{\Phi}(\boldsymbol{\kappa}, \gamma)/[1 - \nu_2 \tilde{\Phi}(\boldsymbol{\kappa}, \gamma)]. \tag{3.13}$$

Also,

$$\Gamma_0^{(3)}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = \nu_3\, \delta(\mathbf{r}_{12})\, \delta(\mathbf{r}_{23}), \tag{3.14}$$

$$\Gamma_1^{(3)}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3; \gamma) = \nu_3\nu_2[\delta(\mathbf{r}_{12})\mathcal{C}(\mathbf{y}_{13}, \gamma) + \delta(\mathbf{r}_{13})\mathcal{C}(\mathbf{y}_{12}, \gamma) + \delta(\mathbf{r}_{23})\mathcal{C}(\mathbf{y}_{13}, \gamma)]. \tag{3.15}$$

The Γ ordering may also be applied directly to the function $\hat{W}$ introduced in Eq. (2.22), where the first few terms become particularly simple since there are no cutting bonds in $\hat{W}$,



$$= \hat{W}_0(\mathbf{r}_{12}, \gamma) + \hat{W}_2(\mathbf{y}_{12}, \gamma; \nu_3) + \hat{W}_3(\mathbf{y}_{12}, \gamma; \nu_3, \nu_4) + \cdots \tag{3.16}$$

with

$$\hat{W}_0(\mathbf{r}) = \nu_2\, \delta(\mathbf{r}), \quad \hat{W}_2(\mathbf{y}, \gamma) = \nu_3^2 \mathcal{C}^2(\mathbf{y}, \gamma), \text{ etc.} \tag{3.17}$$

It is clear from the Γ-ordering scheme mentioned before that $\Gamma_j^{(l)}$ and W_j are of $O(\gamma^{rj})$,

$$\Gamma_j^{(l)} = \gamma^{rj}[\Gamma_{j,0}^{(l)} + \gamma\Gamma_{j,1}^{(l)} + \cdots], \qquad W_j = \gamma^{rj}[W_{j,0} + \gamma W_{j,1} + \cdots]. \tag{3.18}$$

Thus, to different orders in γ,

$$\hat{F}_2(\mathbf{r}) = \nu_2\, \delta(\mathbf{r}) + O(\gamma^r) = [\nu_2^0 + \gamma\nu_{2,1} + \cdots]\, \delta(\mathbf{y}) + O(\gamma^r), \tag{3.19}$$

where $\delta(\mathbf{y}) = 0$ *unless* $\mathbf{r} = \gamma^{-1}\mathbf{y} = 0$, and generally;

$$\begin{aligned}\hat{F}_2(\mathbf{y}) &= \nu_2\,\delta(\mathbf{y}) + \Gamma_1(\mathbf{y},\gamma) + \cdots \\ &\quad + \Gamma_i(\mathbf{y},\gamma) + O(\gamma^{\nu(i+1)}) \\ &= \left(\frac{\gamma}{2\pi}\right)^{\nu}\int_{-\pi/\gamma}^{\pi/\gamma}\cdots\int e^{i\boldsymbol{\kappa}\cdot\mathbf{y}} \\ &\quad \times [\nu_2^{-1} - \tilde{\Phi}(\boldsymbol{\kappa},\gamma)]^{-1}\,d\boldsymbol{\kappa} + O(\gamma^{2\nu}) \\ &\equiv \left(\frac{\gamma}{2\pi}\right)^{\nu}\int_{-\pi/\gamma}^{\pi/\gamma}\cdots\int e^{i\boldsymbol{\kappa}\cdot\mathbf{y}} \\ &\quad \times [\bar{W}_0 - \tilde{\Phi}(\boldsymbol{\kappa},\gamma)]^{-1}\,d\boldsymbol{\kappa} + O(\gamma^{2\nu}). \qquad (3.20)\end{aligned}$$

Alternatively,

$$\begin{aligned}\hat{F}_2(\mathbf{y}) &= \left(\frac{\gamma}{2\pi}\right)^{\nu}\int_{-\pi/\gamma}^{\pi/\gamma}\cdots\int e^{i\boldsymbol{\kappa}\cdot\mathbf{y}} \\ &\times \{[\bar{W}_0 + \bar{W}_2(\boldsymbol{\kappa},\gamma) + \cdots + \bar{W}_i(\boldsymbol{\kappa},\gamma)]^{-1} \\ &- \tilde{\Phi}(\boldsymbol{\kappa},\gamma)\}^{-1}\,d\boldsymbol{\kappa} + O(\gamma^{\nu(i+1)}), \qquad (3.21)\end{aligned}$$

where we have used Eq. (2.23) to express $\hat{F}_2$ in terms of $\hat{W}$. This ordering of $\hat{F}_2$ on the basis of the Γ ordering of $\hat{W}$ turns out to be very convenient later when we consider self-consistent type approximations for $\hat{F}_2$.

A. Expansion of the ν_k

A function that appears often in our results is the chain with ν_2^0 as a vertex-function instead of ν_2; we denote this as $\mathfrak{C}_0(\mathbf{y};\gamma)$ or simply $\mathfrak{C}_0(\mathbf{y})$,

$$\mathfrak{C}_0(\mathbf{y}) = \left(\frac{\gamma}{2\pi}\right)^{\nu}\int_{-\pi/\gamma}^{\pi/\gamma}\cdots\int d\boldsymbol{\kappa}\,\frac{\tilde{\Phi}(\boldsymbol{\kappa},\gamma)e^{-i\boldsymbol{\kappa}\cdot\mathbf{y}}}{1 - \nu_2^0\tilde{\Phi}(\boldsymbol{\kappa},\gamma)}. \qquad (3.22)$$

We also find it useful to introduce the functions

$$\begin{aligned}&\Delta_i = \nu_i - \nu_i^0, \quad S_0(\mathbf{y}) = \mathfrak{C}_0^2(\mathbf{y}), \\ &T_0(\mathbf{y}) = \mathfrak{C}_0^3(\mathbf{y}), \quad Q_0(\mathbf{y}) = \mathfrak{C}_0^4(\mathbf{y}),\end{aligned} \qquad (3.23)$$

and the identities

$$\tilde{\mathfrak{C}}(\mathbf{k};\gamma) = \tilde{\mathfrak{C}}_0(\mathbf{k};\gamma)/[1 - \Delta_2\tilde{\mathfrak{C}}_0(\mathbf{k},\gamma)], \qquad (3.24)$$

$$\begin{aligned}&\ln[1 - \nu_2\tilde{\Phi}] \\ &\quad = \ln[1 - \nu_2^0\tilde{\Phi}] + \ln[1 - \Delta_2\tilde{\mathfrak{C}}_0]. \qquad (3.25)\end{aligned}$$

From (2.30) and (2.31), making use of (3.23) and (3.24), we find

$$\begin{aligned}\nu_2 &= \nu_2^0 - (\nu_2^0)^2\mathfrak{C}_0(0) + [2(\nu_2^0)^3 - (\nu_3^0)^2]S_0(0) \\ &\quad + (\nu_2^0)^4\mathfrak{C}_0(0)\tilde{S}_0(0)\,\frac{-(\nu_2^0)^2(\nu_3^0)^2}{2} \\ &\quad \times\left\{\frac{1}{(2\pi)^{\nu}}\int \tilde{S}_0(\mathbf{k})[\tilde{\mathfrak{C}}_0(\mathbf{k})]^2\,d\mathbf{k}\right\} \\ &\quad - (\nu_2^0)(\nu_3^0)^2\tilde{T}_0(0) + O(\gamma^{3\nu}). \qquad (3.26)\end{aligned}$$

For arbitrary ρ, the next higher order in γ^{ν} already includes a considerable number of terms. For $\rho = \frac{1}{2}$, however, there is appreciable simplification owing to the fact that $\partial^k\nu_l^0/\partial\rho^k = 0$ when $\rho = \frac{1}{2}$ if $k + l$ is odd.

We have for $\rho = \frac{1}{2}$

$$\begin{aligned}\nu_2 &= \frac{1}{4} - \frac{1}{16}\,\mathfrak{C}_0(0) + \frac{1}{256}\,\mathfrak{C}_0(0)\tilde{S}_0(0) + \frac{1}{32}\,S_0(0) \\ &\quad - \frac{17}{768}\,T_0(0) - \frac{1}{768}\,\tilde{Q}_0(0) - \frac{1}{6144} \\ &\quad \times\left\{\frac{1}{(2\pi)^{\nu}}\int [\tilde{\mathfrak{C}}_0(\mathbf{k})]^2\tilde{T}(\mathbf{k})\,d\mathbf{k}\right\} - \frac{1}{4096}\,S_0(0) \\ &\quad \times\left\{\frac{1}{(2\pi)^{\nu}}\int [\tilde{C}_0(\mathbf{k})]^3\,d\mathbf{k}\right\} - \frac{3}{512}\,S_0(0)\tilde{S}_0(0) \\ &\quad - \frac{1}{4096}\,\mathfrak{C}_0(0)[\tilde{S}_0(0)]^3 + O(\gamma^{4\nu}). \qquad (3.27)\end{aligned}$$

The expansion of ν_k for $k > 2$ is similar. In order to find $\ln\Xi$ through $O(\gamma^{3\nu})$ for arbitrary ρ, and $O(\gamma^{4\nu})$ for $\rho = \frac{1}{2}$, we need only

$$\begin{aligned}&\nu_3 = \nu_3^0 + \tfrac{1}{2}(\partial^2\nu_3^0/\partial\rho^2)(\nu_2^0)^2\mathfrak{C}_0(0) + 0(\gamma^{2\nu}),\ \rho \text{ arbitrary}, \\ &\nu_4 = \nu_4^0 + \tfrac{1}{2}(\partial^2\nu_4^0/\partial\rho^2)(\nu_2^0)^2\mathfrak{C}_0(0) + 0(\gamma^{2\nu}),\ \rho \text{ arbitrary},\end{aligned} \qquad (3.28)$$

where

$$\nu_3^0 = \rho(1-\rho)(1-2\rho)$$

and

$$\nu_4^0 = \tfrac{1}{2}\rho(1-\rho)[(2\rho-1)^2 - 1].$$

A general Φ-bond, $(\partial^k\nu_l^0/\partial\rho^k)$-hypervertex expansion of the ν_k can be obtained by repeated use of (2.31) to eliminate the $\hat{F}^{\nu L}$ in (2.30). This expansion, which can easily be re-expressed in terms of $\mathfrak{C}_0$ instead of Φ, gives the general term in (3.27) and, when used with (3.24), (2.16), and (2.29), also yields the full expansion of $\ln\Xi$ in terms of $\mathfrak{C}_0$ and $\partial^l\nu_k^0/\partial\rho^l$. In characterizing the $\mathfrak{C}_0$-bonds $(\partial^l\nu_k^0/\partial\rho^l)$-hypervertex expansion of ν_k and $\ln\Xi$ graphically, the distinction between those lines incident upon a hypervertex that are associated with the index "l" and those lines associated with the "k" is important; in Appendix B of I we used the designations "in" and "out" in considering the "l's" and "k's", and we refer the reader to that paper for the graphological details. These expansions are simple enough to be written out explicitly through one higher order of γ^{ν} beyond (3.26), (3.27), and (3.28) without undue labor. Beyond that, the large number of terms makes explicit enumeration awkward.

B. Expansion of the Free Energy

The ordering of $\hat{F}_2$ combined with the expansion of the ν_k introduces, using (2.26), a corresponding expansion of the Helmholtz free energy. We find in analogy with (I-6.14)

$$\beta a = \beta a_0(\beta, \rho) + \tfrac{1}{2}\frac{1}{(2\pi)^\nu}\int_{-\pi}^{\pi}\cdots\int \ln[1 - \nu_2^0\tilde{\Phi}(\mathbf{k}, \gamma)]\, d\mathbf{k}$$
$$- \tfrac{1}{4}\{(\nu_3^0)^2\tfrac{1}{3}\sum_{\mathbf{y}} T_0(\mathbf{y}) - (\nu_2^0)^2 S_0(0)\} + O(\gamma^{2\nu+1}). \quad (3.29)$$

Brout[19] seems to have been the first to suggest developing a formal program using γ^ν as an ordering parameter to investigate the free energy, and he considered the results of retaining only zero and first-order terms (i.e., of dropping ν_k for $k \geq 3$), and evaluating the ν_2 by means of a spherical model-like approximation that we discuss in detail later. A similar analysis can be made of expansions in terms of H and Φ rather than ρ and Φ, and subsequent to Brout's initial work Horwitz and Callen[18] suggested an approximation obtainable by retaining the zero and first-order terms in such expansions. Equation (3.29) agrees with the result obtained from the prescription of Coopersmith and Brout.[5]

C. Illustrative Examples

1. Exponential Potential in One Dimension

We consider a one-dimensional system where $\varphi(y)$ of (3.1) has the form $\frac{1}{2}\alpha e^{-|y|}$. We then have

$$\tilde{\Phi}(k, \gamma) = -\beta(\alpha/2)\lambda\gamma[\sinh\gamma/(\cosh\gamma - \cos k) - 1], \quad (3.30)$$

where

$$\lambda(\gamma) = \gamma^{-1}(e^\gamma - 1) \underset{\gamma\to 0}{\to} 1 \quad (3.31)$$

making α the integrated strength of w, $\tilde{w}(0) = \alpha$. This gives to second order in γ [see (3.6)]

$$\hat{F}_2(y, \gamma; \nu_2) = \nu_2\,\delta(r) + (\nu_2^0)^2 \mathcal{C}_0(y, \gamma; \nu_2^0) + O(\gamma^2), \quad (3.32)$$

where, from (3.9) and (3.22),

$$(\nu_2^0)^2\mathcal{C}_0(y, \gamma; \nu_2^0) = \nu_2^0\left[\frac{1}{1 - \beta(\alpha/2)\lambda\gamma\nu_2^0} - 1\right]\delta(r)$$
$$- \beta\lambda\frac{\alpha}{2}\gamma\left[\frac{\nu_2^0}{1 - \beta(\alpha/2)\lambda\gamma\nu_2^0}\right]^2\frac{\sinh\gamma}{\sinh(\gamma s)}e^{-s|y|}, \quad (3.33)$$

s being determined by the relation

$$\cosh(\gamma s) = \cosh\gamma + \frac{\beta(\alpha/2)\lambda\gamma\nu_2^0}{1 - \beta(\alpha/2)\lambda\gamma\nu_2^0}\sinh\gamma, \quad (3.34)$$

yielding

$$s = (1 + \beta\alpha\lambda\nu_2^0)^{\frac{1}{2}} + O(\gamma), \quad (3.35)$$

and ν_2 is given in terms of ν_2^0 by means of (2.9), (2.30), (3.9), and (3.23):

$$\nu_2 = \nu_2^0 - (\nu_2^0)^2\mathcal{C}_0(0, \gamma; \nu_2^0) + O(\gamma^2). \quad (3.36)$$

In a way similar to (3.33), we get with (3.34)

$$\nu_2 = \nu_2^0\{1 - \nu_2^0\beta(\alpha/2)\lambda\gamma[1 - s^{-1}]\} + O(\gamma^2). \quad (3.37)$$

After expanding (3.33) to second order in γ, we obtain

$$\Gamma_1(y, \gamma; \nu_2^0) = (\nu_2^0)^2\beta(\alpha/2)\lambda\gamma[\delta(r) - s^{-1}$$
$$\times \exp[-s\,|y|]] + O(\gamma^2), \quad (3.38)$$

and finally,

$$\hat{F}_2(y, \gamma; \nu_2^0) = \nu_2^0\,\delta(r) + (\nu_2^0)^2\beta(\alpha/2)\lambda\gamma$$
$$\times [\delta(r) - \exp[-s\,|y|]]s^{-1} + O(\gamma^2). \quad (3.39)$$

This expression coincides with the one derived by Kac and Helfand[9] [their (5.8)] for $r \neq 0$, and $\rho = \frac{1}{2}$.

For the same one-dimensional potential we obtain, for the free energy per unit volume given by (3.29), the expression

$$\beta a(\beta, \rho, \gamma) = \rho\ln\rho + (1 - \rho)\ln(1 - \rho)$$
$$+ \beta(\alpha/2)\rho^2 - \frac{\gamma}{2}\{1 + \beta(\alpha/2)\lambda\rho - [1 + \beta\alpha\lambda\rho(1 - \rho)]^{\frac{1}{2}}\}$$
$$+ \frac{\gamma^2}{2}\cdot\frac{[\beta(\alpha/2)\lambda\rho(1 - \rho)]^2}{1 + \beta\alpha\lambda\rho(1 - \rho)}\left\{\frac{1}{2} + \frac{1}{9}\beta(\alpha/2)\lambda(1 - 2\rho)^2\right.$$
$$\left.\times\frac{1}{[1 + \beta\gamma\alpha\rho(1 - \rho)]}\right\} + O(\gamma^3). \quad (3.40)$$

This free energy may be compared after expanding λ, (3.31), in powers of γ with the one derived for a continuum in (I-6.15). Both can be written as

$$\beta a(\beta, \rho) = \beta a^0(\beta, \rho) + \beta(\alpha/2)\rho^2 - (\gamma/2)$$
$$\times\{1 + \beta(\alpha/2)\rho - [f(\beta, \rho)]^{\frac{1}{2}}\} + f^{-1}(\beta, \rho)0(\gamma^2), \quad (3.41)$$

where $f(\beta, \rho) = 0$ is the boundary of the meta-stable region in the van der Waal–Maxwell equation of state [cf. Eq. (4.1) and Fig. 1], and follows from (1.15) for the lattice and from (I-1.2) for the continuum system.

Equation (3.39) can also be compared, after the appropriate transcription, with the free energy per lattice site of the spin system, and it coincides at $\rho = \frac{1}{2}$ with the free-field expression derived by Siegert,[6] including the terms of $O(\gamma^2)$.

2. Exponential Potential in Two and Three Dimensions

Following Baker[8] we consider a potential $w_\nu(\mathbf{r}, \gamma)$, $\nu = 2, 3$, as given by (3.1) of the explicit form

$$w_\nu(\mathbf{r}, \gamma) = \alpha\gamma^\nu \frac{\lambda_\nu(\gamma)}{2^\nu} \prod_{i=1}^{\nu} e^{-\gamma|r^i|} \qquad (3.42)$$

for $\mathbf{r} \neq 0$ and vanishes for $\mathbf{r}=0$, where $\mathbf{r}=(r^1, \cdots, r^\nu)$ and

$$\lambda_2(\gamma) = \gamma^{-2}(e^\gamma - 1)^2/e^\gamma \underset{\gamma\to 0}{\to} 1, \qquad (3.43)$$

$$\lambda_3(\gamma) = 8\gamma^{-3}[(e^\gamma - 1)^3/(6e^{2\gamma} + 2)] \underset{\gamma\to 0}{\to} 1. \qquad (3.44)$$

Then

$$\tilde{\Phi}_\nu(\mathbf{k}, \gamma) = -\beta\alpha \frac{\gamma^\nu\lambda_\nu(\gamma)}{2^\nu}\left[(e^{2\gamma} - 1)^\nu - \prod_{i=1}^{\nu}(1 + e^{2\gamma} - 2e^\gamma \cos k_i)\right] \Big/ \prod_{i=1}^{\nu}(1 + e^{2\gamma} - 2e^\gamma \cos k_i), \qquad (3.45)$$

where $\mathbf{k} = (k_1, \cdots, k_\nu)$ and $\tilde{w}_\nu(0) = \alpha$.

For the free energy per unit volume we obtain from (3.29), after expanding in powers of γ, the expression

$$\beta a_\nu(\beta, \rho) = \beta a^0(\beta, \rho) + \tfrac{1}{2}\beta\alpha\rho^2 - \tfrac{1}{2}\gamma^\nu \int_0^{\nu\pi^\nu} d\xi\, I_\nu(\xi) + O(\gamma^{\nu+1}), \qquad (3.46)$$

where the form of the integrand $I_\nu(\xi)$ depends explicitly on the dimension ν.

Two-dimensions ($\nu = 2$). In analogy to the result obtained by Baker[8] for the corresponding $I_2(\xi)$, we have for $\alpha < 0$,

$$\beta a(\beta, \rho) = \beta a^0(\beta, \rho) + \tfrac{1}{2}\beta\alpha\rho^2 - \tfrac{1}{4}\beta\alpha\gamma^2 \times \left[\tfrac{1}{2}\rho(2-\rho) - \frac{1}{\pi}\int_0^{\nu\pi^\nu} d\xi\, \mathbf{K}(\beta\,|\alpha|\,\xi)\right] + O(\gamma^3), \qquad (3.47)$$

where $\mathbf{K}(k)$ is the complete elliptic integral of the first kind which diverges as $k \to 1$, i.e., under an appropriate upper limit in (3.47), when the relation $\beta\,|\alpha|\,\nu_2^0 = 1$ holds. However, the integral in (3.47) is still finite in this case as one can see from the expansion of $\mathbf{K}(k)$ for k near 1. The same expansion allows also to see that the coefficient of γ^3 is already divergent under the same relation. Therefore, we obtain the behavior of the free energy predicted after (1.14). Finally, from the properties of elliptic integrals, it can be seen that for $\beta\,|\alpha|\,\nu_2^0 \to 1$ there is a singularity in the γ^2 term of the specific heat at constant density for $\rho = \frac{1}{2}$, of the form $(T - T_c)^{-1}$, where T_c is the van der Waals–Maxwell critical temperature $\beta_c = -(\nu_2^0\alpha)^{-1} = -4/\alpha$.

Three Dimensions ($\nu = 3$). In this case the expression for the free energy per unit volume differs from (3.47) in that the integral in (3.46) has the form[8]

$$\frac{\gamma^3}{\pi^2}\int_0^{3\pi^3} d\xi \int_0^{\pi} dk_3\, \mathbf{K}\left(\frac{\gamma^2\beta\,|\alpha|\,\xi}{1 + e^{2\gamma} - 2e^\gamma \cos k_3}\right) \Big/ (1 + e^{2\gamma} - 2e^\gamma \cos k_3),$$

giving again a finite result for the coefficient of γ^3 even when $\beta\,|\alpha|\,\nu_2^0 \to 1$.

IV. DISCUSSION OF γ-EXPANSION AND SELF-CONSISTENT APPROXIMATIONS

Before discussing the usefulness of the γ-expansion developed in the last section, we discuss first the known, or conjectured, behavior of our system for different values of γ. In the van der Waals limit $\gamma \to 0$, the free energy per unit volume $a(\beta, \rho, 0+) = \lim_{\gamma\to 0} a(\beta, \rho, \gamma)$ is obtained[7] by applying the double-tangent construction to the generalized van der Waals free energy $a_0(\beta, \rho) = a^0(\beta, \rho) + \frac{1}{2}\alpha\rho^2$. The latter is, of course, the zero-order term in our expansion of $a(\rho, \gamma)$ in powers of γ. In Fig. 1 the exterior of curve I is the region in which $a_0(\beta, \rho) = a(\beta, \rho, 0+)$ for $\alpha < 0$. [For $\alpha \geq 0$, $a(\beta, \rho, 0+)$ coincides with $a_0(\beta, \rho)$ for all β and ρ.] Inside curve I, the system will exist in two phases and have its thermodynamic properties described by $a(\rho, 0+)$, a linear combination of its properties in the two phases; $a_0(\rho)$, on the other hand, describe the properties of the system when the system is in a state

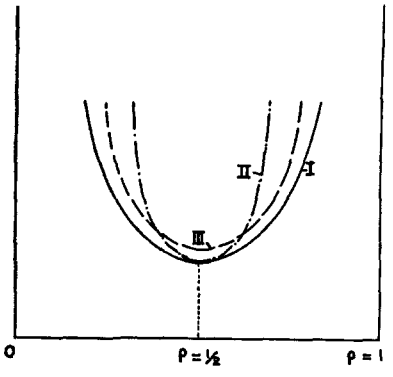


FIG. 1. I. Boundary of two-phase region for $\gamma \to 0$. II. Boundary of metastable region for $\gamma \to 0$. III. "Expected" boundary of two-phase region for small γ in two and three dimensions.

of uniform density. This coincides with the metastable state in the region between curve I and curve II. This latter curve is determined by the equation

$$\beta = -[\nu_2^0 \tilde{w}(0)]^{-1} = -[\rho(1-\rho)\alpha]^{-1}. \quad (4.1)$$

We have also drawn in Fig. 1 the "expected" two-phase region of the system in two and three dimensions for some fixed small γ (keeping the integrated strength of the potential α fixed). (In one dimension there will be no transition for $\gamma \neq 0$.)

Now, the coefficients of the expansion in γ [of $a(\beta, \rho, \gamma)$ or $\hat{F}_2(\mathbf{y}, \gamma)$] will diverge as the boundary of the metastable region, curve II, is approached from the outside and will be meaningless inside curve II, where $dp_0(\rho)/d\rho$ is negative, cf. (3.40). The expansion in γ can therefore be meaningful, as an approximation to the real state of the system, only outside curve I. Inside curve I, the correct $a(\beta, \rho, \gamma)$ is not analytic in γ for small γ. The coefficients of our expansion will become large as the boundary of curve II is approached (this curve coincides with curve I at the critical density $\rho = \frac{1}{2}$).

The first few terms in the expansion may therefore be used as an approximation to the properties of the system for small values of γ only in the region—call it R—outside curve I, and not too close to the critical point [outside a region of $O(\gamma^{\frac{1}{3}})$ in one dimension[9]]. In the region R, the expansion in γ is straightforward (we have not, however, investigated the question of the convergence of the expansion at all).

The above analysis indicates that the straightforward γ-expansion is incapable of yielding reliable information inside the critical (or two-phase) region for finite γ. In order to overcome this limitation on the γ-expansion, as well as to consider situations where γ is not very small, one can use approximation methods for $\hat{F}_l$ which do not assume analyticity in γ; $a(\rho, \gamma)$ may then be obtained from $\hat{F}_2$ via (2.26). Now, our expansion procedure in the last section consisted of two parts: (1) Γ ordering of the graphs

$$\hat{F}_l(\mathbf{r}_1, \cdots, \mathbf{r}_l) = \sum_{k=0}^{\infty} \Gamma_k^l(\mathbf{r}_1, \cdots, \mathbf{r}_l; \gamma, \nu_2, \cdots, \nu_{k+l}) \quad (4.2)$$

and (2) an expansion of each Γ_k^l in powers of γ. It was this second step, which involved the expansion of the ν_k in powers of γ, that led to the singularities discussed before. It seems therefore sensible to avoid the expansion of the Γ_k^l in powers of γ. This requires some method for evaluating the ν_k to each order in the Γ ordering (without any reference to γ at all). We shall now describe such a method but leave its detailed analysis to Part III of this series.

A. Self-Consistent Method for the Determination of the ν_k

Since our interest lies primarily in $\hat{F}_2$, and the thermodynamic properties which may be computed from it, we consider the following iterative self-consistent method for the evaluation of the ν_k appearing in $\hat{F}_2$. [This method is based on Γ ordering but does not make any explicit reference to the range of $w(\mathbf{r})$.] The first-order step in our approximation scheme consists of retaining *only* the Γ_0 and Γ_1 graphs in $\hat{F}_2$ and evaluating the ν_2 appearing in them from the exact relation (2.9),

$$\nu_2^0(\rho) \equiv \rho(1-\rho) = \Gamma_0(\mathbf{r}_1, \mathbf{r}_1; \nu_2) + \Gamma_1(\mathbf{r}_1, \mathbf{r}_1; \nu_2) = \Omega^{-1} \sum_{\mathbf{k}} \frac{\nu_2}{1 - \nu_2 \tilde{\Phi}(\mathbf{k})}. \quad (4.3)$$

In the next order of our approximation we retain the Γ_0, Γ_1, and Γ_2 graphs in $\hat{F}_2$ and the Γ_0^3 and Γ_1^3 graphs in $\hat{F}_3$. We then determine the ν_2 and ν_3 which appear in these graphs from the set of two coupled equations taken from (2.9),

$$\nu_2^0(\rho) = \nu_2 + \Gamma_1(\mathbf{r}_1, \mathbf{r}_1; \nu_2) + \Gamma_2(\mathbf{r}_1, \mathbf{r}_1; \nu_2, \nu_3), \quad (4.4)$$

$$\nu_3^0(\rho) = \nu_3 + \Gamma_1^3(\mathbf{r}_1, \mathbf{r}_1, \mathbf{r}_1; \nu_2, \nu_3). \quad (4.5)$$

In general, the kth order in our scheme consist in evaluating the Γ_i^l for $j + l = 2, \cdots, k+1$, as functions of the ν_i, $i = 2, \cdots, k+1$ and then solving k-equations of the form

$$\nu_j^0(\rho) = f_j(\nu_2, \cdots, \nu_{k+1}), \quad j = 2, \cdots, k+1. \quad (4.6)$$

The self-consistency in our scheme refers to the fact that in each order we have $\hat{F}_2(\mathbf{r}_1, \mathbf{r}_1)$ equal to its exact value at the given density ρ. In other words $n_2(\mathbf{r}_1, \mathbf{r}_1) = 0$ in each order, i.e., the average pair density vanishes when the positions of the two particles coincide. In spin language this means that $\langle[\sigma(\mathbf{r}_1)]^2\rangle = \frac{1}{4}$ in all orders. The discussion following Eq. (2.10) shows that the symmetry properties of $\hat{F}_k$ and ν_k as a function of $\rho - \frac{1}{2} = \langle\sigma\rangle$ remain valid in all orders of our iterative scheme.

A slight modification of this method is to use the auxiliary function $\hat{W}$ (from which $\hat{F}_2$ may then be determined), Eq. (2.23), in the above iterative scheme. According to (3.16), the lth term in the Γ ordering of $\hat{W}$, $\hat{W}_l$ will only contain ν_k with $k \leq l+1$ for $l > 0$, $\hat{W}_0 = \nu_2 \delta(\mathbf{r})$. Thus, in the lowest order, we equate $\hat{W}$ to $\hat{W}_0$ and obtain the corresponding $\hat{F}_2$, (3.20), and again determine ν_2 from (2.9). This leads again to (4.2). In the next order

we retain $\hat{W}_0$ and $\hat{W}_2$ in $\hat{W}$, compute the resulting $\hat{F}_2$ from (3.21), and determine ν_2 and ν_3 from (4.5) and the equation

$$\nu_2^0(\rho) = \Omega^{-1} \sum_{\mathbf{k}} \{[\nu_2 + \nu_3^2 \tilde{S}(\mathbf{k}; \nu_2)]^{-1} - \tilde{\Phi}(\mathbf{k})\}^{-1} \quad (4.7)$$

[$\tilde{S}$ defined in (3.12)], which replaces Eq. (4.4). This process may be continued, yielding at each step equations similar to (4.6). The set of graphs contained in $\hat{F}_2$ obtained from $\hat{W}$ in the nth order is larger than the set $\Gamma_0 + \cdots + \Gamma_n$ for $n > 1$. This appears to have advantages for systems where $w(\mathbf{r})$ is not very long range, to which these methods are also applicable.

The results of this iteration scheme will be analyzed in Part III of this series, where it is shown, in particular, that our lowest approximation for the pair-distribution function is identical with that obtained from the mean spherical model of Lewis and Wannier[10,21] for spin systems which coincide in the thermodynamic limit $\Omega \to \infty$ with the spherical model of Berlin and Kac.[11] This is also similar to the result of Brout.[19]

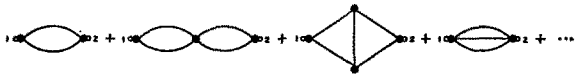
B. Bond Renormalization

We can go a step further in the use of $\hat{W}$ by considering its L-bond expansion instead of its $\mathcal{C}$-bond expansion. For a lattice system the definition of L [Eq. (I-2.19)] is[22]

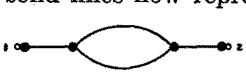
$$L(\mathbf{r}_{12}) = \Phi(\mathbf{r}_{12}) + \sum_{\mathbf{r}_3, \mathbf{r}_4} \Phi(\mathbf{r}_{13}) \hat{F}_2(\mathbf{r}_{34}) \Phi(\mathbf{r}_{24}). \quad (4.8)$$

In terms of graphs with L-bonds rather than $\mathcal{C}$-bonds, the prescription for $\hat{W}$ following (2.23) includes the added restriction that no graph should contain any *articulation pairs of bonds*, i.e., pairs of bonds which, when cut, cause the graph to separate into two or more parts, one of which contains at least one hypervertex, but no labeled hypervertex. Thus, we have

$\hat{W}_2(\mathbf{r}_{12}) = \nu_2 +$



(4.9)

(the solid lines now representing L-bonds). Graphs like  do not appear.

[21] J. L. Lebowitz and J. K. Percus, Phys. Rev. 144, 251 (1966).

[22] Our L coincides, for lattice gases, with the "renormalized interaction" v of R. Abe, Ref. 13.

The introduction of the L-bonds may be thought of as a bond renormalization similar in some ways to the introduction of the ν_k's in place of the μ_k's, which is a vertex renormalization. One reason for introducing the bond renormalization is that it provides a convenient means of analyzing certain aspects of the critical behavior of $\hat{F}_2(\mathbf{r})$ as well as suggesting a class of approximations that appear to exhibit the kind—although not the precise extent—of deviation from the Ornstein–Zernike[15] theory that is actually found in the behavior of the two-dimensional lattice gas and in real three-dimensional systems. Such approximations can be obtained by applying the same recipe that defines the Γ ordering scheme, described in Sec. III, to the L-bond graphs of $\hat{W}$. However, we do not pursue the investigation of any such particular approximation here, but instead restrict our remarks to observations that have a more immediate bearing on the behavior of $\hat{F}_2$ near the critical point.

The L-bond expansion seems better suited to examining critical behavior of $\hat{F}_2$ than the expansions we have previously considered, because one already has the defining relation (2.23) for $\hat{W}$ in terms of $\hat{F}_2$ and any sufficiently simple second relationship between $\hat{W}$ and L—and hence between $\hat{W}$ and $\hat{F}_2$ through Eq. (4.8)—immediately provides a convenient means of determining the spatially asymptotic behavior of $\hat{F}_2$ at the critical point where we assume[13]

$$1 - \tilde{\Phi}(0)\tilde{W}(0) = 0. \quad (4.10)$$

We start by assuming that when (4.10) is satisfied, $\hat{F}_2$ behaves like some inverse power of r for large r, and we use the kind of analysis initiated by Green[23] and generalized by Stillinger and Frisch,[24] Fisher,[25] and Stillinger.[26] We find, then, that (2.23), (4.10), and the assumptions that

$$\hat{W}(\mathbf{r}) \sim A[L(\mathbf{r})]^m, \quad \text{for} \quad r \to \infty \quad (4.11)$$

and

$$\hat{F}_2(\mathbf{r}) \sim B/r^n, \quad \text{for} \quad r \to \infty \quad (4.12)$$

[23] M. S. Green, J. Chem. Phys. 33, 1403 (1960).

[24] F. H. Stillinger, Jr., and H. L. Frisch, Physica 27, 751 (1961).

[25] M. E. Fisher, J. Math. Phys. 5, 944 (1964).

[26] F. H. Stillinger, Jr. (private communication). Stillinger has made a detailed examination of consequences of assuming more general forms than (4.11), including such possibilities as an addition $(\ln r)^t$ term in the denominator and the replacement of the constant by some reasonable angular dependent quantity. He has concluded that such modifications will not change the relationship among constants such as our n, ν, and m.

imply that[27]

$$n = 2\nu/(1 + m), \tag{4.13}$$

where we have also assumed that $\Phi(\mathbf{r})$ is short-ranged enough so that, for large r, $L(\mathbf{r})$ as well as $\hat{F}_2(\mathbf{r})$ is given by (4.12). Here, ν is the dimensionality of the space. Instead of (4.11) and (4.12), we now consider the more general possibility [imposed upon us by the form of (4.17)]:

$$\hat{W}(\mathbf{r}) \sim A[L(\mathbf{r})]^m f_a(\mathbf{r}), \quad \text{for} \quad r \to \infty \tag{4.14}$$

and

$$\hat{F}_2(\mathbf{r}) \sim Bf_b(\mathbf{r})/r^n, \quad \text{for} \quad r \to \infty, \tag{4.15}$$

where f_a and f_b are functions of order less than any positive power and greater than any negative power of r. (When the letter f appears hereafter in this section it always denotes such a function.)

The same arguments that are used to obtain (4.13) from (4.11) and (4.12) suggest that (4.14) and (4.15) also yield (4.13), supplemented by a relation between f_a and f_b. Although a general demonstration for arbitrary f_a or f_b has not been given, and is not attempted here, these remarks can be made more precise[26] for an important class of particular f_a and f_b. For example, in the simple but representative case of $f_a = (\log r)^l$ and $f_b = (\log r)^{-\epsilon}$, we obtain (4.13) and the relation

$$l = \epsilon(1 + m). \tag{4.16}$$

Furthermore, as Green[23] and Abe[13] have pointed out, simple dimensional arguments can be applied to graphs of the sort that appear in the L-bond expansion of $\hat{W}(\mathbf{r})$. These arguments indicate that any graph in the L-bond expansion of $\hat{W}(\mathbf{r})$ consisting of b-bonds and k-hypervertices will behave like

$$r^{-[bn-\nu k+2\nu]} f_c(r) = f_a(r)[L(\mathbf{r})]^m, \quad \text{for} \quad r \to \infty, \tag{4.17}$$

where $L(\mathbf{r})$ is assumed to go asymptotically as $f(r)/r^n$. Such dimensional considerations are somewhat crude but they probably give a reasonable picture of the relative dominance of the various graphs at large r, and we now consider their implications. Equation (4.17) yields

$$\hat{W}(\mathbf{r}) \sim \sum_i A_i r^{-[b_i n-\nu k_i+2\nu]} f_i(\mathbf{r}), \quad \text{for} \quad r \to \infty, \tag{4.18}$$

where the sum is over all graphs in $\hat{W}$, and A_i, b_i, k_i, and f_i are associated with the ith graph. At the critical point, $\rho = \frac{1}{2}$ and any graph containing a ν_k with odd k vanishes [thus, among the graphs actually drawn in (4.9) only the last one remains]. Among the graphs that are left, the ones whose hypervertices are all ν_4's are the ones with the fewest bonds for a given number of hypervertices ($b+1=2k$). We might expect these to be the dominant graphs for large r, and according to (4.18) this expectation is fulfilled as long as

$$\sum_{\{b_i+1=2k_i\}} A_i f_i \neq 0. \tag{4.19}$$

Equations (4.13), (4.17), and (4.19) then yield

$$n = \tfrac{1}{2}\nu. \tag{4.20}$$

This is the case considered by Abe.[13] More generally, it follows that

$$n = \nu k_i/(b_i + 1) \tag{4.21}$$

for the pairs $\{k_i, b_i\}$ that correspond to the set of graphs having the longest range (i.e., the single lowest value of $b_i n - \nu k_i + 2\nu$) such that the sum $\sum A_i f_i$ over these pairs is not zero, provided that the series (4.18) is a valid and convergent representation of $\hat{W}(\mathbf{r})$ at the critical point. This latter stipulation, of course, involves not only the validity of (4.17) but also the validity of the L-bond expansion of $\hat{W}$ at the critical point in the first place.

In the case of a nearest-neighbor interaction on square and cubic lattices, the exact $\hat{F}_2$ at the critical point appears to have the form (4.15) with $n = \frac{1}{4}$ when $\nu = 2$, and $n \simeq \frac{1}{16}$ when $\nu = 3$.[28] In order to be in agreement with these figures, we must have $m = 15$ for $\nu = 2$ and $m \simeq \frac{79}{17}$ for $\nu = 3$ ($m = \frac{14}{3}$ would yield $n = \frac{1}{8}$). In light of (4.21), this suggests that either there is wholesale cancellation among graphs or else the series (4.18) does not provide a valid representation of $\hat{W}$. We further note that the cancellation that would enable us to ignore certain subsums of graphs for $n < \frac{1}{2}\nu$ will necessarily involve cancellation of graphs that are individually divergent [in Eq. (4.18), $b_i n - \nu k_i + 2\nu$ will always be <0 for some b_i and k_i when $n < \frac{1}{2}\nu$]. This means that, strictly speaking, cancellation is not an alternative to the breakdown of the graphical representation (4.9) but rather a special case of this breakdown.

Percus and one of the authors[29] (G.S.) have considered a weakened version of the Ornstein–Zernike theory[14] that does not rest upon the convergence of (4.9). It indicates that the m in Eq. (4.14) may

[27] The argument leading to (4.13) from (2.23), (4.11), (4.12), and the assumption that $1 - \tilde{\Phi}(0)\,\hat{W}(0) = 0$ is identical to the one used by Fisher (Ref. 25) in discussing the hypernetted-chain equation, and we refer the interested reader to that reference for details.

[28] D. S. Gaunt, M. E. Fisher, M. F. Sykes, and J. W. Essam, Phys. Rev. Letters 13, 713 (1964).

[29] J. K. Percus and G. Stell (to be published).

be closely connected with the shape of the critical isotherm in the vicinity of the critical point, and further progress in the direction of associating the m and A in Eq. (4.14) with macroscopic features of the lattice system seems likely. The task of obtaining reliable estimates of these quantities directly in terms of Φ and ρ appears much more difficult, however.

The above considerations are not directly applicable to the one-dimensional system with a Kac potential. However, for this case an explicit computation shows that near the critical point [see Eq. (5.9) of Ref. 9] $\mathfrak{C} \sim \Lambda^{\frac{1}{2}} e^{-\Lambda r}$ where $\Lambda = s\gamma \sim \gamma^{4/3}$ so that we can use Λ instead of γ as an ordering parameter in $\mathfrak{C}$-bond expansions. It can easily be seen that, in the expansion of F_2^L, the graphs whose labeled hypervertices are ν_2's and whose unlabeled hypervertices are ν_4's are all of order $\Lambda^{\frac{1}{2}}$, and that all other graphs are of higher order in Λ. Hence, the sum of these graphs of order $\Lambda^{\frac{1}{2}}$ will yield the

$$\sum_{s \geq 1} a_s \gamma^{\frac{1}{2}} \exp\left[-b_s \gamma^{4/3} r\right],$$

which Kac and Helfand have shown to be the dominant term in $F_2^L(\mathbf{r})$ when the critical point is approached. Similarly, in the $\mathfrak{C}$-bond expansion of $\hat{W}^L$, the graphs whose vertices are all ν_4's are the dominant ones in the critical region. Hence, in the L-bond expansion of $\hat{W}^L$, the graphs whose vertices are all ν_4's are also the dominant ones in the critical region, since, upon expansion of L in terms of $\mathfrak{C}$, all of the $\mathfrak{C}$-bond graphs with ν_4 vertices come only from the L-bond graphs with ν_4 vertices.

Kac[30] has conjectured that, in such graphical representations of F_2 and related functions, the terms that are dominant in the critical region in the one-dimensional case may be the ones that dominate in all dimensions, despite the fact that the degree and even the *kind* of singularities present can be expected to be different in different dimensions. If (4.19) were satisfied, the graph with ν_4 vertices would have a special role in all dimensions and the result would be consistent with Kac's conjecture. However, the confrontation with the known value of n for $\nu = 2$ forces us to either abandon our graphical representation altogether at the critical point or at least conclude that a subtle kind of cancellation among graphs must be occurring so that (4.19) is violated.

ACKNOWLEDGMENT

We would like to thank G. Horwitz for valuable discussions.

[30] Talk presented at Semi-annual Statistical Mechanics Meeting at Belfer Graduate School of Science, Yeshiva University (1964).

A Theorem on the Clebsch–Gordan Series in $A(l)$, $B(l)$, $C(l)$, and $D(l)$

F. ZACCARIA
Istituto di Fisica Teorica dell 'Università, Napoli, Italy
(Received 26 October 1965)

By a general method, conditions are derived for $A(l)$, $B(l)$, $C(l)$, and $D(l)$, in order that terms and multiplicities of the Clebsch–Gordan series for the decomposition of the direct product of two irreducible representations be dependent on only one of them.

I.

THE Clebsch–Gordan (CG) series describes the decomposition of the direct product, always completely reducible, of two irreducible representations of a semi-simple Lie group into a direct sum of irreducible representations. For a group of rank l, we write

$$(p_1 \cdots p_l) \otimes (p_1' \cdots p_l') = \sum\nolimits_{\oplus} m_{pp'}(p_1'' \cdots p_l'')(p_1'' \cdots p_l''), \quad (1)$$

where $(p_1 \cdots p_l)$ denotes the irreducible representation associated to the highest weight

$$M = \sum_1^l{}_i\, p_i M^{(i)}, \quad (2)$$

with p's nonnegative integers and $M^{(i)}$ the l fundamental dominant weights.

The case in which the number of terms of the sum in (1) and their multiplicities m are determined by the set $(p_1' \cdots p_l')$ alone shall be called "the special case".

In $SU(2)$ the CG series is

$$(p_1) \otimes (p_1') = \sum_{-p_1'}^{+p_1'} {}_{\Delta_p} (p_1 + \Delta p)$$

with

$$p_1 \geq p_1',$$

and we are evidently always in the special case.

By using Young diagram techniques, a general analysis of the direct product in $SU(3)$ has been done by Preziosi, Simoni, and Vitale,[1] and the conditions for the special case are

$$p_i \geq p_1' + p_2' \qquad (i = 1, 2).$$

Another, more fruitful, technique[2] starts from the general consideration that a necessary condition, call it (a), for a vector to be the highest weight [corresponding to an irreducible representation contained in the sum of (1)] is that this vector is obtained by adding to M a weight m' of the weight diagram $WD(M')$ having M' as its highest weight. This condition is, however, not sufficient.

The positive Weyl chamber W^+ is the region of the weight space where all the highest weights lie. If $WD(M')$ is translated so that its center is moved from the origin of the weight space to M, and the whole diagram falls inside W^+, it can be shown that all the vectors which are sum of M and of any m' are the highest weights of the irreducible representations of the right-hand side of (1), i.e., the condition (a) is also sufficient. Furthermore, each multiplicity in (1) is the same as the multiplicity of the corresponding added weight m'. In other words, multiplicities and terms of (1) are determined exclusively by the set $(p_1' \cdots p_l')$ and this is the "special case".

For groups of rank 2, $SU(3)$, $C(2)$, $G(2)$, the weight space is bi-dimensional and it is possible to derive geometrically the needed conditions. This has been done by Vitale.[3] What follows is the additional result:

$$C(2): \quad p_1 \geq p_1' + 2p_2', \qquad p_2 \geq p_1' + p_2';$$

$$G(2): \quad p_1 \geq 2p_1' + 3p_2', \qquad p_2 \geq p_1' + 2p_2'.$$

Conditions for $B(2)$ are the same as for $C(2)$. Labeling the fundamental dominant weights as in Sec. III, they become

$$B(2): \quad p_1 \geq p_1' + p_2', \qquad p_2 \geq 2p_1' + p_2'.$$

The external contour of $WD(M')$ is determined

[1] B. Preziosi, A. Simoni, and B. Vitale, Nuovo Cimento **34**, 1101 (1964).

[2] See, for instance, D. Speiser, Istanbul Lecture Notes (1962). For the general structure of semi-simple Lie groups, see G. Racah, *Group Theory and Spectroscopy* (Institute for Advanced Study, Princeton, New Jersey, 1951).

[3] B. Vitale, "On the Structure of the Clebsch–Gordan Series for Semisimple Lie Groups" (National Science Foundation, Summer Institute for Theoretical Physics, University of Wisconsin, preprint, 1965).

by operating on M' with the Weyl reflections S^k, which form the Weyl group. The relation leading to the special case conditions is then

$$M + S^k M' = X^{(k)}, \qquad X^{(k)} \in W^+, \tag{3}$$

i.e., the sum of M and of the weight obtained by applying S^k on M' for any k must be a vector $X^{(k)}$ belonging to W^+.

Detailed calculations to satisfy (3) for each k have been made, yielding the following results:

$$\begin{aligned}
\mathrm{SU}(4)^3\text{:}\quad & p_i \geq p_1' + p_2' + p_3' && (i = 1, 2, 3);\\
B(3)\text{:}\quad & p_i \geq p_1' + 2p_2' + p_3' && (i = 1, 2),\\
& p_3 \geq 2p_1' + 2p_2' + p_3';\\
C(3)\text{:}\quad & p_i \geq p_1' + 2p_2' + 2p_3' && (i = 1, 2),\\
& p_3 \geq p_1' + p_2' + p_3'.
\end{aligned}$$

Due to the rapid increase of the number of the Weyl reflections with the rank, this method soon becomes impractical. General methods are therefore sought to find the special case conditions for an arbitrary l.

Nussinov[4] has recently used Young diagrams and tableaus for $A(l)[SU(l+1)]$, finding the conditions

$$p_i \geq p_1' + p_2' + \cdots + p_i' \qquad (i = 1, 2, \cdots, l).$$

In the following section, a simple method is described. It is in turn applied to $A(l)$, $B(l)$, $C(l)$, and $D(l)$ in Sec. III.

II.

The n-dimensional weight space S_n is referred to a system of orthogonal axes. The components $M_i^{(i)}$ of the fundamental dominant weights, and then of any highest weight, are known. The Weyl group essentially permutes the components of a weight among themselves, with prescriptions about change of sign different for each group.

For $B(l)$, $C(l)$, and $D(l)$, $n = l$. For $A(l)$, $n = l + 1$, but the weight components m_i satisfy the condition

$$\sum_1^{l+1} {}_i\, m_i = 0. \tag{4}$$

From (2) we have

$$M_j = M_j(p_1, \cdots, p_l) \qquad (j = 1, \cdots, n), \tag{5}$$

where the p_i can be interpreted as component of M along $M^{(i)}$, the set $(M^{(1)} \cdots M^{(l)})$ forming a basis for S_l in the case of B, C, D groups and for the l-dimensional subspace of S_{l+1} in which all weights lie, in the case of $A(l)$.

Due to (4), (5) is in any case a system of l linear independent equations in the l p_i's, and for our groups it is possible to solve it by simple inspection. One obtains

$$p_i = p_i(M_1, \cdots, M_l) \qquad (i = 1, \cdots, l). \tag{6}$$

Then (3) yields

$$M_j(p) + [S^k M'(p')]_j = X_j^{(k)}(x_1^{(k)}, \cdots, x_l^{(k)}) \qquad (j = 1, \cdots, n)$$

or

$$\begin{aligned}
X_j^{(k)}(x^{(k)}) - M_j(p) &= X_j^{(k)\prime}(x^{(k)} - p)\\
&= [S^k M'(p')]_j \quad (j = 1, \cdots, n),
\end{aligned}$$

in which $x_i^{(k)}$ is the component of $X^{(k)}$ along $M^{(i)}$.

By making use of (6), we have

$$x_i^{(k)} - p_i = f_i[S^k M'(p')] \qquad (i = 1, \cdots, l). \tag{7}$$

The conditions on $x_i^{(k)}$'s for $X^{(k)}$ to be the highest weight are that they must be nonnegative and integers. That the latter condition is satisfied can be seen from the actual solutions of (7). The first requirement leads to the inequalities

$$p_i \geq -f_i[S^k M'(p')] \qquad (i = 1, \cdots, l). \tag{8}$$

For each i we choose that S^k which makes the right-hand side of (8) as great as possible. This choice is always possible due to the general structure of the Weyl group and corresponds to finding for each i the suitable permutation (and, in case, the change of sign) to give (8) its maximum value.

The final conditions are

$$p_i \geq [-f_i]_{\max} = F_i(p_1', \cdots, p_l') \quad (i = 1, \cdots, l). \tag{9}$$

In Sec. III the explicit expressions of (9) are derived for each of the four groups under consideration.

If the $(p_1' \cdots p_l')$ of the special case is called "representation determining with respect to $(p_1 \cdots p_l)$", we can state:

Theorem: Equation (9) constitutes a set of necessary and sufficient conditions for $(p_1' \cdots p_l')$ to be the representation determining with respect to $(p_1 \cdots p_l)$.

III.

1. $A(l)$

The weight space has $l + 1$ dimensions and the following are the fundamental dominant weights:

[4] S. Nussinov, "A Theorem on the Clebsch-Gordan Series in $SU(n)$" (University of Washington, preprint).

$$M^{(1)}: \ \frac{l}{l+1}, -\frac{1}{l+1}, \cdots, -\frac{1}{l+1},$$

$$M^{(2)}: \ \frac{l-1}{l+1}, \frac{l-1}{l+1}, -\frac{2}{l+1}, \cdots, -\frac{2}{l+1},$$

$$\vdots$$

$$M^{(l)}: \ \frac{1}{l+1}, \frac{1}{l+1}, \cdots, \frac{1}{l+1}, -\frac{l}{l+1}.$$

The components of a highest weight M are

$$M_1 = \frac{1}{l+1}(p_1 l + p_2(l-1) + \cdots + p_{l-1}2 + p_l),$$

$$M_2 = \frac{1}{l+1}(-p_1 + p_2(l-1) + \cdots + p_{l-1}2 + p_l),$$

$$M_3 = \frac{1}{l+1}(-p_1 - 2p_2 + p_3(l-2) + \cdots + p_{l-1}2 + p_l),$$

$$\vdots$$

$$M_{l+1} = \frac{1}{l+1}(-p_1 - 2p_2 - 3p_3 - \cdots - (l-1)p_{l-1} - lp_l).$$

The explicit expressions of (6) are

$$p_i = M_i - M_{i+1} \qquad (i = 1, \cdots, l-1),$$
$$p_l = M_1 + \cdots + M_{l-1} + 2M_l, \tag{10}$$

and those of (8) are

$$p_i \geq (S^k M')_{i+1} - (S^k M')_i \qquad (i = 1, \cdots, l-1),$$
$$p_l \geq -[(S^k M')_1 + \cdots + (S^k M')_{l-1} + 2(S^k M')_l]. \tag{11}$$

The Weyl group permutes all the components of a weight among themselves without any change of sign. The maximum value of (11) is obtained when, for each i, S^k is chosen so that

$$(S^k M')_{i+1} = M'_1 \quad \text{and} \quad (S^k M')_i = M'_{i+1}$$

and for l

$$(S^k M')_1 = M'_2,$$
$$\vdots$$
$$(S^k M')_l = M'_{l+1}.$$

The special case conditions are then

$$y_i \geq p'_1 + p'_2 + \cdots + p'_l \qquad (i = 1, \cdots, l) \tag{12}$$

as already found in Ref. 4.

2. $B(l)$

The weight space has l dimensions and the following are the fundamental dominant weights:

$$M^{(1)} : \ 1, 0, 0, \cdots 0, 0,$$
$$M^{(2)} : \ 1, 1, 0, \cdots 0, 0,$$
$$\vdots$$
$$M^{(l-1)}: \ 1, 1, 1, \cdots 1, 0,$$
$$M^{(l)} : \ \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \cdots \tfrac{1}{2}, \tfrac{1}{2}.$$

The components of a highest weight M are:

$$M_1 = p_1 + p_2 + \cdots + p_{l-1} + \tfrac{1}{2}p_l,$$
$$M_2 = \quad p_2 + \cdots + p_{l-1} + \tfrac{1}{2}p_l,$$
$$\vdots$$
$$M_l = \quad \tfrac{1}{2}p_l.$$

The explicit expressions of (6) are

$$p_i = M_i - M_{i+1} \qquad (i = 1, \cdots, l-1),$$
$$p_l = 2M_l, \tag{13}$$

and those of (8) are

$$p_i \geq (S^k M')_{i+1} - (S^k M')_i \quad (i = 1, \cdots, l-1),$$
$$p_l \geq -2(S^k M')_l. \tag{14}$$

The Weyl group permutes all the components of a weight among themselves with any number of changes of sign. The maximum values of the right-hand side of (14) is obtained when, for each i, S^k is chosen so that

$$(S^k M')_{i+1} = M'_1 \quad \text{and} \quad (S^k M')_i = -M'_2$$

and for l

$$(S^k M')_l = -M'_1.$$

The special case conditions are then

$$p_i \geq p'_1 + 2p'_2 + \cdots + 2p'_{l-1} + p'_l \qquad (i = 1, \cdots, l-1),$$
$$p_l \geq 2p'_1 + \cdots + 2p'_{l-1} + p'_l. \tag{15}$$

3. $C(l)$

The weight space has l dimensions and the following are the fundamental dominant weights:

$$M^{(1)}: \ 1, 0, 0, \cdots 0,$$
$$M^{(2)}: \ 1, 1, 0, \cdots 0,$$
$$\vdots$$
$$M^{(l)}: \ 1, 1, 1, \cdots 1.$$

The components of a highest weight M are

$$\begin{aligned} M_1 &= p_1 + p_2 + \cdots + p_l, \\ M_2 &= \quad p_2 + \cdots + p_l, \\ &\vdots \\ M_l &= \quad p_l. \end{aligned}$$

The explicit expressions of (6) are

$$p_i = M_i - M_{i+1} \quad (i = 1, \cdots, l-1), \qquad (16)$$
$$p_l = M_l,$$

and those of (8) are

$$p_i \geq (S^k M')_{i+1} - (S^k M')_i \quad (i = 1, \cdots, l-1), \qquad (17)$$
$$p_l \geq -(S^k M')_l.$$

The Weyl group is the same as in $B(l)$. Also the choice of the S^k for each i and for l is the same as in $B(l)$.

The special case conditions are

$$p_i \geq p'_1 + 2p'_2 + \cdots + 2p'_l \quad (i = 1, \cdots, l-1), \qquad (18)$$
$$p_l \geq p'_1 + \cdots + p'_l.$$

4. $D(l)$

The weight space has l dimensions and the following are the fundamental dominant weights:

$$\begin{aligned} M^{(1)}&: \quad 1, 0, \cdots, 0, 0, \\ M^{(2)}&: \quad 1, 1, 0, \cdots, 0, 0, \\ &\vdots \\ M^{(l-2)}&: \quad 1, 1, 1, \cdots, 1, 0, 0, \\ M^{(l-1)}&: \quad \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \cdots, \tfrac{1}{2}, \tfrac{1}{2}, \\ M^{(l)}&: \quad \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \cdots, \tfrac{1}{2}, -\tfrac{1}{2}. \end{aligned}$$

The components of a highest weight M are

$$\begin{aligned} M_1 &= p_1 + p_2 + \cdots + p_{l-2} + \tfrac{1}{2}p_{l-1} + \tfrac{1}{2}p_l, \\ M_2 &= \quad p_2 + \cdots + p_{l-2} + \tfrac{1}{2}p_{l-1} + \tfrac{1}{2}p_l, \\ &\vdots \\ M_{l-2} &= \quad p_{l-2} + \tfrac{1}{2}p_{l-1} + \tfrac{1}{2}p_l, \\ M_{l-1} &= \quad \tfrac{1}{2}p_{l-1} + \tfrac{1}{2}p_l, \\ M_l &= \quad \tfrac{1}{2}p_{l-1} - \tfrac{1}{2}p_l. \end{aligned}$$

The explicit expressions of (6) are

$$\begin{aligned} p_i &= M_i - M_{i+1} \quad (i = 1, \cdots, l-2), \\ p_{l-1} &= M_{l-1} + M_l, \\ p_l &= M_{l-1} - M_l, \end{aligned} \qquad (19)$$

and those of (8) are

$$p_i \geq (S^k M')_{i+1} - (S^k M')_i \quad (i = 1, \cdots, l-2),$$
$$p_{l-1} \geq (S^k M')_{l-1} + (S^k M')_l, \qquad (20)$$
$$p_l \geq (S^k M')_l - (S^k M')_{l-1}.$$

The Weyl group is the group of permutations of the components of a weight with an even number of changes of sign. The maximum value of the right-hand side of (20) is obtained when, for each i, S^k is chosen so that

$$(S^k M')_{i+1} = M'_1 \quad \text{and} \quad (S^k M')_i = -M'_2,$$

for $l-1$

$$(S^k M')_{l-1} = M'_1 \quad \text{and} \quad (S^k M')_l = M'_2,$$

and for l

$$(S^k M')_l = M'_1 \quad \text{and} \quad (S^k M')_{l-1} = -M'_2.$$

The special case conditions are

$$p_i \geq p'_1 + 2p'_2 + \cdots + 2p'_{l-2} + p'_{l-1} + p'_l \quad (i = 1, \cdots, l). \qquad (21)$$

IV.

This general method can obviously be applied also to the exceptional groups, once their fundamental dominant weights are known.

Its great simplicity allows hopes for other successful applications. In particular, it is expected to lead to more detailed knowledge on the structure of the CG series in the general case.

Note added in proof: The special case conditions have also been obtained for such groups. This will be given in a forthcoming communication.

ACKNOWLEDGMENTS

The author wishes to thank Professor B. Vitale for his continuous guidance and suggestions. He would also like to thank Dr. B. Gruber for useful discussions.

Probability and Entropy of Macroscopic Fluctuations

ROBERT ZWANZIG*
National Bureau of Standards, Washington, D. C. 20234
(Received 17 June 1965)

A macroscopic fluctuation is a fluctuation of order N in an N-particle system. This article contains a calculation of the probability of a macroscopic fluctuation, and its associated entropy, which is asymptotically correct in the limit of large N. Sufficiently small macroscopic fluctuations are shown to obey the same Gaussian distribution law as spontaneous microscopic fluctuations (of order $N^{\frac{1}{2}}$). The modifications necessary to describe large macroscopic fluctuations are found. The entropy of a macroscopic nonequilibrium state is expressed by means of various moments calculated at equilibrium. The nonlinear thermodynamic force for a nonequilibruim state far from equilibrium is found. The calculation is based on a cumulant expansion of the characteristic function of the probability distribution and a stationary phase estimate of its Fourier transform.

THE statistical mechanical theory of fluctuations is usually developed with small (spontaneous) fluctuations in mind. By "small" is meant fluctuations that are macroscopically unobservable, typically fluctuations of order $N^{\frac{1}{2}}$ in an N-particle system.

Spontaneous macroscopic fluctuations (of order N in an N-particle system) occur with such small probability when N is large that they are never observed experimentally; and for this reason one might suppose that they are of no interest.

But even though they might not occur spontaneously, macroscopic deviations from equilibrium can be *imposed* on a system, for example, by removing a constraint. Macroscopic nonequilibrium states of this kind are of the greatest importance in the study of irreversible processes. Because we are particularly interested in knowing the entropy of such states and because of the familiar relation of probability to entropy, the theory of the probability of macroscopic fluctuations forms an important part of the general statistical mechanical theory of irreversible processes.

Small fluctuations are commonly treated by either of two distinct methods. The first involves use of the central limit theorem of probability theory. This method, developed elegantly by Khinchin,[1] is mathematically rigorous. Unfortunately, however, its useful application is limited to small fluctuations. When fluctuations are of order N, it leads to results that, while rigorous, are so weak as to be useless. Therefore we do not discuss this method any further.

The other common method, found in most textbooks, is based on Boltzmann's principle. The probability $P(a)$ of a state specified by the macroscopic variables[2] a is related to the entropy $S(a)$ of that state by

$$P(a) \sim \exp\,[S(a)/k_B]. \tag{1}$$

At equilibrium, the variables a take on the values $\bar{a}$. Near equilibrium, the entropy may be expanded in a power series,

$$S(a) = S(\bar{a}) - \tfrac{1}{2}\sum_{ij} g_{ij}(a_i - \bar{a}_i)(a_j - \bar{a}_j) + \cdots, \tag{2}$$

where

$$g_{ij} = -(\partial^2 S/\partial a_i\,\partial a_j)_{\text{equil}}. \tag{3}$$

Because the equilibrium state has maximum entropy, the matrix $\mathbf{g}$ is positive definite. Then, for small enough deviations from equilibrium, the probability $P(a)$ is Gaussian in deviations,

$$P(a) \sim \exp\left[-\frac{1}{2k_B}\sum_{ij} g_{ij}(a_i - \bar{a}_i) \times (a_j - \bar{a}_j) + \cdots\right]. \tag{4}$$

The coefficients g_{ij} are found by thermodynamic calculation, or also, as is well known, by calculating second moments.

There is nothing in this method that obviously restricts the validity of the results to fluctuations of order $N^{\frac{1}{2}}$; and in fact the results appear to be correct for macroscopic fluctuations—provided that

* Present address: Institute for Fluid Dynamics and Applied Mathematics, University of Maryland, College Park, Maryland.

[1] A. I. Khinchin, *Mathematical Foundations of Statistical Mechanics* (Dover Publications, Inc., New York, 1949).

[2] For simplicity of notation, we use the single variable a to represent the set of all variables of interest.

they are *small* macroscopic fluctuations, of the form

$$a - \bar{a} \sim \alpha N, \tag{5}$$

where α is small but constant in the limit of large N. To see that this is so, and to extend the theory to large macroscopic fluctuations (where α is large but constant in the limit of large N), another approach seems desirable.

This article is concerned mainly with asymptotic calculation of $P(a)$ for large macroscopic fluctuations, using a kind of "stationary phase" method.

We consider a system in thermal equilibrium, described by specified numerical values[3] a_c of certain phase functions A_c. The appropriate microcanonical ensemble distribution function is

$$f_0(X/a_c) = \delta[A_c(X) - a_c]/W(a_c), \tag{6}$$

where the normalizing denominator is the structure function,

$$W(a_c) = \int dX\ \delta[A_c(X) - a_c]. \tag{7}$$

It must be remembered that $\delta(A_c - a_c)$ is an abbreviation for the product of delta functions in each member of the set denoted by A_c.

The probability that another phase function A_v will have the numerical value a_v is determined by the volume of the intersection of two surfaces in phase space, one surface specified by the equation $A_c(X) = a_c$, and the other by $A_v(X) = a_v$. Thus we have the following expression for the probability (density) of a_v, conditional on specified values of a_c,

$$\text{Prob}\ \{A_v(X) = a_v, \text{ given } a_c\} \equiv P(a_v/a_c)$$
$$= \int dX\ \delta[A_v(X) - a_v] f_0(X/a_c). \tag{8}$$

By the definition of the distribution function $f_0(X/a_c)$ given in (6), this is

$$P(a_v/a_c) = \frac{\int dX\ \delta[A_v(X) - a_v]\ \delta[A_c(X) - a_c]}{W(a_c)}, \tag{9}$$

or, with the obvious generalization of (7),

$$W(a_v, a_c) = \int dX\ \delta(A_v - a_v)\ \delta(A_c - a_c), \tag{10}$$

the probability is

$$P(a_v/a_c) = W(a_v, a_c)/W(a_c). \tag{11}$$

This shows clearly the intimate relation between probabilities and structure functions.

The form given in (8) turns out to be most convenient for our purposes. Let us introduce the simplifying notation

$$\langle B \rangle \equiv \int dX\ B(X) f_0(X/a_c). \tag{12}$$

In this notation, the probability is

$$P(a_v/a_c) = \langle \delta(A_v - a_v) \rangle. \tag{13}$$

Suppose that the set A_v contains n phase functions, $a_1, a_2, \cdots, a_n$. Then the n-fold delta function in (13) may be replaced by an n-fold integral,

$$P(a_v/a_c) = \frac{1}{(2\pi)^n} \int d\xi\ e^{-i\xi \cdot a_v} \langle e^{i\xi \cdot A_v} \rangle. \tag{14}$$

We use ξ as an abbreviation for the set $\xi_1, \xi_2, \cdots, \xi_n$.

Equation (14) shows that the characteristic function of the distribution $f_0(X/a_c)$ is of interest. We make a short detour now, to discuss *the cumulant expansion of a characteristic function.*[4]

The characteristic function has a power series expansion,

$$\langle e^{i\xi \cdot A_v} \rangle = 1 + \sum i\xi_i \langle A_i \rangle + \tfrac{1}{2} \sum i\xi_i i\xi_k \langle A_i A_k \rangle + \tfrac{1}{6} \sum i\xi_i i\xi_k i\xi_l \langle A_i A_k A_l \rangle + \cdots. \tag{15}$$

The logarithm of the characteristic function, denoted here by $F(\xi)$,

$$F(\xi) = \log \langle e^{i\xi \cdot A_v} \rangle \tag{16}$$

also has a power series expansion,

$$F(\xi) = \sum i\xi_i \langle A_i \rangle_c + \tfrac{1}{2} \sum i\xi_i i\xi_k \langle A_i A_k \rangle_c + \tfrac{1}{6} \sum i\xi_i i\xi_k i\xi_l \langle A_i A_k A_l \rangle_c + \cdots. \tag{17}$$

The coefficients in this series are called the *cumulants* of the distribution. (The notation $\langle\ \rangle_c$ appears to be due to Kubo.) Cumulants can be expressed in terms of moments by taking the logarithm of the power series in (15), and then comparing the result with (17). The first three cumulants are

$$\begin{aligned} \langle A_i \rangle_c &= \langle A_i \rangle; \\ \langle A_i A_k \rangle_c &= \langle A_i A_k \rangle - \langle A_i \rangle \langle A_k \rangle; \\ \langle A_i A_k A_l \rangle_c &= \langle A_i A_k A_l \rangle - \langle A_i A_k \rangle \langle A_l \rangle \\ &\quad - \langle A_i A_l \rangle \langle A_k \rangle - \langle A_k A_l \rangle \langle A_i \rangle + 2 \langle A_i \rangle \langle A_k \rangle \langle A_l \rangle. \end{aligned} \tag{18}$$

[3] Lower case letters denote numerical values, and capitals denote functions of points in phase space. The subscripts c and v, appear in the following discussion, are used only to distinguish two sets of variables; but we have in mind that c refers to dynamical constants of motion, of the sort that determine the equilibrium state, while v refers to dynamical variables, of the sort that change systematically with time during the approach to equilibrium. The position of a point in phase space is X, and $\int dX$ means integration over all of phase space.

[4] A convenient summary of cumulant methods has been given by R. Kubo, J. Phys. Soc. Japan 17, 1100 (1962).

The general formula relating cumulants to moments is known (see, for example, Ref. 4); but we do not need it here.

An important property of cumulants is that they are extensive quantities. By this we mean the following. If the system under consideration contains N particles in a volume V, and if the phase functions $A_c(X)$ and $A_s(X)$ have the usual many-body structure, then for large systems all cumulants are of order N. Then the quantity $F(\xi)$ depends on N and V according to

$$F(\xi) \to N\phi(\xi, N/V), \tag{19}$$

where ϕ is a function independent of the size of the system. Although no general proof of this assertion appears to be available, it can be verified by direct calculation in typical cases.[5]

Now we return to (14), which becomes

$$P(a_s/a_c) = \frac{1}{(2\pi)^n} \int d\xi\, e^{-i\xi\cdot a_s} e^{N\phi(\xi)}. \tag{20}$$

In a macroscopic fluctuation, not only the mean values $\bar{a}_i$ but also the deviations $a_i - \bar{a}_i$ from the mean values are of order N. This suggests the substitution

$$a_i = \alpha_i N. \tag{21}$$

Then for macroscopic fluctuations both α and ϕ remain constant as N becomes large.

With this notation, (20) becomes

$$P(a_s/a_c) = \frac{1}{(2\pi)^n} \int d\xi\, e^{N[\phi(\xi) - i\xi\cdot\alpha]}. \tag{22}$$

The structure of the integrand suggests use of the method of stationary phase.[6]

The gist of the method of stationary phase is as follows. When N is large, and ξ varies over the domain of integration, the imaginary part of the exponent varies with extreme rapidity. The integrand is a complex number; its phase is the imaginary part of the exponent. On integration over ξ, the rapid variation of phase gives rise to a substantial cancellation. There is, however, one exception. When the phase is stationary with respect to small changes in ξ, then little cancellation occurs. Thus, the largest contribution to the integral comes from the region where the phase is stationary. In this method, the exact exponent is replaced by the first two terms in its power series expansion about the point of stationary phase.

This point, denoted by ξ_0, is the solution of the variational condition

$$\frac{d}{d\xi_j}[\phi(\xi) - i\sum \xi_k\alpha_k] = 0; \qquad (j = 1, 2, \cdots, n). \tag{23}$$

The expansion of the exponent around this point is

$$\phi(\xi) - i\xi\cdot\alpha = \phi(\xi_0) - i\xi_0\cdot\alpha + \tfrac{1}{2}\sum \left(\frac{\partial^2\phi}{\partial\xi_j\,\partial\xi_k}\right)_0 (\xi_j - \xi_{0j})(\xi_k - \xi_{0k}) + \cdots. \tag{24}$$

For the method to work, the quadratic form in (24) must be negative definite; and we see that this is the case. For convenience, the second derivatives are abbreviated by

$$(\partial^2\phi/\partial\xi_j\,\partial\xi_k)_0 = \phi_{jk}(\xi_0). \tag{25}$$

On substituting the quadratic approximation (24) into the integral (22), one obtains the n-dimensional Fourier transform of a Gaussian function. This can be evaluated easily; the result is

$$(a_s/a_c) = \frac{1}{(2\pi)^n}\left(\frac{2\pi}{N}\right)^{\frac{1}{2}n} \{\det[-\phi_{jk}(\xi_0)]\}^{-\frac{1}{2}} \times \exp N[\phi(\xi_0) - i\xi_0\cdot\alpha]. \tag{26}$$

In our original notation, this is

$$P(a_s/a_c) = [(2\pi N)^n \det\{-\phi_{jk}\}]^{-\frac{1}{2}} \times \exp[F(\xi_0) - i\xi_0\cdot a_s]. \tag{27}$$

We repeat that this result is asymptotically correct as N approaches infinity.

To conclude the calculation, we must find the point of stationary phase as a function $\xi_0(a_s)$ of the macroscopic variables a_s. This can be done by means of the series expansion (17), together with the formulas (18) for the cumulants. On multiplication by N, (23) becomes

$$(d/d\xi_j)F(\xi) = ia_j; \qquad (j = 1, 2, \cdots, n) \tag{28}$$

or

$$\langle A_j\rangle_c + \sum_k \langle A_jA_k\rangle_c i\xi_k + \tfrac{1}{2}\sum_{kl}\langle A_jA_kA_l\rangle_c i\xi_k i\xi_l + \cdots = a_j. \tag{29}$$

[5] The present discussion is based on the asymptotic N dependence given in Eq. (19), and all results are valid only in an asymptotic sense for large N. It might seem at first that the usual theory of spontaneous fluctuations, of asymptotic order $N^{\frac{1}{2}}$, is lost by taking only the leading term, of order N, in (19). This is not so, however. The asymptotic order of spontaneous fluctuations comes in fact from the square root of the asymptotic order of the dominant term in (19).

[6] For an excellent survey and discussion of this method, see A. Erdelyi, *Asymptotic Expansions* (Dover Publications, Inc., New York, 1956).

(We have taken advantage of the symmetry of the cumulants to permutation of their subscripts.) The solution of (29) is found easily by iteration.

The cumulant $\langle A_i \rangle_c$ is identical with the average of A_i, as in (18). This is the equilibrium average, determined by the constants a_c; it is denoted by $\bar{a}_i$,

$$\langle A_i \rangle_c = \langle A_i \rangle = \bar{a}_i. \tag{30}$$

In the course of the calculation we need the inverse of the matrix $\langle A_i A_k \rangle_c$. Here we use the standard notation

$$\langle A_i A_k \rangle_c = k_B (\mathbf{g}^{-1})_{ik}, \tag{31}$$

which defines the matrix $\mathbf{g}$. As a result of these changes in notation, (29) becomes

$$k_B \sum_k (\mathbf{g}^{-1})_{ik} i\xi_k + \tfrac{1}{2} \sum_{kl} \langle A_i A_k A_l \rangle_c i\xi_k i\xi_l + \cdots = a_i - \bar{a}_i. \tag{32}$$

The solution of this equation, carried out explicitly to second order in deviations from equilibrium, is

$$i\xi_{0i} = \frac{1}{k_B} \sum_k g_{ik}(a_k - \bar{a}_k) - \frac{1}{2k_B^3} \sum_{kluvw} g_{iu} g_{kv} g_{lw} \times \langle A_u A_v A_w \rangle_c (a_k - \bar{a}_k)(a_l - \bar{a}_l) + \cdots. \tag{33}$$

The corresponding value of the exponent in (27) is

$$F(\xi_0) - i\xi_0 \cdot a_v = -\frac{1}{2k_B} \sum_{jk} g_{jk}(a_j - \bar{a}_j)(a_k - \bar{a}_k) + \frac{1}{6k_B^3} \sum_{jkluvw} g_{iu} g_{kv} g_{lw} \langle A_u A_v A_w \rangle_c \times (a_j - \bar{a}_j)(a_k - \bar{a}_k)(a_l - \bar{a}_l). \tag{34}$$

The quadratic term is in exact agreement with our earlier (4). The cubic term represents deviations from Gaussian behavior. Evidently, the method given here can be used to generate still higher-order deviations.

It is instructive to estimate the order of magnitude of the various terms appearing in (34). Because of the extensive character of the cumulants,

$$\langle A_i A_k \rangle_c \sim N, \qquad \langle A_i A_k A_l \rangle_c \sim N; \tag{35}$$

and because of (31), we find that

$$F(\xi_0) - i\xi_0 \cdot a_v \sim O\left(\frac{1}{N}\right)(a - \bar{a})^2 + O\left(\frac{1}{N^2}\right)(a - \bar{a})^3 + \cdots. \tag{36}$$

For spontaneous fluctuations from equilibrium, as is well known, the deviations are of order $N^{\frac{1}{2}}$,

$$a - \bar{a} \sim N^{\frac{1}{2}}. \tag{37}$$

Then only the quadratic term in (36) is important. But when the deviation from equilibrium is macroscopic,

$$a - \bar{a} \sim N, \tag{38}$$

then all terms in (36) must be kept. However, when the deviation is macroscopic *but still small*, the cubic term can be neglected relative to the quadratic term. This is how one sees that the Gaussian approximation is valid for small macroscopic fluctuations.

In the course of the derivation, a certain matrix

$$\phi_{ik}(\xi_0) = (\partial^2 \phi / \partial \xi_i \, \partial \xi_k)_0 \tag{39}$$

was assumed to be negative definite. Because of our explicit formula for the root ξ_0, we can now calculate this matrix more explicitly:

$$\phi_{ik}(\xi_0) = -\frac{1}{N} \left\{ \langle A_i A_k \rangle_c + \frac{1}{k_B} \sum_{lm} \langle A_i A_k A_l \rangle_c g_{lm}(a_m - \bar{a}_m) + \cdots \right\}. \tag{40}$$

In the limit of small macroscopic fluctuations, when the third-order cumulant can be neglected, this is clearly a negative definite matrix. [For example, one might have picked the variables $A(X)$ so as to be orthogonal at equilibrium. Then the matrix $\langle A_i A_k \rangle_c$ would be diagonal, and its diagonal elements would be positive.] There will surely be a range of deviations $a_m - \bar{a}_m$ for which the negative definite character is maintained. Whether this range can be extended by considering higher-order deviations is not clear; further discussion appears to depend on specific calculations.

The entropy of the microcanonical ensemble specified by constraints a_c and a_v is given by

$$S(a_v, a_c) = k_B \log W(a_v, a_c). \tag{41}$$

From (11), (27), and (34), we obtain an expression for the extra entropy associated with the constraints a_v,

$$S(a_v, a_c) = S(a_c) - \tfrac{1}{2} k_B \log [(2\pi N)^n \det (-\phi_{ik})] - \tfrac{1}{2} \sum_{jk} g_{jk}(a_j - \bar{a}_j)(a_k - \bar{a}_k) + \tfrac{1}{6} \sum_{jkluvw} g_{iu} g_{kv} g_{lw} \times \langle A_u A_v A_w \rangle_c (a_j - \bar{a}_j)(a_k - \bar{a}_k)(a_l - \bar{a}_l) + \cdots. \tag{42}$$

In this way the calculation of the entropy of a nonequilibrium system is reduced to the calculation of various moments or cumulants at equilibrium.

The thermodynamic force driving macroscopic irreversible processes is defined as[7]

$$F_i = \partial S/\partial a_i . \tag{43}$$

[7] Linear irreversible processes are discussed by S. R. DeGroot and P. Mazur, *Non-Equilibruim Thermodynamics* (North-Holland Publishing Company, Amsterdam, 1962). The generalization to nonlinear irreversible thermodynamics is given by R. Zwanzig, Phys. Rev. **124**, 983 (1961).

From (42) we obtain a generalization of the familiar linear approximation to the next order in macroscopic deviations from equilibrium,

$$F_i = -\sum_k g_{ik}(a_k - \bar{a}_k) + \tfrac{1}{2}\sum_{klmw} g_{iu}g_{kv}g_{lw} \times \langle A_u A_v A_w\rangle_c (a_k - \bar{a}_k)(a_l - \bar{a}_l) + \cdots . \tag{44}$$

The preceding calculations can evidently be carried out to arbitrarily large order by an obvious extension of the method set forth.