JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 5, NUMBER 1 JANUARY 1964

Von Neumann Algebras of Local Observables for Free Scalar Field*

HUZIHIRO ARAKI†
Department of Physics, University of Illinois, Urbana, Illinois

Basic properties of von Neumann algebras of local observables for the free scalar field are investigated. In particular, the duality theorem for regions at a fixed time is proved. It is also shown that von Neumann algebras for regions at a fixed time are factors of type II_∞ or III_∞. An argument which excludes the possibility of a factor of type I for a general class of theory and for a certain class of regions is given.

I. INTRODUCTION

RECENTLY, some attention has been given to von Neumann algebras of local observables in quantum field theory. In order to give a basis for a general discussion, we analyze in this paper the von Neumann algebras of local observables in the quantum theory of the free scalar field, using the results of our previous paper.[1]

With each open set B in the space time, we associate the von Neumann algebra $R(B)$, generated by the spectral projections of smeared out field operators

$$A(h) = \int A(x)h(x)\, d^4x,$$

where the support (the closure of the nonvanishing region) of the test function $h(x)$ is restricted to inside B. This algebra $R(B)$ is interpreted as the von Neumann algebra, generated by those observables, which can be measured in the space time region B.

We will prove the following fundamental properties of $R(B)$ for the free scalar field.

Theorem 1: Let B and B_α be all open.

(1) (Field property)

$$\text{If } B = \bigcup_\alpha B_\alpha, \quad R(B) = \vee_\alpha R(B_\alpha). \tag{1.2}$$

In particular,

$$\text{if } B_1 \supset B_2, \quad R(B_1) \supset R(B_2). \tag{1.3}$$

(2) (Local commutativity)

$$\text{If } B_1' \supset B_2, \quad R(B_1)' \supset R(B_2). \tag{1.4}$$

(3) (Covariance)

$$U(a, \Lambda)R(B)U(a, \Lambda)^{-1} = R((a, \Lambda)B). \tag{1.5}$$

(4) (Completeness) $R(M) = B(\mathfrak{H})$. (1.6)

Here $\vee_\alpha R_\alpha$ is the von Neumann algebra generated by the union of R_α (see Ref. 1, Sec. 3), B' is the causal adjoint of B defined by

$$B' = \{x; (x - y)^2 < 0 \text{ for } y \in B\}, \tag{1.7}$$

where the signature of the metric is $1, -1, -1, -1$, $U(a, \Lambda)$ is the unitary representation of the inhomogeneous Lorentz group in the free scalar field theory, $(a, \Lambda)B$ is the transformed region,

$$(a, \Lambda)B = \{\Lambda x + a; x \in B\}. \tag{1.8}$$

M is the whole-space time and $B(\mathfrak{H})$ denotes the set of all bounded operators on the Hilbert space $\mathfrak{H}$ under consideration.

The proof of this theorem is rather trivial. We explicitly write it down because the properties of $R(B)$ in Theorem 1 may be used as a set of axioms, from which we can deduce many of the results of the conventional axiomatic field theory.[2]

* Supported by the National Science Foundation.

† On leave from Department of Nuclear Engineering, Kyoto University, Kyoto, Japan.

[1] H. Araki, J. Math. Phys. 4, 1343 (1963).

[2] The consequences of the axioms (1.2)–(1.6) will be treated elsewhere.

Some more interesting axioms for $R(B)$, i.e., the time-slice axiom and the duality theorem, have been discussed by Haag and Schroer[3] and, in connection with the duality theorem, the type of the von Neumann algebra $R(B)$ has been of some interest. The time-slice axiom is trivially true for the free scalar field. However, we will show for the free scalar field that the duality theorem, in its most general form, is not true, though it is true for a restricted class of regions B, including the ones considered by Haag and Schroer. The type of von Neumann algebra $R(B)$ is proved to be not of type I for a certain general class of regions B, contrary to the expectation of Haag and Schroer. The conclusion about the duality theorem is probably model sensitive but, as for the type question, we present a general argument which excludes the possibility of a factor of type I for a certain class of regions B and for any $R(B)$ satisfying (1)–(4) of Theorem 1.

To describe our result on the duality theorem more precisely, we define the von Neumann algebra for a wider class of regions. For any $\Delta \subset M$, we define $\bar{R}(\Delta)$ and $\hat{R}(\Delta)$ by

$$\bar{R}(\Delta) = \bigwedge_{B \supset \bar{\Delta}} R(B), \tag{1.9}$$

$$\hat{R}(\Delta) = \bigvee_{\bar{B} \subset \text{int}\Delta} R(B), \tag{1.10}$$

where $\bar{\Delta}$ and int Δ are the closure and the interior of Δ, B is open, and $\wedge$ denotes the intersection. The difference between $\bar{R}(\Delta)$ and $\hat{R}(\Delta)$ is connected with the problem at the boundary of Δ. As we shall see, we have

$$\bar{R}(B) \supset R(B) = \hat{R}(B), \tag{1.11}$$

$$\hat{R}(\Delta_1) \supset \bar{R}(\Delta_2), \quad \text{if } \text{int}\Delta_1 \supset \bar{\Delta}_2. \tag{1.12}$$

Let S_t be the hyperplane $x^0 = t$ in M and let $\mathbf{B}$ be a relatively open subset of S_t, whose relative boundary $\partial\mathbf{B}$ in S_t is a many-times continuously differentiable surface (of two dimensions) or several such surfaces, joined together along many-times continuously differential curves. Let $C(\mathbf{B})$ be the (open) double light cone[4] with the base $\mathbf{B}$. We will prove

Theorem 2:

$$\bar{R}(\mathbf{B}) = \bar{R}(C(\mathbf{B})) = \hat{R}(C(\mathbf{B})) = R(C(\mathbf{B})), \tag{1.13}$$

$$\bar{R}(S_t) = B(\mathfrak{H}). \tag{1.14}$$

The first equality of (1.13) is the Einstein causality and (1.14) is the primitive causality (the time-slice axiom). The rest of (1.13) says that there is no subtle effect of the boundary for the von Neumann algebra, which also shows up in the following theorem.

Theorem 3:

$$\bar{R}(\partial\mathbf{B}) = \{\lambda\mathbf{1}\}. \tag{1.15}$$

The relatively open subsets $\mathbf{B}$ in S_t, satisfying $\text{int}\,\bar{\mathbf{B}} = \mathbf{B}$ and bounded by piecewise many-times differentiable surfaces, form a lattice according to the definition

$$\mathbf{B}_1 \wedge \mathbf{B}_2 = \mathbf{B}_1 \cap \mathbf{B}_2, \tag{1.16}$$

$$\mathbf{B}_1 \vee \mathbf{B}_2 = \text{int}\,\overline{\mathbf{B}_1 \cup \mathbf{B}_2}, \tag{1.17}$$

and $\mathbf{B}_1 \to \bar{\mathbf{B}}_1^c$ (the complementary region of $\bar{\mathbf{B}}_1$ in S_t) is a complementation of this lattice. We will prove

Theorem 4:

$$\bar{R}(\mathbf{B}_1) \wedge \bar{R}(\mathbf{B}_2) = \bar{R}(\mathbf{B}_1 \wedge \mathbf{B}_2), \tag{1.18}$$

$$\bar{R}(\mathbf{B}_1) \vee \bar{R}(\mathbf{B}_2) = \bar{R}(\mathbf{B}_1 \vee \mathbf{B}_2), \tag{1.19}$$

$$\bar{R}(\mathbf{B})' = \bar{R}(\bar{\mathbf{B}}^c). \tag{1.20}$$

The Eq. (1.20) is the duality equation. Equation (1.19) is sharper than Eq. (1.2).

Concerning the type of von Neumann algebras, we immediately see from Theorem 6 of Ref. 1, that $R(B)$ for the free scalar field is of infinite type.[5] Furthermore, we will prove

Theorem 5: $\bar{R}(\mathbf{B})$ is not of type I.

We also give a theorem in the last section concerning the type of $R(B)$, which can be applied not only to the free scalar field but also to the general case.

The above theorems will be proved by applying the results of Ref. 1. In Ref. 1, we have described the normal representation of canonical commutation relations in three different ways by using a complex Hilbert space L or a real Hilbert space K or another real Hilbert space $\hat{L} = H(= K \oplus \beta K)$ and the operator β. In the present case, L is the single-particle space of a free scalar field, which will be described together with $\hat{L}$ in Sec. 2. The formalism using $\hat{L}$ is suited for the discussion of $R(B)$ for a four-dimensional region B, which will be explained in Lemma 1 and Definition 1 of Sec. 3. The rest of Sec. 3 is devoted to the proof of Theorem 1. Since Theorem 1 is concerned with simpler properties of $R(B)$ and not with either the duality theorem or the type question, we prove it directly without using the deeper results of Ref. 1.

For the consideration of $\bar{R}(\mathbf{B})$ for a fixed time, the formalism with K is suited, and K and βK for

[3] R. Haag and B. Schroer, J. Math. Phys. **3**, 248 (1962).

[4] $C(\mathbf{B}) = \{x; (x-y, x-y) < 0 \text{ for any } y \in S_t,\ y \notin \mathbf{B}\}$ where $\mathbf{B} \subset S_t$.

[5] For a general class of local field theory satisfying standard axioms, R. V. Kadison, J. Math. Phys. **4**, 1511 (1963) has shown that $R(B)$ is always of infinite type.

present case are $\mathfrak{F}_\varphi$ and $\mathfrak{F}_\pi$ described in Sec. 4. von Neumann algebras treated in Ref. 1 are specified by two subspaces K_1 and K_2 of K. K_1 and βK_2 for the present case are $F_\varphi(\mathbf{B})$ and $F_\pi(\mathbf{B})$ described in Sec. 5. The actual verification that the von Neumann algebras of Theorem 2, indeed, correspond to these subspaces $F_\varphi(\mathbf{B})$ and $F_\pi(\mathbf{B})$, together with the proof of Theorem 2, will be given in Sec. 6 after we have studied properties of $F_\varphi(\mathbf{B})$ and $F_\pi(\mathbf{B})$ in Sec. 5. In Ref. 1, we have already reduced various properties of von Neumann algebras to the properties of the subspaces K_1 and K_2. Therefore, the problems of duality and type as well as other properties are immediately reduced to certain properties of the subspaces $F_\varphi(\mathbf{B})$ and $F_\pi(\mathbf{B})$, which will be studied in Secs. 5 and 8.

The arguments of Secs. 7 and 9 are independent of the result of Ref. 1. In Sec. 7, we give a counter example to a general duality theorem. In Sec. 9, we give a theorem which excludes the possibility of a factor of type I for a certain class of regions in any theory satisfying standard axioms.

The Appendix is devoted to a proof of some lemmas.

2. THE FREE SCALAR FIELD

Let $\mathcal{S}_4$ be the set of infinitely continuously differentiable real functions of a Minkowski vector with strong decrease and $\{L(a, \Lambda)\}$ be the representation of the inhomogeneous Lorentz group, given by

$$(L(a, \Lambda)h)(x) = h(\Lambda^{-1}(x - a)), \qquad h \in \mathcal{S}_4. \tag{2.1}$$

For the Fourier transform

$$\tilde{h}(p) = (2\pi)^{-\frac{3}{2}} \int h(x) e^{i(p,x)}\, dx, \tag{2.2}$$

where $(p, x) = p^0x^0 - \mathbf{p}\cdot\mathbf{x}$, we have

$$(L(a, \Lambda)\tilde{h})(p) = e^{i(p,a)}\tilde{h}(\Lambda^{-1}p). \tag{2.3}$$

We now introduce a complex-valued inner product

$$\begin{aligned}(h_1, h_2)_L &= \int \tilde{h}_1(p)^*\tilde{h}_2(p)(d^3p/p^0) \\ &= 2i \int h(x)\Delta^{(+)}(x - y)h(y)\, dx\, dy,\end{aligned} \tag{2.4}$$

where p^0 is set equal to $(p^2 + m^2)^{\frac{1}{2}}$ in the first integration and

$$\begin{aligned}\Delta^{(+)}(x) &= -i(2\pi)^{-3} \\ &\times \int e^{-i(p,x)}\, \delta((p, p) - m^2)\theta(p^0)\, d^4p\end{aligned} \tag{2.5}$$

is a well-known distribution. (f, g) is real linear in f and g and $(f, f) \geq 0$. Since (2.4) is determined by the value of $\tilde{h}(p)$ on the hyperboloid $(p, p) = m^2$, $p^0 > 0$, we consider equivalence classes $\mathcal{S}_4/\mathcal{S}_4^0$ where $\mathcal{S}_4^0$ is the set of h with $(h, h) = 0$. On $\mathcal{S}_4/\mathcal{S}_4^0$, we define[6] the multiplication of i by $h' = ih$ if $\tilde{h}'(p) = i\tilde{h}(p)$, on the hyperboloid $(p, p) = m^2$, $p^0 > 0$. With this definition of i, (h_1, h_2) is a complex inner product in $\mathcal{S}_4/\mathcal{S}_4^0$. We complete it and denote the Hilbert space so obtained by L.

As is well known, (2.1) induces an irreducible unitary representation of the inhomogeneous Lorentz group, which we shall denote by the same $L(a, \Lambda)$.

We now use extensively the notations and results of Ref. 1. Since Re $\Delta^{(+)} = \frac{1}{2}\Delta$ and Im $\Delta^{(+)} = \frac{1}{2}\Delta_1$, the inner product in the real Hilbert space $\hat{L}$ and the matrix element of β in $\hat{L}$ are given by

$$\begin{aligned}(h_1, h_2)_{\hat{L}} &= -\int h_1(x)\Delta_1(x - y)h_2(y)\, dx\, dy \\ &= +\int \tilde{h}_1(p)^*\tilde{h}_2(p)\, \delta((p, p) - m^2)\, d^4p.\end{aligned} \tag{2.6}$$

$$(h_1, \beta h_2)_{\hat{L}} = -\int h_1(x)\Delta(x - y)h_2(y)\, dx\, dy. \tag{2.7}$$

The operators $T(L(a, \Lambda)) = U(a, \Lambda)$ give a unitary representation of the inhomogeneous Lorentz group on $\mathfrak{H}_S(\hat{L})$ and the vector $\Omega = \Psi_S(\hat{L})$ is the unique vacuum of the theory. The free scalar field $A(x)$ is defined by

$$A(h) = \int A(x)h(x)\, dx = \chi(h) \mid D_A, \tag{2.8}$$

where h in $\chi(h)$ is to be considered as an element in $\hat{L}$ and $\chi(h) \mid D_A$ is the restriction of $\chi(h)$ to the dense domain D_A, consisting of finite linear combinations of $A(h_1) \cdots A(h_n)\Omega$.

The Hilbert space $\mathfrak{H} = \mathfrak{H}_S(\hat{L})$, the unitary representation $U(a, \Lambda)$, the vacuum Ω and the field $A(x)$ satisfy Wightman's axioms.[7] In addition, we have $A(f) = 0$ for $f \in (\Box + m^2)\mathcal{S}$ (the Klein–Gordon equation) and $[A(h_1), A(h_2)] = -i(h_1, \beta h_2)_{\hat{L}}$ (the commutation relation). Due to (7.12) of Ref. 1, the closure of $A(h)$ is the self-adjoint operator $\chi(h)$, i.e., $A(h)$ is essentially self-adjoint.

3. VON NEUMANN ALGEBRAS OF LOCAL OBSERVABLES FOR THE FREE SCALAR FIELD

Lemma 1: Let H_1 be any linear subset of $\hat{L}$. Each of the following three conditions defines the same von Neumann algebra R.

[6] For example, let $\eta(\lambda)$ be an infinitely continuously differentiable odd real function of a real variable λ, being 1 for $\lambda \geq m$ and -1 for $\lambda \leq m$. If $f \in \mathcal{S}$, then g defined by $\tilde{g}(p) = i\tilde{f}(p)\eta(p^0)$ is also in $\mathcal{S}$ and $g = if$ in $\mathcal{S}_4/\mathcal{S}_4^0$.

[7] For example, see A. S. Wightman, Phys. Rev. **101**, 860 (1956).

(1) $R = R_S(H_1/\hat{L})$ as defined by (3.1) of Ref. 1.

(2) R is generated by the spectral projections of $\overline{A(h)}$, $h \in H_1$.

(3) R' (the commutant of R) is the set of all bounded operators C, satisfying

$$(C\Phi, A(h)\Psi) = (A(h)\Phi, C^*\Psi), \tag{3.1}$$

for all $h \in H_1$ and $\Phi, \Psi \in D_A$.

Proof. The spectral projections of $\chi(h) = \overline{(Ah)}$ and $W_S(h) = \exp i\chi(h)$, $h \in H_1$ generate the same von Neumann algebra, which proves the equivalence of (1) and (2). (2) obviously implies that $C \in R'$ satisfies (3.1). On the other hand, $W(h) = \lim_{N\to\infty} \sum_{n=0}^{N} (n!)^{-1}A(h)^n$ holds[8] on D_A in the strong sense. Hence, (3.1) implies $(C\Phi, W(h)\Psi) = (W(h)^*\Phi, C^*\Psi)$, from which we have $[W(h), C] = 0$ for all $h \in H_1$. Therefore, (3) is equivalent to (1) and (2).

Definition 1: Let B be an open subset of M. Let $\mathcal{S}(B)$ be the set of $f(x) \in \mathcal{S}_4$ with its support in B. We define $R(B)$ as $R_S(\mathcal{S}(B)/\hat{L})$ where $\mathcal{S}(B) \subset \mathcal{S}_4$ is considered as a linear subset of $\hat{L}$.

By (3.4) of Ref. 1, $R_S(\mathcal{S}(B)/\hat{L}) = R_S(\overline{\mathcal{S}(B)}/\hat{L})$ where $\overline{\mathcal{S}(B)}$ is the closure in $\hat{L}$. The Proof of theorem 1 is now trivial.

Proof of Theorem 1. Let $\mathfrak{D}(B)$ be the set of infinitely continuously differentiable functions with compact supports in B. If B and B_α are open and $B = \bigcup_\alpha B_\alpha$, any $h(x) \in \mathfrak{D}(B)$ can be written as $h = \sum_{i=1}^{N} h_{\alpha_i}$, $h_{\alpha_i} \in \mathfrak{D}(B_{\alpha_i})$. Since $\mathfrak{D}(B)$ is dense in $\mathcal{S}(B)$, we have (1) due to (3.12) of Ref. 1.

It is well known that (2.7) vanishes if the supports of f and g are spacelike to each other. Hence we have (2) due to the trivial part of (3.14) of Ref. 1. The support of $L(a, \Lambda)f$ is in $(a, \Lambda)B$ if that of f is in B. Hence, we have (3) due to the formula (7.16) of Ref. 1 and the definition of $U(a, \Lambda)$.

(4) follows from the (5.1) of Ref. 1.

Proof of (1.11) and (1.12). If $f \in \mathcal{S}(B)$, then there exists an open set B_1 such that $B_1 \supset$ supp. f, $B \supset \bar{B}_1$, due to the separation theorem. Hence, we have $R(B) = \hat{R}(B)$. The rest of (1.11) and (1.12) are obvious consequences of the definition.

4. A FORMULATION AT A FIXED TIME

In order to discuss von Neumann algebras of local observables at a fixed time, the formulation with $\mathfrak{H}_F(K)$, $\Psi_F(K)$, U_F and V_F is convenient. We clarify the connection between K and $\hat{L}$ in this section.

Let $\mathcal{S}_3$ be the set of real test functions of three-dimensional vectors $\mathbf{x}$, in the class $\mathcal{S}$. We denote the Fourier transform by

$$\tilde{f}(\mathbf{p}) = (2\pi)^{-\frac{3}{2}} \int e^{-i\mathbf{p}\mathbf{x}} f(\mathbf{x})\, d\mathbf{x}. \tag{4.1}$$

We equip $\mathcal{S}_3$ with the inner product of L_2 space,

$$(f, g)_{L_2} = \int \tilde{f}(\mathbf{p})^* \tilde{g}(\mathbf{p})\, d\mathbf{p} = \int f(\mathbf{x}) g(\mathbf{x})\, d\mathbf{x}. \tag{4.2}$$

The completion of $\mathcal{S}_3$ with respect to (4.2) is the Hilbert space L_2. We define the operators ω and ω^{-1} by

$$\widetilde{\omega f}(\mathbf{p}) = \omega(\mathbf{p})\tilde{f}(\mathbf{p}), \qquad \omega(\mathbf{p}) = (\mathbf{p}^2 + m^2)^{\frac{1}{2}}, \tag{4.3}$$

$$\widetilde{\omega^{-1} f}(\mathbf{p}) = \omega(\mathbf{p})^{-1}\tilde{f}(\mathbf{p}). \tag{4.4}$$

We now consider two more real inner products in $\mathcal{S}_3$,

$$(f_1, f_2)_\varphi = (f_1, \omega^{-1} f_2)_{L_2}, \tag{4.5}$$

$$(g_1, g_2)_\pi = (g_1, \omega g_2)_{L_2}. \tag{4.6}$$

The completion of $\mathcal{S}_3$ with respect to (4.5) and (4.6) are denoted by $\mathfrak{F}_\varphi$ and $\mathfrak{F}_\pi$, respectively. Each of them is considered as a Hilbert space equipped with the inner product (4.5) or (4.6). The mapping $f \to f$ for $f \in \mathcal{S}_3$ can be extended to bounded injections j_1 and j_2 of $\mathfrak{F}_\pi$ into L_2 and L_2 into $\mathfrak{F}_\varphi$ because of $(f, f)_\pi \geq m(f, f)$ and $(f, f) \geq m(f, f)_\varphi$. $j_1\mathfrak{F}_\pi \subset L_2$, $j_2 L_2 \subset \mathfrak{F}_\varphi$. The mapping $f \to \omega^{\frac{1}{2}} f$, $f \in \mathcal{S}_3$, on the other hand, induces a unitary mapping of $\mathfrak{F}_\pi$ onto L_2 and L_2 onto $\mathfrak{F}_\varphi$. $f \to \omega f$, $f \in \mathcal{S}_3$ induces a unitary mapping of $\mathfrak{F}_\pi$ onto $\mathfrak{F}_\varphi$. The extended mappings are denoted by $(\omega^{\frac{1}{2}} j_1)$, $(\omega^{\frac{1}{2}} j_2)$, and $(\omega j_2 j_1)$.

Let us now consider the splitting of $f(x) \in \mathcal{S}_4$ into symmetric and antisymmetric parts with respect to x^0.

$$f = f_+ + f_-, \qquad f_\pm(x) = \tfrac{1}{2}(f((x^0, \mathbf{x})) \pm f((-x^0, \mathbf{x}))). \tag{4.7}$$

In terms of the Fourier transform,

$$\tilde{f}_\pm((p^0, \mathbf{p})) = \tfrac{1}{2}(\tilde{f}((p^0, \mathbf{p})) \pm \tilde{f}((p^0, -\mathbf{p}))^*) \tag{4.8}$$

where we have used the fact that f, f_+, and f_- are real.

Let K be the subspace of $\hat{L}$, generated by time-symmetric $f = f_+$. Then βK is exactly the subspace generated by the time antisymmetric $f = f_-$ (see

[8] Consider the decomposition $\mathfrak{H}_s(\hat{L}) = \mathfrak{H}_s(\hat{L}_1) \otimes \mathfrak{H}_s(\hat{L}_2)$ where $\hat{L}_1$ is generated by f and βf and $\hat{L}_2$ is $\hat{L}_1^\perp$. The space $\mathfrak{H}_s(\hat{L}_1)$ can be identified with $L_2(R^1)$ where $W(h)$ is a multiplication operator $e^{ix||h||}$, $A(h)$ is a multiplication operator of $||h||\, x$, and $\Psi_s(L_1)$ is $\pi^{-\frac{1}{4}} e^{-x^2/2}$. Any vector in D is a finite linear combination of vectors $P(x)e^{-x^2/2} \otimes \Psi$, where $P(x)$ is a polynomial. By Taylor's theorem

$$|e^{ix\lambda} - \textstyle\sum_{n=0}^{N} (n!)^{-1}(ix\lambda)^n| \leq 2(\lambda x)^{N+1}/N!.$$

From this, the strong convergence is obvious.

footnote 6). Due to the symmetry of Δ_1, we have $K \perp \beta K$ and $L = K \oplus \beta K$. Hence, we may take K as K of Ref. 1.

Because of (4.8) and the realness of f, we have

$$\tilde{f}_\pm(p^0, \mathbf{p})^* = \tilde{f}_\pm(-p^0, -\mathbf{p}) = \pm\tilde{f}_\pm(p^0, -p). \tag{4.9}$$

Due to this equation, (2.6) reduces to

$$(h_1, h_2) = \int h_1(\omega(\mathbf{p}), \mathbf{p})^* h_2(\omega(\mathbf{p}), \mathbf{p})(\omega(\mathbf{p}))^{-1}\, d^3\mathbf{p}, \tag{4.10}$$

if h_1 and h_2 are both in K or both in βK. Because of (4.9) and (4.10), the mappings defined by

$$\widetilde{\delta_0 h}(\mathbf{p}) = \tilde{h}(\omega(\mathbf{p}), \mathbf{p}), \qquad h \in K \tag{4.11}$$

$$\widetilde{\delta_1 h}(\mathbf{p}) = (i\omega(\mathbf{p}))^{-1}\tilde{h}(\omega(\mathbf{p}), \mathbf{p}), \qquad h \in \beta K \tag{4.12}$$

are unitary mappings of K and βK onto $\mathfrak{F}_\varphi$ and $\mathfrak{F}_\pi$, respectively. In the following, we use $\mathfrak{F}_\varphi$ and $\mathfrak{F}_\pi$ for K and βK through this identification, and we write $\delta_0\beta\delta_1^{-1}$ and $\delta_1^{-1}\beta\delta_0$ simply as β. We then have the relations $\beta = -\omega j_2 j_1$ on $\mathfrak{F}_\pi$ and $\beta = j_1^{-1}j_2^{-1}\omega^{-1}$ on $\mathfrak{F}_\varphi$.

For $f \in \mathfrak{F}_\varphi$, we write $\chi(\delta_0^{-1}f)$ as $\varphi(f)$ and for $g \in \mathfrak{F}_\pi$, we write $\chi(\delta_1^{-1}g)$ as $\pi(g)$ as in Ref. 1.

By considering a sequence $f_n(x) = f_n^0(x^0)f(x)$ and $g_n(x) = g_n^0(x^0)g(x)$ where $f, g \in \mathcal{S}_3$, $f_n^0(x^0) \to \delta(x^0)$ and $g_n^0(x^0) \to -\delta'(x^0)$ in a suitable sense,[9] we may write a symbolic equation

$$\varphi(f) = \int A(0, \mathbf{x})f(\mathbf{x})\, d\mathbf{x},$$

$$\pi(g) = \int \dot{A}(0, \mathbf{x})g(\mathbf{x})\, d\mathbf{x},$$

where $\dot{A}$ is the time derivative of A.

For any $h \in \mathcal{S}_4$, we obtain by (4.5) and by (2.8) of Ref. 1,

$$W_S(h) = U_F(\delta_0 h_+)V_F(\delta_1 h_-) \times \exp \tfrac{1}{2}(j_2^{-1}\delta_0 h_+, j_1\delta_1 h_-)_{L_2}. \tag{4.13}$$

If we define a real solution of the Klein–Gordon equation by

$$F_h(x) = (-i/(2\pi)^{\frac{3}{2}}) \int e^{-i(p,x)} \times \delta((p, p) - m^2)\epsilon(p^0)\tilde{h}(p)\, d^4p \tag{4.14}$$

$$= \int \Delta(x - y)h(y)\, dy,$$

where $\epsilon(p^0) = 1$ if $p^0 > 0$ and -1 if $p^0 < 0$, then

$$(\delta_1 h_-)(\mathbf{x}) = F_h(0, \mathbf{x}), \tag{4.15}$$

$$(\delta_0 h_+)(\mathbf{x}) = -(\partial F_h/\partial x^0)(0, \mathbf{x}). \tag{4.16}$$

Because $\Delta(x)$ vanishes for space-like x, we have

$$F_h(x) = 0 \quad \text{if} \quad x \in (\text{supp. } h)', \tag{4.17}$$

where supp. h denotes the support of h.

5. THE SPACES $F_\varphi(\mathbf{B})$ AND $F_\pi(\mathbf{B})$

Let $\mathbf{B}$ be a measurable subset of S_0. We define $L_2(\mathbf{B})$ as the subspace of L_2 consisting of functions which vanish outside of $\mathbf{B}$. We define

$$F_\pi(\mathbf{B}) = j_1^{-1}(L_2(\mathbf{B}) \cap j_1\mathfrak{F}_\pi), \tag{5.1}$$

$$F_\varphi(\mathbf{B}) = \overline{j_2 L_2(\mathbf{B})} \qquad (\text{in } \mathfrak{F}_\varphi). \tag{5.2}$$

$F_\varphi(\mathbf{B})$ is a subspace of $\mathfrak{F}_\varphi$ by definition and $F_\pi(\mathbf{B})$ is also a subspace of $\mathfrak{F}_\pi$ due to $(f, f)_\pi \geq m(f, f)$.

We first prove simple properties of these spaces.

Lemma 2:

(1) If $\Delta_1 \supset \Delta_2$, $\quad \underset{\pi}{F_\varphi}(\Delta_1) \supset \underset{\pi}{F_\varphi}(\Delta_2)$; (5.3)

(2) $\vee F_\varphi(\Delta_\alpha) = F_\varphi(\cup \Delta_\alpha)$; (5.4)

(3) $\beta F_\pi(\Delta)^\perp = F_\varphi(\Delta^c)$, $\quad \beta F_\varphi(\Delta)^\perp = F_\pi(\Delta^c)$; (5.5)

(4) $\wedge F_\pi(\Delta_\alpha) = F_\pi(\cap \Delta_\alpha)$. (5.6)

Proof. (1) follows directly from the definition. (2). If Δ's are disjoint, then any $f \in L_2(\Delta)$ can be split uniquely as $f = \sum_\alpha f_\alpha$, $f_\alpha(x) = \varphi_\alpha(x)f(x)$, where $\varphi_\alpha(x)$ is the characteristic function of Δ and $\sum_\alpha$ converges in the strong topology of L_2. Hence, $L_2(\Delta) = \vee_\alpha L_2(\Delta_\alpha)$. Since j_2 is bounded, $j_2L_2(\Delta) = \vee_\alpha j_2L_2(\Delta_\alpha)$. Hence, we have (2). (3) $L_2(\Delta) \perp L_2(\Delta^c)$ and $L_2(\Delta) + L_2(\Delta^c) = L_2$, i.e., $L_2(\Delta^c) = L_2(\Delta)^\perp$ in L_2. Hence, $F_\pi(\Delta^c) = \{g \in \mathfrak{F}_\pi;\ f \in L_2(\Delta) \to (f, j_1g)_{L_2} = 0\}$. Since $(f, j_1g)_{L_2} = 0$ is equivalent to $(j_2f, \omega j_2 j_1 g)_\varphi = 0$, we have $F_\pi(\Delta^c) = \beta(j_2L_2(\Delta))^\perp$. Namely, $F_\pi(\Delta^c) = \beta F_\varphi(\Delta)^\perp$. Taking the orthogonal complement, we also have $F_\varphi(\Delta^c) = \beta F_\pi(\Delta)^\perp$. (4) follows from (2) and (3). This completes the proof of Lemma 2.

We now consider a relatively open subset $\mathbf{B}$ of S_0 with the property int $\bar{\mathbf{B}} = \mathbf{B}$, where int denoted the relative interior of $\bar{\mathbf{B}}$ in S_0. We define

$$\underset{\pi}{\bar{F}_\varphi}(\mathbf{B}) = \underset{\mathbf{B}' \supset \bar{\mathbf{B}}}{\wedge} \underset{\pi}{F_\varphi}(\mathbf{B}'), \tag{5.7}$$

$$\underset{\pi}{\hat{F}_\varphi}(\mathbf{B}) = \underset{\mathbf{E} \subset \mathbf{B}}{\vee} \underset{\pi}{F_\varphi}(\mathbf{E}), \tag{5.8}$$

where $\mathbf{B}'$ is a relatively open subset of S_0 and $\mathbf{E}$ is a closed subset of S_0.

Lemma 3: Let $\mathbf{B}$ and $\mathbf{B}_\alpha$ be relatively open subsets of S_0 satisfying int $\bar{\mathbf{B}} = \mathbf{B}$, int $\bar{\mathbf{B}}_\alpha = \mathbf{B}_\alpha$.

[9] For example, $\tilde{f}_n^0 \to 1$, $\tilde{g}_n^0 \to ip^0$ uniformly over compacts where $|\tilde{f}_n^0|$ and $|\tilde{g}_n^0|$ should be bounded by 1 and p^0 (or their multiples), respectively.

(1) $\bar{F}_{\varphi}(\mathbf{B}) \supset \hat{F}_{\varphi}(\mathbf{B})$; (5.9)

(2) $\bar{F}_{\varphi}(\mathbf{B}_1) \supset \bar{F}_{\varphi}(\mathbf{B}_2)$, $\hat{F}_{\varphi}(\mathbf{B}_1) \supset \hat{F}_{\varphi}(\mathbf{B}_2)$, if $\mathbf{B}_1 \supset \mathbf{B}_2$; (5.10)

(3) $\hat{F}_{\varphi}(\mathbf{B}_1) \supset \bar{F}_{\varphi}(\mathbf{B}_2)$, if $\mathbf{B}_1 \supset \bar{\mathbf{B}}_2$; (5.11)

(4) $\hat{F}_{\varphi}(\mathbf{B}) = F_{\varphi}(\mathbf{B})$, $F_{\pi}(\mathbf{B}) = F_{\pi}(\mathbf{B})$; (5.12)

(5) $\bigvee_{\alpha} \hat{F}_{\pi}(\mathbf{B}_\alpha) = \hat{F}_{\pi}(\mathbf{B})$, if $\mathbf{B} = \bigcup_{\alpha} \mathbf{B}_\alpha$; (5.13)

(6) $\beta\hat{F}_{\varphi}(\mathbf{B})^{\perp} = \bar{F}_{\varphi}(\bar{\mathbf{B}}^c)$. (5.14)

Proof. (1) and (2) are immediate consequences of the definition. If $\mathbf{B}_1$ is open and $\mathbf{B}_1 \supset \bar{\mathbf{B}}_2$ there exists another open set $\mathbf{B}_3$ with $\mathbf{B}_1 \supset \bar{\mathbf{B}}_3$ and $\mathbf{B}_3 \supset \bar{\mathbf{B}}_2$ due to the separation theorem. Hence (3) also follows from the definition.

Since

$$\bigcap_{\mathbf{B}' \supset \mathbf{B}} L_2(\mathbf{B}') = L_2(\bar{\mathbf{B}}) = L_2(\mathbf{B}), \quad j_1\bar{F}_{\pi}(\mathbf{B}) = j_1 F_{\pi}(\mathbf{B}).$$

Since j_1 is an injection, $\bar{F}_{\pi}(\mathbf{B}) = F_{\pi}(\mathbf{B})$. By using (3) of Lemma 2, and the equations $(\bar{\mathbf{B}})^c = \text{int } \mathbf{B}^c$, $\text{int } (\mathbf{E})^c = \overline{\mathbf{E}^c}$, we obtain (6). By using (6) and $\bar{F}_{\pi}(\mathbf{B}) = F_{\pi}(\mathbf{B})$, we have $\hat{F}_{\varphi}(\mathbf{B}) = F_{\varphi}(\mathbf{B})$.

Finally, we prove (5). Since $\mathbf{B}_\alpha \subset \mathbf{B}$, $\vee \hat{F}_{\pi}(\mathbf{B}_\alpha) \subset \hat{F}_{\pi}(\mathbf{B})$ by (2). Now consider $f \in F_{\pi}(\mathbf{B}')$, $\bar{\mathbf{B}}' \subset \mathbf{B}$. Since the closed set supp. $f (\subset \bar{\mathbf{B}}' \subset \mathbf{B})$ is covered by a set of open sets $\mathbf{B}_\alpha$, we have supp. $f \subset \bigcup_{i=1}^{n} \mathbf{B}_{\alpha_i}$ for some $\alpha_1 \cdots \alpha_n$. Furthermore, there exists $\varphi_i \in \mathcal{S}_3$ such that supp. $\varphi_i \subset \mathbf{B}_{\alpha_i}$, $\sum_{i=1}^{n} \varphi_i = 1$ on supp. f. Since the multiplication of $\varphi_i \in \mathcal{S}_3$ on a vector of $\mathfrak{H}_{\pi}$ is a bounded operator in $\mathfrak{H}_{\pi}$,[10] we have $f = \sum \varphi_i f$, $\varphi_i f \in F_{\pi}(\mathbf{B}_{\alpha_i})$. Hence, we have (5). This completes the proof of Lemma 3

We use the definition (5.7) also for closed subsets of S_0. Namely,

$$\overline{F_{\varphi}}(\mathbf{E}) = \bigwedge_{\mathbf{B}' \supset E} F_{\varphi}(\mathbf{B}'),$$

where $\mathbf{B}'$ is open. By definition $\bar{F}_{\varphi}(\mathbf{B}) = \bar{F}_{\varphi}(\bar{\mathbf{B}})$.

We now want to prove the following lemma, which deals with problems at the boundary of a region.

Lemma 4: Let $\mathbf{B}$ and $\mathbf{B}_\alpha$ be relatively open subsets of S_0 satisfying $\text{int } \bar{\mathbf{B}} = \mathbf{B}$ and $\text{int } \bar{\mathbf{B}}_\alpha = \mathbf{B}_\alpha$. Furthermore, their boundaries $\partial\mathbf{B}$ and $\partial\mathbf{B}_\alpha$ shall be many times differentiable surfaces or the union of several such surfaces which are joined together at many times differentiable curves. Then

(1) $\bar{F}_{\varphi}(\mathbf{B}) = \hat{F}_{\varphi}(\mathbf{B})$, $\bar{F}_{\pi}(\mathbf{B}) = \hat{F}_{\pi}(\mathbf{B})$; (5.15)

(2) $\bar{F}_{\varphi}(\partial\mathbf{B}) = \{0\}$, $\bar{F}_{\pi}(\partial\mathbf{B}) = \{0\}$; (5.16)

(3) $\hat{F}_{\varphi}(\mathbf{B}_1) \vee \hat{F}_{\varphi}(\mathbf{B}_2) = \hat{F}_{\varphi}(\text{int } \overline{(\mathbf{B}_1 \cup \mathbf{B}_2)})$; $\hat{F}_{\pi}(\mathbf{B}_1) \vee \hat{F}_{\pi}(\mathbf{B}_2) = \hat{F}_{\pi}(\text{int } \overline{(\mathbf{B}_1 \cup \mathbf{B}_2)})$. (5.17)

Proof. The first equation of (3) follows from (5.12), (5.4), (5.2), and $L_2(\mathbf{B}_1 \cup \mathbf{B}_2) = L_2 \text{ int } \overline{(\mathbf{B}_1 \cup \mathbf{B}_2)}$. By (5.10) and (5.14),

$$\bar{F}_{\pi}(\partial\mathbf{B}) \subset \bar{F}_{\pi}(\mathbf{B}) \wedge \bar{F}_{\pi}(\bar{\mathbf{B}}^c) = \beta(\hat{F}_{\varphi}(\bar{\mathbf{B}}^c) \vee \hat{F}_{\varphi}(\mathbf{B}))^{\perp}.$$

Hence, the second equation of (2) follows from the first of (3). Similarly, the first equation of (2) follows from the second of (3). The two equations of (1) are equivalent due to (5.14).

To prove the second equation of (1), let $f(x) \in \bar{F}_{\pi}(\mathbf{B})$ where we may assume[11] that supp. f is compact. Furthermore, by a partition of unity $(\varphi_\alpha(x) \in \mathfrak{D}$, supp. φ_α is small, $\sum \varphi_\alpha(x) = 1$ on supp. $f)$, we see that we have only to prove that $f \in \bar{F}_{\pi}(\mathbf{B})$ with sufficiently small support belongs to $\hat{F}_{\pi}(\mathbf{B})$. If the boundary of $\mathbf{B}$ is regular, and if supp. f is sufficiently small, there exists a vector $\mathbf{b}$ such that supp. $f(\mathbf{x} - \lambda\mathbf{b}) \subset \mathbf{B}$ for sufficiently small $\lambda > 0$. Now

$$||f(\mathbf{x} - \lambda\mathbf{b}) - f(\mathbf{x})||_{\pi}^2 = \int |1 - e^{i\lambda(\mathbf{b},\mathbf{p})}|^2 \, |\tilde{f}(\mathbf{p})|^2 \, \omega(\mathbf{p}) \, d^3\mathbf{p},$$

where $|\tilde{f}(\mathbf{p})|^2 \omega(\mathbf{p}) \in L_1$ and $|1 - e^{i\lambda(\mathbf{b},\mathbf{p})}|$ is uniformly bounded and tends to zero uniformly on compacts as $\lambda \to 0$. Hence, $\lim f(\mathbf{x} - \lambda\mathbf{b}) = f(\mathbf{x})$ in $\mathfrak{H}_{\pi}$. This proves the second equation of (1).

To prove the second equation of (3), we need the following lemmas.

Lemma 4.1:

$$\overline{j_1^{-1}\mathfrak{D} \cap \hat{F}_{\pi}(\mathbf{B})} = \hat{F}_{\pi}(\mathbf{B}). \tag{5.18}$$

Proof. By the footnote (11) and the definition (5.8), the function $f \in \mathfrak{H}_{\pi}$ with a compact support in $\mathbf{B}$ is dense in $\hat{F}_{\pi}(\mathbf{B})$. Let $g_\lambda \in \mathfrak{D}$, $g_\lambda(\mathbf{x}) = 0$ for $|\mathbf{x}| > \lambda$, $g_\lambda(\mathbf{x}) > 0$, $\int g_\lambda(\mathbf{x})d\mathbf{x} = 1$. Then the convolu-

[10] Since $\omega(\mathbf{p}_1 - \mathbf{p}_2)^{\frac{1}{2}} + \omega(\mathbf{p}_2)^{\frac{1}{2}} \geq \omega(\mathbf{p}_1)^{\frac{1}{2}}$, $||\omega(\mathbf{p})^{\frac{1}{2}} \int d\mathbf{p}' \, \bar{\varphi}(\mathbf{p} - \mathbf{p}')\tilde{f}(\mathbf{p}')||_{L_2}^2 \leq A^2||f||_{L_2}^2 + B^2||\omega^{\frac{1}{2}}f||_{L_2}^2$, where $A = \sup ||\int e^{i(\mathbf{x},\mathbf{p})}\omega(\mathbf{p})^{\frac{1}{2}}|\bar{\varphi}(\mathbf{p})|d\mathbf{p}| = \int \omega^{\frac{1}{2}}|\bar{\varphi}| \, d\mathbf{p}$ and $B = \sup \int e^{i(\mathbf{x}\mathbf{p})}|\bar{\varphi}(\mathbf{p})|d\mathbf{p}| = \int|\bar{\varphi}| \, d\mathbf{p}$. Since $||f||_{L_2} \leq m^{-1}||\omega^{\frac{1}{2}}f||_{L_2} = m^{-1}||f||_{\pi}$, the multiplication of $\varphi(\mathbf{x})$ is a bounded operator in $\mathfrak{H}_{\pi}$.

[11] If $h'(t) \in \mathfrak{D}_t$, $h(t) = 1$ for $t \geq 1$, $= 0$ for $t \leq 0$, and $g_{\lambda^a}(\mathbf{x}) = h((\mathbf{x}, \mathbf{a}) - \lambda)$, then it is easy to see that $||g_{\lambda^a}(\mathbf{x})f(\mathbf{x})||_{\mathfrak{H}_{\pi}}$ for any $f \in \bar{F}_{\pi}(B)$ is bounded in λ (footnote 9) and $\lim_{\lambda\to+\infty} \cdot(\varphi, g_{\lambda^a}(\mathbf{x})f(\mathbf{x}))_{\pi} = 0$ for any $\varphi \in \mathcal{S}$. Since $\mathcal{S}$ is dense in $\mathfrak{H}_{\pi}$, these two properties imply $\mathrm{W} - \lim_{\lambda\to\infty} g_{\lambda^a}(\mathbf{x})f(\mathbf{x}) = 0$ in $\mathfrak{H}_{\pi}$. For sufficiently many directions a_j,

$$\Pi_j(1 - g_{\lambda_j{}^{a_j}}(\mathbf{x}))f(\mathbf{x})$$

has a compact support and its limit as $\lambda_j \to \infty$ is $f(\mathbf{x})$. Hence, $f \in \mathfrak{H}_{\pi}$ with compact support is dense in $F_{\pi}(\mathbf{B})$.

tion $(g_\lambda \circ f)(\mathbf{x}) = \int g_\lambda(\mathbf{x} - \mathbf{y})f(\mathbf{y})\,d\mathbf{y}$ is in $\hat{F}_\pi(\mathbf{B})$ for sufficiently small λ and since $\tilde{g}_\lambda(\mathbf{p})$ is uniformly bounded and tends to 1 uniformly on compacts as $\lambda \to 0$, we have $\lim_{\lambda\to 0}(g_\lambda \circ f) = f$. Hence, we have (5.18).

Lemma 4.2: Let U be an open sphere, $S = \{(xyz);\ z = z(x, y)\}$ be a many times continuously differentiable surface which divides U into two parts U_+ and U_-, such that the two principal curvatures have constant signs (+, −, or 0) in U, and let g be any function in $\mathfrak{D}$ with its support in U. Then there exist two functions f_α^+ and f_α^- depending on $\alpha > 1$ such that $f_\alpha^+, f_\alpha^- \in \mathfrak{F}_\pi$, supp. $f_\alpha^\pm \subset \overline{U}_\pm$ and

$$\lim_{\alpha\to 1} ||f_\alpha^+ + f_\alpha^- - g||_\pi^2 = 0. \tag{5.19}$$

The proof of this lemma is given in the Appendix.

For the second equation of (3), we have to prove that any $f(x) \in \hat{F}$ (int $(\overline{\mathbf{B}_1 \cup \mathbf{B}_2})$) is a limit of a sum of functions in $\bar{F}_\pi(\mathbf{B}_1) = \underline{\hat{F}}_\pi(\mathbf{B}_1)$ and $\bar{F}_\pi(\mathbf{B}_2) = \hat{F}_\pi(\mathbf{B}_2)$, where we may take supp. f as small as we please, and we may assume $f \in \mathcal{S}$ due to the lemma 4.1. Of course, the nontrivial problem occurs at the boundary $\partial\mathbf{B}_1 \cap \partial\mathbf{B}_2$, where B_1 and B_2 occupy the opposite sides of the surface. If the boundary surface in question has principal curvatures of definite signs in a neighborhood of the supp. f, then the lemma 4.2 immediately proves the desired result. If the boundary surface consists of several parts S_k, each of which has principal curvatures of definite signs, then we consider many-times continuously differentiable surfaces S_α' through boundary curves $\partial S_{k_1} \cap \partial S_{k_2}$, which have principal curvatures of definite signs in a neighborhood of the boundary curve in question. We can use the Lemma 4.2 twice, first across each S_k and then across each S_α', and we obtain the desired result. (If there are some exceptional points on the boundary curves, we may consider an additional surface through the point, so that we may use the lemma 4.2 three times.)

Thus, if the boundary of each $\mathbf{B}_1$ and $\mathbf{B}_2$ is many-times continuously differentiable surfaces, we have the second equation of (3). This completes the proof of Lemma 4.

6. THE VON NEUMANN ALGEBRAS AT A FIXED TIME

We now give a connection between $\bar{R}(\mathbf{B})$ defined by (1.9) and the spaces $F_\varphi(\mathbf{B})$ and $F_\pi(\mathbf{B})$ of the preceding section. Namely, we prove the equation

$$\bar{R}(\mathbf{B}) = R_F(F_\varphi(\mathbf{B}), \beta F_\pi(\mathbf{B})/K) \tag{6.1}$$

together with the Theorem 2. Theorems 3 and 4 immediately follow from (6.1), the lemma 4, and the Theorem 1′ of Ref. 1.

First we prove

$$\bar{R}(\mathbf{B}) \supset R_F(F_\varphi(\mathbf{B}), \beta F_\pi(\mathbf{B})/K). \tag{6.2}$$

Let $f \in F_\varphi(\mathbf{B})$, $g \in F_\pi(\mathbf{B})$. Since $j_2L_2(\mathbf{B})$ is dense in $F_\varphi(\mathbf{B})$ and $\mathfrak{D}(\mathbf{B})$ is dense in $L_2(\mathbf{B})$, $j_2\mathfrak{D}(\mathbf{B})$ is dense in $F_\varphi(\mathbf{B})$ (in the strong topology of $\mathfrak{F}_\pi$). By the lemma 4.1, $j_1^{-1}\mathfrak{D}(\mathbf{B})$ is also dense in $F_\pi(\mathbf{B})$. Hence we consider $f(\mathbf{x}), g(\mathbf{x}) \in \mathfrak{D}(\mathbf{B})$. We define $f_n(x) = f_n^0(x^0)f(\mathbf{x})$ and $g_n(x) = g_n^0(x^0)g(\mathbf{x})$ where f_n^0 and g_n^0 are specified in the footnote 9. Considering f_n and g_n as elements of $\hat{L} = K \oplus \beta K$ and identifying K and βK with $\mathfrak{F}_\varphi$ and $\mathfrak{F}_\pi$, respectively, we have $\lim_{n\to\infty} f_n = f$ and $\lim_{n\to\infty} g_n = g$. Since supp. f_n and supp. g_n is in the interior of any given neighborhood of $\mathbf{B}$ for large n (note that supp. f and supp. g are compact), $U(f)$ and $V(g) \in \bar{R}(\mathbf{B})$. Hence, by (3.5) of Ref. 1, we have (6.2).

Let $\mathbf{B}_1 \supset \overline{\mathbf{B}_2}$, where $\mathbf{B}_1$ and $\mathbf{B}_2$ are relatively open on $x^0 = 0$ plane. We then have

$$\bar{R}(C(\mathbf{B}_1)) \supset \bar{R}(\mathbf{B}_1) \supset R_F(F_\varphi(\mathbf{B}_1), F_\pi(\mathbf{B}_1)/K) \tag{6.3}$$

from $C(\mathbf{B}_1) \supset \mathbf{B}_1$ and (6.2), and

$$R_F(F_\varphi(\mathbf{B}_1), F_\pi(\mathbf{B}_1)/K) \supset R(C(\mathbf{B}_2)) \tag{6.4}$$

from (4.13), (4.15), (4.16), and (4.17). Since there exists a relatively open set $\mathbf{B}_3$ such that $\mathbf{B}_1 \supset \overline{\mathbf{B}_3}$ and $\mathbf{B}_3 \supset \overline{\mathbf{B}_2}$, we have

$$R_F(F_\varphi(\mathbf{B}_1), F_\pi(\mathbf{B}_1)/K) \supset \bar{R}(C(\mathbf{B}_2)), \tag{6.5}$$

from (6.4) and (1.12) and

$$\hat{R}(C(\mathbf{B}_1)) \supset R_F(F_\varphi(\mathbf{B}_1), F_\pi(\mathbf{B}_1)/K), \tag{6.6}$$

from (6.3) and (1.12). The desired results (6.1) and (1.13) are an immediate consequence of (6.2)–(6.6), and (5.15), if we remember (3.17) and (3.18) of Ref. 1. This completes the proof of (6.1) and the Theorem 2.

Although we have made our discussion only for the $x^0 = 0$ plane, all the conclusion obviously holds for any $x^0 = t$ plane because of (1.5).

7. A COUNTEREXAMPLE FOR THE GENERAL CASE OF THE DUALITY THEOREM

By the general case of the duality theorem we mean the equation

$$\bar{R}(B')' = \hat{R}(B). \tag{7.1}$$

We have seen in the previous section that (7.1) holds for domains of the type $C(\mathbf{B})$ for the free scalar field. We now give an example of B for which (7.1) does not hold for the free scalar field.

(*Note added in proof.* T_1, T_2, and T are intervals on the time axis $\mathbf{x} = 0$.)

Let T_1 be the open interval (t_1, t_2) and similarly $T_2 = (-t_2, -t_1)$ where we take $t_2 > t_1 > 0$. $C(T_1)$ and $C(T_2)$ shall be the open double light cones spanned by T_1 and T_2 and $B = C(T_1) \cup C(T_2)$. We have $B' = C(T)'$ where $T = (-t_2, t_2)$. Since we already know the equality $R(C(T)')' = R(C(T))$, we see that (7.1) can not hold for this B if we can prove that $R(C(T)) \neq R(C(T_1)) \cup R(C(T_2))$. Due to the Theorem 1 of Ref. 1, we can show this by constructing g such that $g \notin \mathcal{S}(C(T))^\perp$ but $g \in (\mathcal{S}(C(T_1)) \cup \mathcal{S}(C(T_2)))^\perp$.

We first consider a hyperbolic equation in two variables,

$$\partial^2 h/\partial x^2 - \partial^2 h/\partial t^2 - m^2 h = 0. \tag{7.3}$$

For any given initial data $h(t, 0)$ and $(\partial h/\partial x)(t, 0)$ belonging to $\mathfrak{D}_t$ at $x = 0$, there exists a unique solution $h(t, x)$ which is infinitely often continuously differentiable. Furthermore, $h(t, x) = 0$ if $x^2 < (t - t')^2$ for all $t' \in$ (supp. $h(t, 0)$) $\cup$ (supp. $(\partial h/\partial x)(t, 0)$). Thus there exists an infinitely continuously differentiable solution $h_0(t, x)$ such that $h_0(t, x) = 0$ in a neighborhood of $C(T_1)$ and $C(T_2)$ and nonzero in a neighborhood of some point $(t, 0)$, $|t| < t_1$. Let h_1 be a function in $\mathfrak{D}_\mathbf{x}$ being equal to 1 for $|\mathbf{x}| \leq t_2$ and let

$$g_1(\mathbf{x}) = h_0(0, x_1)h_1(\mathbf{x}),$$

$$g_2(\mathbf{x}) = \frac{\partial h_0}{\partial t}(0, x_1)h_1(\mathbf{x}).$$

Then the solution $G(x)$ of the Klein–Gordon equation $(\Box_x + m^2)G = 0$ with the initial values $G(0, \mathbf{x}) = g_1(\mathbf{x})$, $(\partial G/\partial x^0)(0, \mathbf{x}) = g_2(\mathbf{x})$ coincides with $h_0(x^0, x^1)$ on $C(T)$. Since g_1 and g_2 are in $\mathfrak{D}_\mathbf{x}$, there exists $g \in \mathcal{S}_x$ such that $\delta_1 g_- = g_1$ and $\delta_0 g_+ = -g_2$. (For example, $\tilde{g}(p) = -\tilde{g}_2(p)\alpha((p, p)) + i\omega(p)\tilde{g}_1(p)\alpha((p, p))\epsilon(p^0)$ with an appropriate function α.) Then $\int \Delta(x - y)g(y)\, dy = G(x)$ is zero in $C(T_1)$ and $C(T_2)$ but nonzero in some point in $C(T)$. Hence, $\beta g \notin \mathcal{S}(C(T))^\perp$ but $\beta g \in (\mathcal{S}(C(T_1)) \cup \mathcal{S}(C(T_2)))^\perp$. This then disproves (7.1) for our choice of B.

Actually our argument above shows that for the regions in question not only the duality theorem fails but even the weaker statement

$$R(B'') = R(B) \tag{7.4}$$

is untrue, where $B'' = (B')'$. This property (sometimes referred to as the diamond theorem) is known to hold[12] for a certain type of regions in any theory satisfying standard axioms. It probably holds for any timewise convex region. It then appears reasonable to conjecture that for the free scalar field the duality theorem will be true for those regions for which (7.4) holds. We shall, however, not pursue these problems further in this paper.

[12] H. J. Borchers, Nuovo Cimento **19**, 787 (1961); H. Araki, Helv. Phys. Acta **36**, 132 (1963).

8. THE TYPE OF VON NEUMANN ALGEBRAS (I)

We will prove the Theorem 5 in this section. We first prove that the spaces $K^{(2)}$, $K^{(3)}$, K_4, and K_5 are $\{0\}$ for $K_1 = F_\varphi(\mathbf{B})$, $K_2 = \beta F_\pi(\mathbf{B})$ and $K = \mathfrak{F}_\varphi$, where $\mathbf{B}$ is as described before Theorem 2. It is enough to show that $K_1 \vee K_2^\perp = K$, $K_1^\perp \vee K_2 = K$, $\beta K_1 \wedge \beta K_2 = \{0\}$, $K_1^\perp \wedge K_2^\perp = \{0\}$. The first two equations follow from (5.17) and (5.14). The last two equations follow from the following lemma.

Lemma 5: If $f \in F_\varphi(\mathbf{B})$, then $(\beta f)(\mathbf{x})$ never vanishes in the entire neighborhood of any point in (supp. $f)^c$.

Proof. We define

$$F(z) = (2\pi)^{-\frac{3}{2}} \times \int \exp\,[i(\mathbf{p}\cdot\mathbf{z} - \omega(\mathbf{p})z^0)]\tilde{f}(\mathbf{p})\omega(\mathbf{p})^{-1}\, d^3\mathbf{p}. \tag{8.1}$$

$F(x)$, for a real value $z = x$, is a solution of the Klein–Gordon equation $(\Box_z + m)F = 0$ with the initial data $F(0, \mathbf{x}) = (\beta f)(\mathbf{x})$, $(\partial F/\partial x^0)(0, \mathbf{x}) = -if(\mathbf{x})$. Furthermore, $F(z)$ is analytic for Im z in the past light cone. Now if $(\beta f)(\mathbf{x})$ vanishes in an entire three-dimensional real neighborhood of a point $\mathbf{y} \in$ (supp. $f)^c$, then, since $f(\mathbf{x})$ also vanishes there, the solution $F(x)$ of the Klein–Gordon equation vanishes in a four-dimensional real neighborhood of the point $(0, \mathbf{y})$. Hence, the analytic function $F(z)$ identically vanishes, which implies $f(\mathbf{x}) = 0$. This proves the lemma.

We are now in the situation where $K_1' = K_1 = F_\varphi(\mathbf{B})$, $K_2' = K_2 = F_\pi(\mathbf{B})$. By virtue of the Theorem 4′ of Ref. 1, we can prove the Theorem 5 by showing that $\varphi(\beta F_\pi(\mathbf{B}); F_\varphi(\mathbf{B})/\mathfrak{F}_\varphi)$ is unbounded.

For any $f \in F_\varphi(\mathbf{B})$, $\omega^{-1}f(\, = \beta f)$ is in $\mathfrak{F}_\pi$ and, hence, is an L_2 function. It has a unique decomposition

$$\omega^{-1}f = f_1 + f_2, \tag{8.2}$$

where $f_1 \in L_2(\mathbf{B})$ and $f_2 \in L_2(\mathbf{B}^c)$. If f_1 and f_2 are in $\mathfrak{F}_\pi$, then

$$-\beta f_1 = f + \beta f_2 \tag{8.3}$$

is a decomposition of $-\beta f_1 \in \beta F_\pi(\mathbf{B})$ into $f \in F_\varphi(\mathbf{B})$ and $\beta F_\pi(\mathbf{B}^c) = F_\varphi(\mathbf{B})^\perp$. Hence,

$$\beta f_2 = \varphi(\beta F_\pi(\mathbf{B}); F_\varphi(\mathbf{B})/\mathfrak{F}_\varphi)f. \tag{8.4}$$

Conversely, if f is in the domain of $\varphi(\beta F_\pi(\mathbf{B}); F_\varphi(\mathbf{B})/\mathfrak{F}_\varphi)$, then f_2 defined by (8.4) is in $F_\pi(\mathbf{B}^c)$, f_1 subsequently defined by (8.3) is in $F_\pi(\mathbf{B})$ and they satisfy (8.2). Hence, if the closed operator $\varphi(\beta F_\pi(\mathbf{B}); F_\varphi(\mathbf{B})/\mathfrak{F}_\varphi)$ is bounded, then every f is in its domain and f_1 and f_2 in (8.2) must always belong to $\mathfrak{F}_\pi$. We now construct an example of $f \in \mathfrak{F}_\varphi$ for which $f_1 \notin \mathfrak{F}_\pi$.

Consider any infinitely differentiable function f with a compact support in $\mathbf{B}$. Because of the lemma 5, $\beta f_{\mathbf{a}}(\mathbf{x}) = \beta f(\mathbf{x} + \mathbf{a})$ for some arbitrarily small $\mathbf{a}$ must be nonzero at a given point on $\partial\mathbf{B}$. Therefore, we may assume that $(\beta f)(\mathbf{y}) \neq 0$ at a point $\mathbf{y} \in \partial\mathbf{B}$ such that, in a neighborhood of $\mathbf{y}$, $\partial\mathbf{B}$ has principal curvatures of definite signs. We further choose another function $g(\mathbf{x}) \in \mathfrak{D}$ with $g(\mathbf{y}) \neq 0$ and we choose supp. g small enough so that any relevant quantity concerning βf and $\partial\mathbf{B}$ change only slightly on supp. g and, in particular, $\partial\mathbf{B}$ has principal curvatures of definite signs in supp. g. By choosing a suitable coordinate, $\partial\mathbf{B}$ may be represented by the equation $x_3 = z(x_1, x_2)$, where z is many times continuously differentiable. Suppose B is on the side $x_3 > z(x_1, x_2)$.

If $f_1 \in \mathfrak{F}_\pi$, then $g(\mathbf{x})f_1(\mathbf{x}) = \theta(x_3 - z(x_1, x_2))g(\mathbf{x})\beta f(\mathbf{x}) \in \mathfrak{F}_\pi$. We disprove this by showing that $h_1 = \theta(x_3 - z(x_1, x_2))h(\mathbf{x}) \notin \mathfrak{F}_\pi$ for any h such that $h \neq 0$ on some point of $x_3 = z(x_1, x_2)$.

We define

$$h_{N+2}(\mathbf{x}) = \theta(x_3 - z(x_1, x_2)) \times \left\{h(\mathbf{x}) - e^{-x_3} \sum_{t=0}^{N} \frac{1}{t!}(x_3 - z(x_1, x_2))^t g_t(x_1, x_2)\right\}, \tag{8.5}$$

$$g_t(x_1, x_2) = \left(\frac{\partial}{\partial x_3}\right)^t (e^{x_3} h(\mathbf{x}))\Big|_{x_3 = z(x_1, x_2)}. \tag{8.6}$$

Since h_{N+2} is an N-times continuously differentiable function with an exponential decrease, it belongs to $\mathfrak{F}_\pi$ for sufficiently large N. (Actually $N = 0$ suffices.) We now investigate the function

$$\alpha_t(\mathbf{x}) = (1/t!)\theta(x_3 - z(x_1, x_2))e^{-x_3} \times (x_3 - z(x_1, x_2))^t g_t(x_1, x_2). \tag{8.7}$$

Its Fourier transform is

$$\tilde{\alpha}_t(\mathbf{p}) = (1 + ip_3)^{-(t+1)} G_t(\mathbf{p}), \tag{8.8}$$

$$G_t(\mathbf{p}) = (2\pi)^{-\frac{3}{2}} \int e^{-i(p_1x_1 + p_2x_2 + p_3z(x_1, x_2))} \times (e^{-z(x_1, x_2)} g_t(x_1, x_2))\, dx_1\, dx_2. \tag{8.9}$$

We now use the following lemma, which will be proved in the Appendix.

Lemma 6: Let $\mathcal{O}$ be a bounded open set in R^2 and

$$S = \{\mathbf{x}; x_3 = z(x_1, x_2), (x_1, x_2) \in \mathcal{O}\} \tag{8.10}$$

be a many times continuously differentiable surface whose principal curvatures have constant signs. Let $g(x_1, x_2)$ be a many times continuously differentiable function with a compact support in $\mathcal{O}$ and

$$G(p) = \int e^{-i(p_1x_1 + p_2x_2 + p_3z(x_1, x_2))} \times g(x_1, x_2)\, dx_1\, dx_2. \tag{8.11}$$

Let $\mathbf{n}(x_1, x_2)$ be a unit normal of S at (x_1, x_2) in the direction of $x_3 = +\infty$, and $\mathfrak{N} = \{\mathbf{n}(x_1, x_2); (x_1, x_2) \in \mathcal{O}\}$, $\mathfrak{N}_0 = \{\mathbf{n}(x_1, x_2); (x_1, x_2) \in \text{supp. } g\}$.

(1) *The case where both principal curvatures of S are nonzero.* $G(\mathbf{p})\ |\mathbf{p}|^N$ is bounded uniformly over the angle provided $\boldsymbol{\alpha} \equiv \mathbf{p}/|\mathbf{p}|$ lies outside of a neighborhood of $\mathfrak{N}_0$, N depending on the degree of the differentiability assumption. For $\boldsymbol{\alpha}$ in a neighborhood of $\mathfrak{N}_0$, $|\mathbf{p}|\ |G(\mathbf{p})|$ is bounded uniformly in $\boldsymbol{\alpha}$. Furthermore,

$$\lim_{\lambda\to\infty} \lambda\ |G(\lambda\mathbf{p})| = \sum \pi\ |g(x_1(\boldsymbol{\alpha}), x_2(\boldsymbol{\alpha}))|/|p_3|\ [R(\boldsymbol{\alpha})]^{\frac{1}{2}} \tag{8.12}$$

where the limit is uniform in $\boldsymbol{\alpha}$, $R(\boldsymbol{\alpha})$ is a nonvanishing differentiable function, $x_1(\boldsymbol{\alpha})$ and $x_2(\boldsymbol{\alpha})$ are (x_1, x_2) satisfying $\mathbf{n}(x_1, x_2) = \boldsymbol{\alpha}$, the number of such (x_1, x_2) for a given $\boldsymbol{\alpha}$ is uniformly bounded, and the summation is over all such (x_1, x_2).

(2) *The case where one principal curvature is zero.* The parametric equation for S is any one of the following three.

(i) $\mathbf{x} = \xi(s) + u\xi'(s)$, $|\xi'(s)| = 1$, $u \neq 0$, $\kappa(s) \neq 0$, $\tau(s) \neq 0$,

where $\kappa(s)$ and $\tau(s)$ are the curvature and the torsion of the curve $\xi(s)$.

(ii) $\mathbf{x} = \xi_0 + u\xi(s)$, $|\xi(s)| = |\xi'(s)| = 1$, $|\xi(s) + \xi''(s)| \neq 0$, $u \neq 0$.

(iii) $\mathbf{x} = \xi(s) + u\xi_0'$, $|\xi_0'| = |\xi'(s)| = 1$, $(\xi'(s), \xi_0') = 0$, $\xi''(s) \neq 0$.

In these cases $\mathfrak{N}$ becomes a curve. Let $\boldsymbol{\nu}(\mathbf{n})$ be the normal to $\mathfrak{N}$ at $\mathbf{n}$ along the surface of the unit sphere. In a sufficiently small neighborhood of $\mathfrak{N}$, any unit vector $\boldsymbol{\alpha}$ can be decomposed as

$$\boldsymbol{\alpha} = a(\boldsymbol{\alpha})\boldsymbol{\nu}(\mathbf{n}(\boldsymbol{\alpha})) + b(\boldsymbol{\alpha})\mathbf{n}(\boldsymbol{\alpha}), \tag{8.13}$$

where $\mathbf{n}(\boldsymbol{\alpha})$, $a(\boldsymbol{\alpha})$, and $b(\boldsymbol{\alpha}) = [1 - a(\boldsymbol{\alpha})^2]^{\frac{1}{2}}$ is uniquely defined by this equation. Then $|G(\mathbf{p})|\ (1 + |\mathbf{p}| \times$

$|a(\alpha)|)^N(1 + |\mathbf{p}|\ |b(\alpha)|)^{\frac{1}{2}}$ is bounded uniformly over α. For a finite $\tau = |\mathbf{p}|\ a(\alpha)$,

$$\lim_{|\mathbf{p}|\to\infty} (|\mathbf{p}|)^{\frac{1}{2}}\ |G(\mathbf{p})| = \sum (2\pi)^{\frac{1}{2}} \left| \int e^{i\tau u}\, du\varphi(u) \right.$$
$$\left. \times R(s(\alpha))g(x_1(u, s(\alpha)), x_2(u, s(\alpha))) \right| , \qquad (8.14)$$

where $s(\alpha)$ is the root of $\alpha = \mathbf{n}(x_1(u, s), x_2(u, s))$ ($\mathbf{n}(x_1(u, s), x_2(u, s))$ is independent of u), the summation is over all such s, $\varphi(u) = u^{-\frac{1}{2}}$ for the cases (i) and (ii), $\varphi(u) = 1$ for the case (iii), and $R(s(\alpha))$ is a bounded differentiable function.

(3) *The case where both principal curvatures are zero.* S is a plane and by a suitable choice of x_3 axis, $|G(\mathbf{p})|\ (|p_1|^2 + |p_2|^2)^N$ is bounded uniformly and is independent of p_3.

Due to this lemma, $\tilde{\alpha}_t(p)$ for $t \geq 1$ belongs to $\mathfrak{F}_\tau$. We now investigate $\tilde{\alpha}_0(p)$ for three cases.

(1) For large λ,

$$\int \omega(\lambda\alpha)\ |\tilde{\alpha}_0(\lambda\alpha)|^2\, d\alpha$$
$$\sim \lambda^{-3} \int (\pi^2\ |g(x_1(\alpha)x_2(\alpha)|^2/\alpha_3 R(\alpha))\, d\alpha.$$

Since the coefficient of λ^{-3} is nonvanishing $\tilde{\alpha}_0(p) \notin \mathfrak{F}_\tau$.

(2) For large λ,

$$\int \omega(\lambda\alpha)\ |\tilde{\alpha}_0(\lambda\alpha)|^2\, d\alpha \sim \lambda^{-3} \int (2\pi^2\varphi_1(u)^2\varphi_2(s)$$
$$\times R(s)^2\ |g(x_1(u, s), x_2(u, s))|^2)\, du\, ds,$$

where $\varphi_2(s)$ is a nonvanishing Jacobian of the transformation $d\alpha \to da\, ds$. Again $\tilde{\alpha}_0(p) \notin \mathfrak{F}_\tau$.

(3) Considering $p_3 \to \infty$, p_1, p_2 finite, we see that $\tilde{\alpha}_0(p) \notin \mathfrak{F}_\tau$. This completes the proof of Theorem 5.

9. THE TYPE OF VON NEUMANN ALGEBRAS (II)

In this section we give a quite general argument concerning the type of von Neumann algebras $R(B)$, which can be applied not only to the free scalar field under consideration but also to any $R(B)$ satisfying a set of axioms. The axioms we want to use are (1), (3), (4) of the Theorem 1 (excluding the locality) and the following.

Spectrum Condition. The spectral projections E defined by

$$U(a, 1) = \int e^{i(a,p)}\, dE(p)$$

has its spectrum in the future light cone $\bar{V}_+ = \{p;\ (p, p) \geq 0,\ p^0 \geq 0\}$. Furthermore, it does not have any point spectrum apart from the point $p = 0$ which has multiplicity 1 and corresponds to the unique vacuum Ψ_0.

We use the following lemma due to Reeh and Schlieder.[13]

Lemma 7: $\overline{R(B)\Psi_0} = \mathfrak{H}$ for any open set B.

Proof. Let $\Psi \in (R(B)\Psi_0)^\perp$. Take an open set B_1 satisfying $\overline{B_1} \subset B$. For any set of operators $Q_1 \cdots Q_n$ in $R(B_1)$, we consider

$$Q_i(x) = U(x, 1)Q_iU(x, 1)^{-1},$$
$$f(x_1 \cdots x_n) = (\Psi, Q_1(x_1) \cdots Q_n(x_n)\Psi_0).$$

By the spectrum condition, $f(x_1 \cdots x_n)$ is a boundary value of a function $f(z_1 \cdots z_n)$ analytic for Im z_1, Im $(z_2 - z_1) \cdots$ Im $(z_n - z_{n-1}) \in V_+$ (the interior of $\bar{V}_+$). By (1.5), $Q_i(x) \in R(B)$ if x is sufficiently small. Hence, $f(x_1 \cdots x_n) = 0$ for sufficiently small x_i. This means that $f(x_1 \cdots x_n) = 0$ identically. Since $Q_1(x_1) \cdots Q_n(x_n)\Psi_0$ generate $\vee_x R((x, 1)B)\Psi_0 = R(M)\Psi_0 = \mathfrak{H}$ due to (1.5), (1.2), and (1.6), we have $\Psi \in \mathfrak{H}^\perp$, i.e., $\Psi = 0$. Hence, $\overline{R(B)\Psi_0} = \mathfrak{H}$.

We now have the following theorem.

Theorem 6: Consider a theory satisfying (1), (3), and (4) of Theorem 1 and the spectrum condition. Let B be an open set in M, invariant under some translation a. If $R(B)$ is a factor and not equal to $B(\mathfrak{H})$, then $R(B)$ is not of type I.

Proof. If $R(B)$ is a factor of type I, we have a decomposition $\mathfrak{H} = \mathfrak{H}_1 \otimes \mathfrak{H}_2$, $R(B) = B(\mathfrak{H}_1) \otimes \mathbf{1}$. Furthermore, $U(a, 1)R(B)U(a, 1)^{-1} = R(B)$ due to (1.5) and the assumption on B. The vacuum Ψ_0 is a unique discrete eigenvector of $U(a, 1)$ and is a cyclic vector of $R(B)$ due to Lemma 7. Hence, by Lemma 10.3 of Ref. 1, $R(B)$ cannot be of type I.

The trivial possibility $R(B) = B(\mathfrak{H})$ is, of course, excluded for any region B whose causal complement B' has nonempty interior if the locality postulate (1.4) is used.

ACKNOWLEDGMENT

The author would like to thank Professor R. Haag for helpful discussions and critical reading of the manuscript.

APPENDIX

Proof of Lemmas 4.2 and 6

We first prove Lemma 4.2 assuming Lemma 6. Let G be a given function in $\mathfrak{D}$ and let $g(x_1, x_2) = G(x)\,|_{x_3=s(x_1,x_2)}$. We define

[13] H. Reeh and S. Schlieder, Nuovo Cimento **22**, 1051 (1961).

$$g_\alpha(\mathbf{x}) = h_\alpha(x_3 - z(x_1, x_2))g(x_1, x_2), \tag{A1}$$

$$h_\alpha(x_3) = \gamma(\alpha)^{-1} \int (p^2 + m^2)^{-\alpha/2} e^{ipx_3}\, dp, \tag{A2}$$

$$\gamma(\alpha) = \int (p^2 + m^2)^{-\alpha/2}\, dp$$

$$= m^{-(\alpha-1)}\Gamma((\tfrac{1}{2}\alpha) + \tfrac{1}{2})\Gamma(\tfrac{1}{2})\Gamma((\tfrac{1}{2}\alpha))^{-1}((\tfrac{1}{2}\alpha) - \tfrac{1}{2})^{-1}, \tag{A3}$$

where $\alpha > 1$. By definition,

$$g_\alpha(\mathbf{x}) = G(\mathbf{x}) \quad \text{if} \quad x_3 = z(x_1, x_2). \tag{A4}$$

The Fourier transform of g_α is given by

$$\tilde{g}_\alpha(\mathbf{p}) = \gamma(\alpha)^{-1}(p_3^2 + m^2)^{-\alpha/2} f(\mathbf{p}), \tag{A5}$$

$$f(\mathbf{p}) = (2\pi)^{-\frac{3}{2}}$$

$$\times \int e^{-i(p_1x_1 + p_2x_2 + p_3z(x_1, x_2))} g(x_1, x_2)\, dx_1\, dx_2. \tag{A6}$$

If the surface $x_3 = z(x_1, x_2)$ has principal curvatures of constant signs, then it follows from Lemma 6 that $|f(\mathbf{p})| < A(\mathbf{p}^2 + m^2)^{-N}$ if $|p_3|/|\mathbf{p}| < \epsilon$ for some $\epsilon > 0$ and

$$\int_{\alpha_3 \geq \epsilon} |f(|\mathbf{p}|\ \alpha)|^2\, d\alpha < \text{const}\ (\mathbf{p}^2 + m^2)^{-1}. \tag{A7}$$

Hence,

$$||g_\alpha||_\pi^2 < \gamma(\alpha)^{-2}\ \text{const} \int (\mathbf{p}^2 + m^2)^{-(2\alpha+1)/2}\, d^3\mathbf{p}$$

$$< \text{const}\ \gamma(\alpha)^{-2}\gamma(2\alpha - 1)$$

and in the limit $\alpha \to 1$, g_α tends to 0 in $\mathfrak{F}_\pi$. Namely, we have

$$\lim_{\alpha \to 1} (G - g_\alpha) = G. \tag{A8}$$

We now define

$$f_\alpha^+(\mathbf{x}) = \theta(x_3 - z(x_1, x_2))(G - g_\alpha)(\mathbf{x}). \tag{A9}$$

The lemma under discussion is proved if $f_\alpha^+ \in \mathfrak{F}_\pi$. We already know from the discussion of Sec. 8 that

$$f_\alpha'^+(\mathbf{x}) = \theta(x_3 - z(x_1, x_2))$$

$$\times (G - e^{-(x_3 - z(x_1, x_2))} g(x_1, x_2)) \tag{A10}$$

is in $\mathfrak{F}_\pi$. Hence we have only to prove that $f_\alpha^+ - f_\alpha'^+$ is in $\mathfrak{F}_\pi$.

The Fourier transform of $f_\alpha^+ - f_\alpha'^+$ is given by $\varphi(p_3)f(\mathbf{p})$ where

$$\varphi(p_3) = \frac{1}{1 + ip_3} - \frac{1}{i} \lim_{\epsilon \to 0} \int \frac{dp_3'}{p_3 - p_3' - i\epsilon}$$

$$\times (p_3'^2 + m^2)^{-\alpha/2}\gamma(\alpha)^{-1} \tag{A11}$$

$$= \gamma(\alpha)^{-1} \frac{1}{1 + ip_3} I,$$

$$I = \lim_{\epsilon \to 0} \int \frac{-p_3' - i}{p_3 - p_3' - i\epsilon} (p_3'^2 + m^2)^{-\alpha/2}\, dp_3'. \tag{A12}$$

We now demonstrate that $|\varphi(p_3)|\ p_3^\mu$ for some $\mu > 1$ is bounded. This together with the majorization of $f(p)$ due to Lemma 6 will yield the desired result: $f_\alpha^+ - f_\alpha'^+ \in \mathfrak{F}_\pi$.

Without loss of generality, we may assume $p_3 > m$. We divide the integration of I into 6 intervals:

$$I_1 = \int_{3p_3/2}^{\infty}, \quad I_2 = \int_{p_3 + \frac{1}{2}m}^{3p_3/2}, \quad I_3 = \int_{p_3 - \frac{1}{2}m}^{p_3 + \frac{1}{2}m},$$

$$I_4 = \int_{\frac{1}{2}p_3}^{p_3 - \frac{1}{2}m}, \quad I_5 = \int_{-\frac{1}{2}p_3}^{\frac{1}{2}p_3}, \quad I_6 = \int_{-\infty}^{-\frac{1}{2}p_3}.$$

In I_1 and I_6, we may majorize $|(-p_3' - i)/(p_3 - p_3' - i\epsilon)|$ by a finite constant and we have $|I_1| + |I_6| <$ const $p_3^{-(\alpha-1)}$. In I_5, we majorize $(p_3 - p_3' - i\epsilon)^{-1}$ by $2p_3^{-1}$ and we obtain $|I_5| <$ const $p_3^{-(\alpha-1)}$. In I_2 and I_4, we majorize $(p_3'^2 + m^2)^{-\alpha/2}(-p_3' - i)$ by const $p_3^{-(\alpha-1)}$ and we obtain $|I_2| + |I_4| <$ const $p_3^{-(\alpha-1)} \log p_3$. In I_3 we first shift the integration path in the complex p_3' plane so that the new path is a semicircle $m/2$ away from the singular point $p_3 = p_3'$. Then we majorize $(p_3'^2 + m^2)^{-\alpha/2}(-p_3' - i)$ by const $p_3^{-(\alpha-1)}$ and we obtain $|I_3| <$ const $p_3^{-(\alpha-1)}$. Thus we have $|I| <$ const $p_3^{-(\alpha-1)} \log p_3$ and Lemma 4.2 is proved.

We now prove Lemma 6.

(1) The case where both principal curvatures are nonzero.

Let us first consider $\mathbf{p}$ for which $\boldsymbol{\alpha} = \mathbf{p}/|\mathbf{p}|$ is outside of a neighborhood of $\mathfrak{N}_0$. For each point $(a, b) \in$ supp. g, either $\alpha_1 + \alpha_3(\partial z/\partial x_1)(a, b)$ or $\alpha_2 + \alpha_3(\partial z/\partial x_2)(a, b)$ is nonzero by the assumption because $\mathbf{n}(x_1, x_2)$ is proportional to $(-\partial z/\partial x_1, -\partial z/\partial x_2, 1)$. If

$$\alpha_1 + \alpha_3(\partial z/\partial x_1)(a, b) \neq 0,$$

there is a neighborhood $\mathfrak{N}(a, b)$ of (a, b) in which $\alpha_1 + \alpha_3(\partial z/\partial x_1)(x_1, x_2) \neq 0$ and the transformation

$$x_1' = (x_1 - a)(\alpha_1 + \alpha_3(x_1 - a)^{-1}(z(x_1, x_2) - z(a, x_2))$$

in nonsingular. If supp. $g \subset \mathfrak{N}(a, b)$, then by the change of variables $(x_1, x_2) \to (x_1', x_2)$ and by the partial integration with respect to x_1', we obtain

$$|\mathbf{p}|^N\ |G(\mathbf{p})| \leq \int \left| \left(\frac{\partial}{\partial x_1} \left(\alpha_1 + \alpha_3 \frac{\partial z}{\partial x_1}(x_1, x_2) \right) \right)^{-1} \right)^N$$

$$\times\ g(x_1, x_2) \Bigg| \, dx_1\, dx_2. \tag{A13}$$

For a general g, we choose a finite covering $\mathfrak{N}(a_i, b_i)$ of supp. g and consider the decomposition $g = \sum g_i$,

$g_i = \varphi_i g$, $\varphi_i \in \mathfrak{D}$, supp. $\varphi_i \subset \mathfrak{N}(a_i, b_i)$, $\sum \varphi_i(x_i, x_2) = 1$ if $(x_1, x_2) \in$ supp. g. Then we can apply the above result for each g_i. The integrand on the right-hand side of (A13) is bounded continuous function of (α, x_1, x_2) with a compact support with respect to x_1, x_2. Hence, (A13) is uniformly bounded and we have $|G(\mathbf{p})| < \text{const} \quad |\mathbf{p}|^{-N}$.

Next, let us consider $\mathbf{p} = |\mathbf{p}|\ \alpha$ for which α is in a neighborhood of $\mathfrak{N}_0$. Because $\mathbf{n}(x_1x_2)_3 \neq 0$ on $\mathfrak{N}_0$, we may take a neighborhood of $\mathfrak{N}_0$ in which $|\alpha_3| > \delta$. Since both the principal curvatures are nonzero, the mapping $(x_1, x_2) \to \mathbf{n}(x_1, x_2)$ is locally one to one and we may consider an inverse mapping $\alpha \to n^{-1}\alpha = (x_1(\alpha), x_2(\alpha))$. We consider the Taylor expansion of $z(x_1, x_2)$ around $(x_1(\alpha), x_2(\alpha))$:

$$\alpha_3 z(x_1, x_2) = \alpha_3 z(x_1(\alpha), x_2(\alpha)) - \alpha_1(x_1 - x_1(\alpha)) - \alpha_2(x_2 - x_2(\alpha)) + \frac{A}{2}\xi_1^2 + \frac{B}{2}\xi_2^2 + z_3(\xi_1, \xi_2), \quad \text{(A14)}$$

$$\xi_i = \sum_{j=1}^{2} a_{ij}(x_j - x_j(\alpha)), \quad \sum_k a_{ik}a_{jk} = \delta_{ij}, \quad \text{(A15)}$$

where (ξ_1, ξ_2) are the variables which diagonalize the second-order term and $A \cdot B \neq 0$ by the assumption of nonvanishing principal curvatures. We decompose z_3 as

$$z_3(\xi_1, \xi_2) = \xi_1^2 z_1(\xi_1, \xi_2) + \xi_2^2 z_2(\xi_1, \xi_2), \quad \text{(A16)}$$

$$z_1(\xi_1, \xi_2) = \xi_1^{-2}\left(z_3(\xi_1, 0) + \xi_2 \frac{\partial z_3}{\partial \xi_2}(\xi_1, 0)\right) = 0((\xi_1^2 + \xi_2^2)^{\frac{1}{2}}), \quad \text{(A17)}$$

$$z_2(\xi_1, \xi_2) = \xi_2^{-2}(z_3 - z_1\xi_1^2) = 0((\xi_1^2 + \xi_2^2)^{\frac{1}{2}}). \quad \text{(A18)}$$

We now consider the change of variables

$$u = \xi_1(1 + z_1(\xi_1, \xi_2))^{\frac{1}{2}}, \quad v = \xi_2(1 + z_2(\xi_1, \xi_2))^{\frac{1}{2}}, \quad \text{(A19)}$$

which is nonsingular in a neighborhood of $(x_1(\alpha), x_2(\alpha))$. Since $z_i(\xi_1, \xi_2)$ is a continuously differentiable function of x_1, x_2, and α vanishing at $x_1 = x_1(\alpha)$, $x_2 = x_2(\alpha)$, we may choose a neighborhood $\mathfrak{N}(\alpha_0)$ of any given α_0 on n^{-1} (supp. g) such that the above transformation $(x_1, x_2) \leftrightarrow (u, v)$ is nonsingular for any $\alpha \in \mathfrak{N}(\alpha_0)$ and $(x_1, x_2) \in \{(x_1(\alpha'), x_2(\alpha'))\ ;\ \alpha' \in \mathfrak{N}(\alpha_0)\}$. By this change of variables, we obtain

$$|G(\mathbf{p})| = \left| \int \exp\left[i\,|\mathbf{p}|\,(Au^2 + Bv^2)\right] \times g(x_1, x_2) \left|\frac{\partial(u, v)}{\partial(x_1, x_2)}\right|^{-1} du\, dv \right|. \quad \text{(A20)}$$

By changing variables to (u^2, v^2) and by partial integrations, we obtain

$$|\mathbf{p}|\ |G(\mathbf{p})| < \text{const}, \quad \text{(A21)}$$

if supp. $g \subset \{(x_1(\alpha'), x_2(\alpha'))\ ;\ \alpha' \in \mathfrak{N}(\alpha_0)\}$.

We choose a finite covering $\mathfrak{N}(\alpha_{0_i})$ of $\mathbf{n}$ (supp. g) and consider the decomposition $g = g_i$, $g_i = \varphi_i g$, $\varphi_i \in \mathfrak{D}$, supp. $\varphi_i \subset n^{-1}\mathfrak{N}(\alpha_{0_i})$, $\sum \varphi_i = 1$ on supp. g. We then apply (A21) as well as the earlier estimate $|G(\mathbf{p})| < \text{const}\ |\mathbf{p}|^N$ to each g_i and we obtain (A21) for a general g. The main contribution to (A20) comes from the point $u = 0$, $v = 0$ and is easily integrated. The remainder after subtracting out the main contribution is evaluated as follows. We subtract from $g(x_1, x_2)\ |\partial(u, v)/\partial(x_1, x_2)|^{-1}$ its value at $v = 0$, add the same expression again and subtract its value at $u = 0$, $v = 0$, the last term being the subtraction of the main contribution. After the change of variables to (u^2, v^2), we make the partial integration of the first two terms with respect to v and that of the last two terms with respect to u. The main contribution to the expression so obtained comes from the point $u = 0$, $v = 0$ again and is the order of $|p|^{-2}$ uniformly over α. The remainder can be evaluated in the same way and is smaller. Hence we have

$$G(\mathbf{p}) = C\ |\mathbf{p}|^{-1} + O(|\mathbf{p}|^{-2})$$

$$C = \pi g(x_1(\alpha)x_2(\alpha)) \exp i(p_1x_1(\alpha) + p_2x_2(\alpha) + p_3z(x_1(\alpha), x_2(\alpha))) \times \alpha_3^{-1} \left|\frac{\partial^2 z}{\partial x_1^2}\frac{\partial^2 z}{\partial x_2^2} - \left(\frac{\partial^2 z}{\partial x_1\, \partial x_2}\right)^2\right|^{-\frac{1}{2}}_{x_1 = x_1(\alpha),\, x_2 = x_2(\alpha)}$$

This proves (8.11).

(2) The case where one of the principal curvatures is zero.

If $\alpha = \mathbf{p}/|\mathbf{p}|$ is outside of a neighborhood of $\mathfrak{N}_0$, the same argument as in the case of (1) shows that $|\mathbf{p}|^N\ |G(\mathbf{p})|$ is uniformly bounded. We now consider an α which is near to $\mathfrak{N}_0$ in each of the cases (i)–(iii).

(i) In this case, the normal $\mathbf{n}((x_1(u, s), x_2(u, s))$ is given by $\xi'(s) \times \xi''(s)/|\xi''(s)|$ (up to a sign) and we will denote it by $\mathbf{n}(s)$. The equation $\mathbf{n}(\mathbf{p}) = \mathbf{n}(s)$ defines the function $s = s(\mathbf{p}) = s(\alpha)$. The vector $\nu(\mathbf{n}(s))$ is given by $\xi'(s)$ (up to a sign). We now define $\Xi_i(s)$ by

$$(\xi'(s), \mathbf{n}(s(\alpha))) = \tfrac{1}{2}(s - s(\alpha))^2\Xi_1(s), \quad \text{(A22)}$$

$$(\xi(s) - \xi(s(\alpha)), \mathbf{n}(s(\alpha))) = \tfrac{1}{2}(s - s(\alpha))^2\Xi_2(s), \quad \text{(A23)}$$

$$(\xi(s) - \xi(s(\alpha)), \xi'(s(\alpha))) = (s - s(\alpha))\Xi_3(s). \quad \text{(A24)}$$

We have $|\Xi_1(s(\alpha))| = |\kappa(s(\alpha))\tau(s(\alpha))| \neq 0$, $\Xi_2(s(\alpha)) = 0$, $\Xi_3(s(\alpha)) = 1$. We now consider the transformation

$$w = \left(s - s(\alpha) - \frac{a(\alpha)}{b(\alpha)}\,\Xi_3(s)(u\Xi_1(s) + \Xi_2(s))^{-1}\right) \times |u\Xi_1(s) + \Xi_2(s)|^{\frac{1}{2}}, \quad \text{(A25)}$$

$$v = u(\xi'(s), \xi'(s(\alpha))) + \frac{a(\alpha)}{2b(\alpha)}\,\Xi_3(s)^2(u\Xi_1(s) + \Xi_2(s))^{-2}, \quad \text{(A26)}$$

which is nonsingular in a neighborhood $\mathfrak{N}_\alpha$ of the point $s = s(\alpha)$. We obtain

$$|G(\mathbf{p})| = \left|\int \exp\,[i(|\mathbf{p}|\,a(\alpha)v + |\mathbf{p}|\,b(\alpha)w^2)] \times g(x_1, x_2)\left|\frac{\partial(w, v)}{\partial(s, u)}\right|^{-1}\left|\frac{\partial(x_1, x_2)}{\partial(s, u)}\right| dv\,dw\right|, \quad \text{(A27)}$$

which can be majorized by $\text{const}(1+|\mathbf{p}|\,|a(\alpha)|)^{-N}(1+|\mathbf{p}|\,|b(\alpha)|)^{-\frac{1}{2}}$ if supp. $g \subset \{(x_1(u, s), x_2(u, s)); u \neq 0, s \in \mathfrak{N}_\alpha\}$. The case of general g can be discussed exactly as before and we have

$$(1 + |\mathbf{p}|\,|a(\alpha)|)^N \times (1 + |\mathbf{p}|\,|b(\alpha)|)^{\frac{1}{2}}\,|G(\mathbf{p})| < \text{const}. \quad \text{(A28)}$$

The main contribution to (A29) can be evaluated as in (1), and we have (8.13).

(ii) This case is very similar to (i). We have $\mathbf{n}(s) \equiv \mathbf{n}(x_1(s, u), x_2(s, u)) = \xi(s) \times \xi'(s)$ and we define $s(\alpha)$ by $\mathbf{n}(\alpha) = \mathbf{n}(s(\alpha))$. We also have $\nu(\mathbf{n}(s)) = \xi(s(\alpha))$. We define $\Xi_1(s)$ by

$$(\xi(s)\mathbf{n}(s(\alpha))) = \tfrac{1}{2}(s - s(\alpha))^2\Xi_1(s). \quad \text{(A29)}$$

We have $\Xi_1(s(\alpha)) = |\xi''(s(\alpha)) + \xi(s(\alpha))| \neq 0$. We consider the change of variables

$$w = (s - s(\alpha))(u\Xi_1(s))^{\frac{1}{2}}, \quad \text{(A30)}$$

$$v = u(\xi(s)\xi(s(\alpha))). \quad \text{(A31)}$$

Then we obtain (A27) and (A28) follows. From (A27), we also have (8.13).

(iii) In this case, $\mathbf{n}(s) \equiv \mathbf{n}(x_1(s, u), x_2(s, u)) = \xi_0' \times \xi'(s)$ and $s(\alpha)$ is defined by $\mathbf{n}(s(\alpha)) = \mathbf{n}(\alpha)$. Then $\nu(\mathbf{n}(s)) = \xi_0'$. $\Xi_1(s)$ is defined by (A29) where we have $\Xi_1(s(\alpha)) = |\xi''(s(\alpha))| \neq 0$. We then consider the change of variables $v = u$, $w = (s - s(\alpha))\Xi_1(s)^{\frac{1}{2}}$ and again we obtain (A27) from which (A28) follows. From (A27), we also have (8.13).

(3) The case where both principal curvatures vanish. Here we have a plane and the stated result is obvious.

This completes the proof of Lemma 6.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 1 JANUARY 1964

Stationary Entropy Principle and Renormalization in Normal and Superfluid Systems. I. Algebraic Formulation*

CYRANO DE DOMINICIS AND PAUL C. MARTIN†
Service de Physique Théorique, Centre d'Etudes Nucléaires de Saclay (S.O.) France and
Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts
(Received 2 August 1963)

This paper shows how the dynamical and thermodynamical properties of an interacting quantum mechanical system with many degrees of freedom may be expressed and calculated solely in terms of renormalized propagators and renormalized vertices or interactions. The formulation employed is sufficiently general to encompass systems which have several components, with Fermi or Bose statistics, whether or not they exhibit superfluidity or superconductivity. The process of renormalization is the functional generalization of the thermodynamic transformation from the chemical potential and temperature to the energy and matter densities. With each set of variables (here, functions) is associated a natural thermodynamic function (here a functional). The natural functional for the unrenormalized potentials which occur in the Hamiltonian is the logarithm of the grand partition function; the natural functional for the fully renormalized variables, the distribution functions, is the entropy. In particular, a stationarity principle for a functional $F^{(2)}$ of distribution functions subject to constraints is shown to provide a fully renormalized description of the system. The numerical value of this functional, at the stationarity point at which the distribution functions take their actual value, is the entropy of the system. The equations of stationarity are expressions for the unrenormalized ν-body potentials v_ν in terms of the ν'-body distribution functions $G_{\nu'}$. The functionals $F^{(2)}$ and v_ν (of the distribution functions $G_{\nu'}$) are expressed as the solutions of closed functional differential equations which may be used to generate their power-series expansions. For a superfluid Bose system, as for the electromagnetic field interacting with matter, it is necessary to consider expectation values of odd, as well as even, numbers of field operators. In particular it is necessary to employ the expectation values G_ν for $2\nu = 1, 2, 3, 4$ field operators. For a fermion system, even if it is superconducting, only the functions G_ν for $2\nu = 2, 4$ are required. In contrast to other thermodynamical functionals, the entropy functional $F^{(2)}$ makes no reference to equilibrium parameters such as temperature and chemical potential.

I. INTRODUCTION

A. Renormalization and Entropy

MODERN methods for systematically determining the properties of many-particle systems are based on studying equations or series involving the interaction potentials of the Hamiltonian and either the true or the noninteracting one-body distribution function. The purpose of this paper and the companion paper is to extend these techniques to obtain equations or series which involve two-body correlation functions instead of the interaction potentials. In other words, we transfer attention from the potentials which cause the interaction between particles to the distribution functions which describe their effect.

The one-particle distribution function or propagator describes the true density and single-particle excitation spectrum. Thus, a formulation in terms of it is akin to the renormalization of mass in elementary particle physics. Correspondingly, the elimination of the interaction potential in terms of an effective interaction may be likened to charge renormalization. The techniques by which the renormalizations are performed here, can be applied to all formulations of the classical and quantum mechanical many-particle problem. However, they are particularly suited to the so-called Green's function formulation. The techniques apply, with essentially no change, to a multicomponent system (as for example, charges in interaction with an electromagnetic field). They will be applied, in this paper and its companion, to a superfluid system since it embodies as a special case all other less complicated quantum mechanical systems. Further extensions to describe nonequilibrium properties of classical and quantum systems have been studied and will be dealt with elsewhere.

In a certain sense, this scheme has a form anticipated long ago, in that all the properties of the many-particle system are expressed entirely in terms of quantities like scattering matrices. However these matrices (the two-particle correlation functions) depend both on the gross properties of the medium (e.g., its density and temperature), and on finer properties such as the energies of excitations

* A preliminary account of this work was presented at the Brown Conference on Phase Transitions and Irreversible Thermodynamics, 11–15 June 1962. See J. Math. Phys. 4, 208, 255 (1963).

† Sloan Foundation Fellow.

in the medium. The present work makes this idea into a precise working tool.

Furthermore, as we shall see in the example of superfluids, there may also occur correlations whose description requires another kind of scattering matrix or correlation function. This possibility is an important feature of our work.

When the theory is cast in a fully renormalized form, the entropy plays a central role and appears as a multiply stationary functional depending only on some distribution functions of the system.

B. Legendre Transformation

In the present paper, we point out and continually make use of the observation that renormalization amounts to a Legendre transformation from the potentials in the Hamiltonian to the distribution functions. In fact, from this viewpoint, it is possible to unify much of the diverse research on the many-particle problem. To make this notion more precise, we illustrate it schematically.

We consider a system whose Hamiltonian is the sum of a one-body potential $\hat{v}_1$ (including the kinetic energy and the chemical potential), and a two-body potential $\hat{v}_2$. The grand partition function

$$Z = \mathrm{Tr}\,[\exp(-\beta H)]$$

is then a functional of the potentials $\beta\hat{v}_1 = v_1$ and $\beta\hat{v}_2 = v_2$,

$$\ln Z = W\{v_1, v_2\}. \tag{1}$$

The one- and two-particle distribution functions G_1, G_2 (which are equal "time" limits of the time-dependent distribution functions introduced explicitly later) are defined by

$$G_1 = -(\delta/\delta v_1)W\{v_1, v_2\}, \tag{2}$$

$$G_2 = -(\delta/\delta v_2)W\{v_1, v_2\}. \tag{2'}$$

Explicit expansions in powers of v_1, v_2 for W, and hence G_1, G_2, could be obtained through perturbation theory.

Instead of W we may consider the function

$$F^{(1)} = W\{v_1, v_2\} + \int v_1 G_1 \tag{3}$$

obtained by a Legendre transformation. Although slightly different from its usual definition, $F^{(1)}$ is like a free energy and we consider it as a functional of its natural variables G_1 and v_2, with v_1 being expressed in terms of G_1 through (2). This transformation is the functional analog of the one used in thermodynamics to express the free energy in terms of its natural variables. From Eqs. (3) and (2) we obtain

$$v_1 = (\delta/\delta G_1)F^{(1)}\{G_1, v_2\}. \tag{4}$$

Equation (2) expresses G_1 as a functional of v_1; Eq. (4) inverts this relation, expressing v_1 as a functional of G_1. Constructing explicitly v_1 as a functional of G_1 (and v_2) is tantamount to performing explicitly the propagator renormalization.[1] We also observe that the functional

$$W = F\{G_1, v_2\} - \int v_1 G_1, \tag{5}$$

in which v_1 (and v_2) is kept *fixed*, remains *stationary* under variations of G_1.

In classical equilibrium statistical mechanics, $W\{v_1, v_2\}$ is more conveniently described as and expanded in terms of the function e^{-v_1}, which is the unperturbed local density. The expansion in powers of e^{-v_1} is the Ursell[2] fugacity expansion when the system is uniform. The importance of expressing the potentials in terms of the distribution functions rather than the converse has been repeatedly emphasized by Yvon[3] who first obtained the renormalized form (4) in 1935. The classical virial expansion[3,4] in terms of a local density results from substituting (4) into (5).

A stationary property of the type (5) was first pointed out by Lee and Yang.[5] For the quantum mechanical problem, they explicitly constructed W as a stationary functional of the average occupation number multiplied by the fugacity. Stationary functional forms of the same general character have been constructed in two other cases: first, as a functional of the average occupation number in a form which tends to the virial expansion in the classical limit[6]; second, as a functional of a function which might be called a quasiparticle occupation number.[7] In this case, W has the form one might expect from the Landau theory of Fermi liquids.[8]

---

[1] Propagator renormalization is a more appropriate terminology than mass renormalization. In this illustrative example, it reduces to a density renormalization, since G_1 is the equal time value of the one-body distribution function (the propagator).

[2] H. Ursell, Proc. Cambridge Phil. Soc. **23**, 685 (1927).

[3] J. Yvon, Actualités Sci. Ind. **203**, (1935); Cahiers Phys. **28**, (1945); Colloque de Thermodynamique, Bruxelles (January, 1948); J. Phys. Radium **10**, 373 (1949).

[4] J. Mayer, J. Chem. Phys. **5**, 67 (1937).

[5] T. D. Lee and C. N. Yang, Phys. Rev. **113**, 1165 (1959); **117**, 22 (1960).

[6] R. Balian, C. Bloch, and C. De Dominicis, Nucl. Phys. **25**, 529 (1961); **27**, 294 (1961).

[7] R. Balian and C. De Dominicis, Compt. Rend. **250**, 3885, 4111 (1960).

[8] L. Landau, Zh. Eksperim. i Teor. Fiz. **30**, 1058 (1956); **32**, 59 (1957); **34**, 262 (1958); [English transl.: Soviet Phys.—JETP **3**, 920 (1957); **5**, 101 (1957); **7**, 182 (1958)].

The three W functionals mentioned so far for quantum mechanical systems do not possess, actually, all the features desired here. Indeed v_1, $F^{(1)}$, and W become functionals of G_1 (or of an equivalent quantity) but remain also, partially, functionals of v_1; the elimination could of course be carried out by iterative substitutions of (4) into (5) but no resulting explicit functional of G_1 only is known so far. The full elimination of v_1 can be easily carried out, however, if one is willing to use a time-dependent distribution function G_1.[9-11] It appears to be a viewpoint by which one treats most directly equilibrium and transport properties; it is also well adapted to the zero temperature limit. This is the point of view that we shall use in this and the following paper.

To complete the renormalization, we now carry out a double Legendre transformation and consider the function

$$F^{(2)} = W\{v_1, v_2\} + \int [v_1 G_1 + v_2 G_2]. \qquad (3')$$

$F^{(2)}$ is identical with the entropy of the system and is considered here as a functional of its natural variables G_i , v_i being expressed in terms of G_i through (2) and (2'). Again we have the relation

$$v_i = (\delta/\delta G_i) F^{(2)}\{G_1, G_2\}, \qquad (4')$$

and the statement that

$$W = F^{(2)}\{G_1, G_2\} - \int [v_1 G_1 + v_2 G_2] \qquad (5')$$

is a functional stationary under variations of G_1 or G_2 when v_1 *and* v_2 *are kept fixed* in it. Equivalently we may say that the functional $F^{(2)}\{G_1 , G_2\}$ is stationary under variations of G_1, G_2 with the constraint that

$$\int [v_1 G_1 + v_2 G_2]$$

has a fixed numerical value, the stationarity conditions being identical with Eq. (4').

If the Hamiltonian contains only a one-body and a two-body potential, as we assumed here, the entropy is only a functional of the one- and two-body distribution functions and no longer depends upon the potentials or the thermodynamical parameters.

This is true when the Legendre transformation can be carried out completely explicitly as it can be for classical systems.[12,13] In quantum mechanical systems the only known form in which the potentials and the equilibrium parameters are absent[13] (the one we discuss here) involves "time"-dependent distribution functions.

It is straightforward to continue this procedure when there is a general n-body potential, and, by performing a n-fold Legendre transformation, express the entropy as a functional of $G_1, G_2, \cdots G_n$. While this is a necessity to achieve the elimination of the potentials when $\hat{v}_1$, $\hat{v}_2$, $\cdots$ $\hat{v}_n$ are actually present in the Hamiltonian, it is also feasible and useful even when there is no n-body potential. Some of the equations generalizing (4, 4') will then have a vanishing left-hand side.

C. Superfluid Systems

This situation arises in a very simple and interesting fashion in superfluid systems.

For treating this problem we have introduced elsewhere[14] an ensemble in which the thermodynamic properties are amenable to calculation. In this ensemble one specifies the average value of the field over regions large enough to contain many particles. The hierarchy of equations for the distribution functions in this ensemble is the same as the hierarchy obtained by introducing an external source potential in a grand canonical ensemble,[15]

[9] E. Montroll and J. Ward, Phys. Fluids **1**, 55 (1958).

[10] A. Abrikosov, L. Gorkov, and I. Dzyaloshinskii, Zh. Eksperim. i Teor. Fiz. **36**, 900 (1959) [English transl.: Soviet Phys.—JETP **9**, 636 (1959)]. E. Fradkin, Nucl. Phys. **12**, 465 (1959).

[11] P. Martin and J. Schwinger, Phys. Rev. **115**, 1342 (1959).

[12] T. Morita and K. Hiroike, Progr. Theoret. Phys. (Kyoto) **25**, 537 (1961).

[13] C. De Dominicis, J. Math. Phys. **3**, 983 (1962).

[14] P. Martin and P. Hohenberg (to be published).

[15] The procedure used here does not assume the form of the condensate wavefunction, or, in particular, that it is spatially uniform. Indeed, it permits one to determine the form of the interacting condensate wavefunction; it will have structure and be affected by impurities and boundaries [Cf. E. Gross, Ann. Phys. **4**, 57 (1958), and subsequent publications]. The special case in which a spatially uniform condensate is assumed has been discussed by several people. The pioneer work is that of N. Bogoliubov, J. Phys. (U.S.S.R.) **9**, 23 (1947).

The first systematic calculations at zero temperature are those of T. D. Lee and C. N. Yang, Phys. Rev. **105**, 1119 (1957), and of S. T. Beliaev, Zh. Eksperim. i Teor. Fiz. **34**, 417 (1958) [English transl.: Soviet Phys.—JETP **7**, 289 (1958)]. For superconductors, the pioneer calculations are the ones of J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **108**, 1175 (1957).

N. M. Hugenholz and D. Pines [Phys. Rev. **116**, 489 (1959)] have introduced an ensemble at zero temperature which can be used to systematically rederive the results of Beliaev. The techniques of these three authors are closest to the ones of the present paper. For superconductors, the systematization is commonly known but it has never to our knowledge been made available in the literature.

At nonzero temperature, another ensemble has been introduced by T. D. Lee and C. N. Yang [Phys. Rev. **117**, 897 (1960)] to discuss boson condensed systems. In their ensemble they have expressed the thermodynamical functions in terms of the condensate density and the average occupation

and allowing for the possibility that, as the external potential goes to zero, the average value of the field may remain finite. The external potential which is coupled to the quantum field is more appropriately called a source or "half-body" potential $v_{\frac{1}{2}}$. The average value of the quantum field in the ensemble is $G_{\frac{1}{2}}$. The equation akin to (4) determining $G_{\frac{1}{2}}$ becomes the homogeneous equation for the mean value of the field when $v_{\frac{1}{2}}$ tends to zero.

One may also perform additional transformations from the functional of the one- and two-body potentials and $G_{\frac{1}{2}}$, to a functional of the three distribution functions $G_{\frac{1}{2}}$, G_1, G_2. In so doing, it is convenient to introduce also a "$\frac{3}{2}$-body" potential which is eliminated by a Legendre transformation in favor of the three-point distribution function[16] $G_{\frac{3}{2}}$. This function also will not vanish in a superfluid system when $v_{\frac{1}{2}}$ and $v_{\frac{3}{2}}$ vanish. The result of performing the complete transformation is an expression for the entropy as a stationary functional of four functions which correspond physically to the condensate wavefunction; a generalized one-body distribution function; a three-point vertex connected with processes in which the number of excitations is changed, a single excitation dividing into two or two combining to one; and a four-point vertex like the one that ordinarily describes the scattering or correlation of a pair of particles. A similar but simpler technique applies for superconductors. In fact, it is found in a superfluid that except for one simple term, which is present whether or not the system is interacting, the functional depends on only two quantities which are essentially three- and four-point renormalized vertices. Thus indeed, the thermodynamic properties have, even in this complex system, an expansion purely in terms of true "scattering matrices".

In this paper we carry out the analysis using functional techniques to derive closed differential equations whose solution performs the renormalization and whose iterative solution would generate a perturbation expansion in powers of the renormalized vertices. By functional integration we also express the thermodynamical functions in the variational form described in this introduction. In the companion paper, diagrammatic techniques are applied to analyze and resum perturbation expansions leading to a simple characterization of each term of the various functional expansions.

II. STATEMENT OF PROBLEM AND DEFINITIONS

We consider a system described by a single Heisenberg field operator[17] $\Psi_\alpha(\mathbf{r})$ whose components are $\Psi_1(\mathbf{r}) = \psi(\mathbf{r})$ and $\Psi_2(\mathbf{r}) = \psi\dagger(\mathbf{r})$, the usual Heisenberg destruction and creation operators for the system. The dynamics of the system we consider is described by a Hamiltonian[18]

$$\hat{H} = \int (dx)\hat{w}_{\frac{1}{2}}(x_1)\Psi(x_1) + \int (dx)\hat{w}_1(x_1, x_2)\Psi(x_1)\Psi(x_2) + \int (dx)\hat{w}_{\frac{3}{2}}(x_1, x_2, x_3)\Psi(x_1)\Psi(x_2)\Psi(x_3) + \int (dx)\hat{w}_2(x_1, x_2, x_3, x_4)\Psi(x_1)\Psi(x_2)\Psi(x_3)\Psi(x_4).$$

The indices x_i refer to the spatial variables $\mathbf{r}_i$ (or in momentum space, $\mathbf{k}_i$) and the "spinor" index α_i. The spatial variables $\mathbf{r}_i$ also include implicitly any internal degree of freedom like spin. Thus an integration over x_i involves a spatial integration, a summation over internal degrees of freedom, and a summation over the two components α_i. We shall enploy the convention that repeated indices x_i are summed over. The Hamiltonians ordinarily considered are clearly of the form assumed above.

By using the equal time commutation (or anticommutation) relations it is always possible to rewrite the Hamiltonian so that it only involves ϵ-symmetrized potentials. By that we mean potentials which are completely symmetrized with respect to all of their arguments (for bosons), or completely antisymmetrized (for fermions). For example the ϵ-symmetrized form of $\hat{w}_1(x_1,x_2)$ contains the terms

number. However, this procedure presents the same problems we noted with respect to references 5, 6, and 7.

Here an ensemble is used in which the expectation value of the quantum field does not vanish. When such an ensemble is degenerate, as in a ferromagnet when there is a degeneracy in the spin direction, the degeneracy may be conveniently removed by a coupling to the conjugate variable. In the ferromagnet, this is the magnetic field, and as stressed by Schwinger (reference 19) for the quantum field this variable is a source. N. Bogoliubov, Physica **26**, 81 (1960) has proven rigorously in a simple model that if one allows the volume to approach ∞, and then a source to approach zero, in a grand canonical ensemble, one obtains not the correlation functions of the grand canonical ensemble, but the ones obtained in an ensemble like the one considered here if the volume, which is here taken finite is allowed to approach ∞. He has suggested that in other systems where the long-range order is known, the same effect will occur if one lets the appropriate field go to zero after taking the limit of infinite volume.

[16] The meaning of $v_{\frac{1}{2}}$, $v_{\frac{3}{2}}$ is made less mysterious by considering the two-component system mentioned earlier. In this case the external current is $v_{\frac{1}{2}}$ and the unrenormalized charge is the nonvanishing component of $v_{\frac{3}{2}}$. The corresponding distribution function is the electromagnetic form factor.

[17] Y. Nambu, Phys. Rev. **117**, 648 (1960).

[18] We have placed carets on these functions in anticipation of the fact that we shall employ slightly different related functions in most of our analysis, and wish to reserve the least cumbersome notation for them.

$$\hat{w}_1(x_1, x_2) + \epsilon\hat{w}_1(x_2, x_1),$$

and the ϵ-symmetrized form of $\hat{w}_{\frac{3}{2}}$ contains the terms

$$\hat{w}_{\frac{3}{2}}(x_1, x_2, x_3) + \hat{w}_{\frac{3}{2}}(x_2, x_3, x_1) + \hat{w}_{\frac{3}{2}}(x_3, x_1, x_2)$$
$$+ \epsilon\hat{w}_{\frac{3}{2}}(x_3, x_2, x_1) + \epsilon\hat{w}_{\frac{3}{2}}(x_2, x_1, x_3) + \epsilon\hat{w}_{\frac{3}{2}}(x_1, x_3, x_2),$$

where $\epsilon = +1$ for bosons and, $\epsilon = -1$ for fermions. We shall suppose that this symmetrization has been performed so that the Hamiltonian is given by

$$\hat{H} = \sum_{2\nu=0}^{4} \frac{1}{(2\nu)!}\hat{v}_\nu(x_1, \cdots x_{2\nu})\Psi(x_1) \cdots \Psi(x_{2\nu}), \tag{6}$$

where the potentials $\hat{v}_\nu$ are ϵ symmetrical. The identity by which this ϵ-symmetrization is accomplished is exhibited in the appendix. Likewise, the components of the potentials $\hat{v}_\nu$ for the usual two-body Hamiltonian are explicitly listed there.

The Heisenberg operators satisfy the equal time commutation relation

$$[\Psi(x), \Psi(x')]_+ = [\Psi_\alpha(\mathbf{r}), \Psi_{\alpha'}(\mathbf{r}')]_{-\epsilon}$$
$$= \delta(\mathbf{r} - \mathbf{r}')(\tau^\epsilon)_{\alpha\alpha'}, \tag{7}$$

where $[\ ,\ ]_{-\epsilon}$ is the commutator for bosons and anti-commutator for fermions and where τ^ϵ is the Pauli matrix $i\tau^2$ for bosons and τ^1 for fermions. Explicitly, they are given by

$$\tau^\epsilon \equiv (\tau^3)^{\frac{1}{2}(1+\epsilon)}\tau^1 \equiv \begin{bmatrix} 0 & 1 \\ -\epsilon & 0 \end{bmatrix}. \tag{8}$$

We shall also use

$$\sigma^\epsilon \equiv (\tau^\epsilon)^{-1} \equiv \begin{bmatrix} 0 & -\epsilon \\ 1 & 0 \end{bmatrix}.$$

As a function of time, the Heisenberg operator[18] $\hat{\Psi}(x, t)$ transforms according to

$$\hat{\Psi}(x, t) = \exp(i\hat{H}t)\Psi(x)\exp(-i\hat{H}t). \tag{9}$$

The distribution functions, or Green's functions of the system are defined by the relations

$$e^W\hat{G}_\nu(x_1, t_1; \cdots; x_{2\nu}, t_{2\nu})$$
$$= \epsilon^P \operatorname{Tr}[\exp(-\beta\hat{H})(\hat{\Psi}(x_1, t_1) \cdots \hat{\Psi}(x_{2\nu}, t_{2\nu}))_+], \tag{10}$$
$$\hat{G}_\nu(x_1, t_1; \cdots; x_{2\nu}, t_{2\nu})$$
$$\equiv \langle T(\hat{\Psi}(x_1, t_1) \cdots \hat{\Psi}(x_{2\nu}, t_{2\nu}))\rangle,$$

where

$$e^W \equiv \operatorname{Tr}[\exp(-\beta\hat{H})]. \tag{11}$$

In (10) the symbol $(\ )_+$ denotes that the operators $\hat{\Psi}(x_i, t_i)$ are to be arranged from right to left in order of increasing time. P is the order of the permutation which transforms the natural ordering into the time ordering. The combined operation $\epsilon^P(\ )_+$ is denoted by T. We also define the dimensionless quantities

$$H \equiv \beta\hat{H},$$
$$v_\nu(x_1, \cdots x_{2\nu}) \equiv \beta\hat{v}_\nu(x_1, \cdots x_{2\nu}), \tag{12}$$
$$\Psi(x, z) = \hat{\Psi}(x, -i\beta z),$$

and

$$G_\nu(x_1, z_1; \cdots; x_{2\nu}, z_{2\nu})$$
$$\equiv \hat{G}_\nu(x_1, -i\beta z_1; \cdots; x_{2\nu}, -i\beta z_{2\nu}),$$

and denote the variables (x_i, z_i) by j, writing

$$G_\nu(12 \cdots 2\nu)$$
$$\equiv e^{-W}\operatorname{tr}[e^{-H}T(\Psi(1)\Psi(2) \cdots \Psi(2\nu))]. \tag{13}$$

Finally we shall employ extensively the correlation functions $\tilde{G}_\nu$, cumulants of the distribution functions G_ν, explicitly defined in the next section and again in the companion paper. The periodicity (antiperiodicity) of the functions on the interval $z = 0$ to $z = 1$ is easily established and is connected with the fluctuation–dissipation theorem. This periodicity permits one to introduce Fourier series representations. For the developments in this paper, there is no need to carry out this Fourier transformation. It is useful, however, for concisely writing certain formulas and will be used in the following paper. It is also invaluable for performing calculations at finite temperatures.

III. GENERALIZED GREEN'S FUNCTIONS

A. Equation of Motion

In order to apply the first method, it is convenient to employ a procedure discussed in several other places.[11,19,20]

We introduce additional time-varying external potentials $\hat{u}_{\frac{3}{2}}$ and $\hat{u}_1$, and suppose that the system is described by the generalized Hamiltonian

---

[19] J. Schwinger, Proc. Natl. Acad. Sci. U. S. **37**, 952 (1951). The method we use for fermion sources is the one contained in this paper, where the fermion sources anti-commute with fermion operators. The same equations can be obtained without this artifice. In particular, the equations can be derived by the techniques of reference 9, Sec. 6, without using sources having anticommuting properties relative to the operator Ψ. In the equations preceding (32), the anticommutivity is implicit in the combination $[\delta/\delta u(1) + G_{\frac{3}{2}}(1))]$; the multiplication by $G_{\frac{3}{2}}(1)$ is also to be understood as a right multiplication.

[20] L. P. Kadanoff and G. Baym, *Quantum Statistical Mechanics* (W. A. Benjamin Company, Inc., New York, 1962).

$$\hat{H}_g(t) = \hat{H}_{ext}(t) + \sum_{2\nu=0}^{4} \frac{1}{(2\nu)!} \hat{v}_\nu(x_1, \cdots, x_{2\nu}) \times \hat{\Psi}^H(x_1, t) \cdots \hat{\Psi}^H(x_{2\nu}, t) \tag{14}$$

$$\hat{H}_{ext}(t) = \hat{u}_1(x, t)\hat{\Psi}^H(x, t) + \tfrac{1}{2}\hat{u}_1(x, t; x', t)\hat{\Psi}^H(x, t)\hat{\Psi}^H(x', t).$$

For a fermion system, the additional potential $\hat{u}_1$ is understood to be a completely anticommutative algebra having one element associated with each point in space and time. The operators $\hat{\Psi}^H(x, t)$ are the Heisenberg operators for the generalized Hamiltonian $H_g(t)$. The distribution functions for this generalized system could be defined in a manner identical to (10), but are equivalently and more conveniently written as

$$e^W \hat{G}_\nu(x_1, t_1; \cdots x_{2\nu}, t_{2\nu}) \equiv \epsilon^P \operatorname{Tr} [\exp(-\beta\hat{H}) \times (\hat{S}(-i\beta, t_1)\hat{\Psi}(x_1, t_1)\hat{S}(t_1, t_2) \cdots \hat{S}(t_{2\nu}, 0))_+] \times \exp_+[W\{\beta\hat{u}\}] = \operatorname{Tr} [\exp(-\beta\hat{H})\hat{S}(-i\beta, 0)]. \tag{15}$$

The operators $\hat{\Psi}(x, t)$ are still defined by (9), and are here the field operators for the generalized system in the interaction representation where $\hat{H}_{ext}(t)$ is the interaction. (Equivalently, they are of course the field operators in the Heisenberg representation for the system governed by $\hat{H}$ only). The quantity

$$\hat{S}(t_1, t_2) \equiv \left[\exp\left(-i\int_{t_2}^{t_1} \hat{H}^I_{ext}(t)\, dt\right)\right]_+$$
$$\hat{H}^I_{ext}(t) \equiv \hat{u}_1(x, t)\hat{\Psi}(x, t) + \tfrac{1}{2}\hat{u}_1(x, t; x', t)\hat{\Psi}(x, t)\hat{\Psi}(x', t) \tag{16}$$

is the transformation function from the interaction to the Heisenberg representation, for the generalized system.

The time-dependent external potentials have been introduced in order to generate $\hat{G}_\nu$ by functional differentiations of W or lower-order distribution functions. This reason is also behind the choice of the form (15) to define $\hat{G}_\nu$ where the dependence upon the time dependent potentials $\hat{u}$ appears explicitly (in the functions $\hat{S}$). Eventually, when the desired relations are established, the time-dependent potentials $\hat{u}$ can be set to zero, and the results for the original Hamiltonian H recovered. Notice that the functions $\hat{G}$ and W as defined by Eq. (15) reduce to those defined by (10) and (11) when we set $u = 0$. In the rest of this paper we work only with the functions $\hat{G}$ and W defined by (15), retaining the $\hat{u}$ dependence without bothering to exhibit it explicitly. The Heisenberg equation of motion for the operator $\hat{\Psi}^H$ is

$$i\sigma^\epsilon(\partial/\partial t)\hat{\Psi}^H(x, t) = \sigma^\epsilon[\hat{\Psi}^H(x, t), \hat{H}_g(t)] = \hat{u}_1(x, t) + \hat{u}_1(x, t; x_2, t)\hat{\Psi}^H(x_2, t) + \sum_{2\nu=0}^{4} \frac{1}{(2\nu-1)!} \hat{v}_\nu(x, x_2, \cdots, x_{2\nu}) \times \hat{\Psi}^H(x_2, t)\hat{\Psi}^H(x_3, t) \cdots \hat{\Psi}^H(x_{2\nu}, t). \tag{17}$$

We use now the relation between the Heisenberg and interaction representation of an operator $\hat{A}$ of the generalized system

$$\hat{A}^H(t) = \hat{S}(0, t)\hat{A}^I(t)\hat{S}(t, 0), \tag{18}$$

implying in particular

$$\operatorname{Tr} \hat{U}(-i\beta, 0)\hat{A}^H(t) = \operatorname{Tr} [\exp(-\beta\hat{H})\hat{S}(-i\beta, t)\hat{A}^I(t)\hat{S}(t, 0)], \quad \hat{U}(-i\beta, 0) = \exp(-\beta\hat{H})\hat{S}(-i\beta, 0). \tag{19}$$

Thus if we multiply both sides of Eq. (17) by $\hat{U}(-i\beta, 0)$, take the trace, and use (19) and (15) we obtain

$$i\sigma^\epsilon(\partial/\partial t_1)\hat{G}_1(x_1, t_1) = \hat{u}_1(x_1, t_1) + \hat{u}_1(x_1, t_1; x_2, t_1)\hat{G}_1(x_2, t_1) + \sum_{2\nu} \frac{\hat{v}_\nu(x_1, \cdots x_{2\nu})}{(2\nu-1)!} \hat{G}_{\nu-1}(x_2, t_1; \cdots; x_{2\nu}, t_1). \tag{20}$$

With this equation we can generate simply the hierarchy of equations for the functions $\hat{G}_\nu$ by using variational derivatives. As above it is convenient to transform from these physical quantities to dimensionless ones. We let

$$u_\nu = \beta\hat{u}_\nu$$
$$\exp W[u] = \operatorname{Tr}\left\{e^{-H}\left[\exp\left(-\int_0^1 H^I_{ext}(z)\, dz\right)\right]_+\right\}$$
$$H^I_{ext}(z) = u_1(x, z)\Psi(x, z) + \tfrac{1}{2}u_1(x, z; x', z)\Psi(x, z)\Psi(x', z)$$
$$G_\nu(x_1, z_1; \cdots; x_{2\nu}, z_{2\nu}) = \hat{G}_\nu(x_1, -i\beta z_1; \cdots; x_{2\nu}, -i\beta z_{2\nu}).$$

We also extend the matrix notation for the potentials u and v in (x, z) space writing them as

$$v_\nu(x_1, z_1; \cdots; x_{2\nu}, z_1) \to v_\nu(x_1, z_1; \cdots; x_{2\nu}, z_{2\nu}) \times \delta(z_2 - z_1) \cdots \delta(z_{2\nu} - z_1) \equiv v_\nu(1, \cdots, 2\nu). \tag{21}$$

With this matrix notation and a summation convention for repeated values of (x_i, z_i) or j, Eq. (20) becomes

$$-\sigma^{\epsilon}(\partial/\partial z_1)G_{\frac{1}{2}}(1) = u_{\frac{1}{2}}(1) + u_1(12)G_{\frac{1}{2}}(2) + \sum_{2\nu}\frac{v_\nu(1\cdots 2\nu)}{(2\nu-1)!}G_{\nu-\frac{1}{2}}(2,\cdots,2\nu). \quad (22)$$

B. Derivative Relations

From the above definition we can derive several useful differential relations. For example we have the relations

$$e^{W}G_{\frac{1}{2}n}(1\cdots n) = [-\delta/\delta u_{\frac{1}{2}}(n)]\cdots[-\delta/\delta u_{\frac{1}{2}}(1)]e^{W} \quad (23)$$

($n = 2\nu$ runs over integral values). For fermions, where the derivative is with respect to the anti-commutative object $u_{\frac{1}{2}}$, the derivatives are understood to be right-hand derivatives. Indeed, (23) enables us to derive the hierarchy of field distribution equations from (22).

In carrying out the analysis, it is convenient not only to employ the Green's functions G_ν but the correlated parts or cumulants of these functions, $\tilde{G}_\nu$, defined by

$$\exp\left[\sum_{n=0}^{\infty}\frac{1}{n!}u_{\frac{1}{2}}(1)\cdots u_{\frac{1}{2}}(n)\tilde{G}_{\frac{1}{2}n}(1\cdots n)\right] = \sum_{n=0}^{\infty}\frac{1}{n!}u_{\frac{1}{2}}(1)\cdots u_{\frac{1}{2}}(n)G_{\frac{1}{2}n}(1\cdots n), \quad (24)$$

or equivalently, the functions generated by

$$\tilde{G}_0 \equiv G_0 = 1; \qquad \tilde{G}_{\frac{1}{2}}(1) \equiv G_{\frac{1}{2}}(1);$$
$$[-\delta/\delta u_{\frac{1}{2}}(n)]\tilde{G}_{\frac{1}{2}(n-1)}(1,\cdots,n-1) = \tilde{G}_{\frac{1}{2}n}(1\cdots n), \qquad n \geq 2. \quad (24')$$

We may express G_ν in terms of $\tilde{G}_\nu$ either by comparing terms in the power series or by using the recursion relation

$$G_{\frac{1}{2}n}(1\cdots n) = \left(-\frac{\delta}{\delta u_{\frac{1}{2}}(n)} - \frac{\delta W}{\delta u_{\frac{1}{2}}(n)}\right)\cdots\left(-\frac{\delta}{\delta u_{\frac{1}{2}}(1)} - \frac{\delta W}{\delta u_{\frac{1}{2}}(1)}\right) = \left(-\frac{\delta}{\delta u_{\frac{1}{2}}(n)} - \frac{\delta W}{\delta u_{\frac{1}{2}}(n)}\right)G_{\frac{1}{2}(n-1)}(1\cdots n-1), \quad n \geq 1 \quad (25)$$

together with (24). For fermions, the terms in $\delta W/\delta u$ are understood to be ordered in the same way as the corresponding derivatives.

For the correlation functions we have the relations

$$G(1) = \tilde{G}(1),$$
$$G(12) = \tilde{G}(12) + \frac{1}{2!}\sum_P \epsilon^P G(1)G(2) = \tilde{G}(12) + \tfrac{1}{2}[G(1)G(2) + \epsilon G(2)G(1)],$$
$$G(123) = \tilde{G}(123) + \frac{1}{3!}\sum_P \epsilon^P[G(1)G(2)G(3) + 3G(1)\tilde{G}(23)],$$
$$G(1234) = \tilde{G}(1234) + \frac{1}{4!}\sum_P \epsilon^P[G(1)G(2)G(3)G(4) + 4G(1)\tilde{G}(234) + 6G(1)G(2)\tilde{G}(34) + 3\tilde{G}(12)\tilde{G}(34)]. \quad (26)$$

The summation extends over all permutations of the subsequently exhibited arguments. In these equations we have omitted, and henceforth we shall occasionally omit, the index which tells how many arguments a function G, $\tilde{G}$ or v has when we explicitly indicate its arguments.

In addition to the derivative relations indicated above, we can introduce similar relations involving the potential u_1,

$$-(\delta W/\delta u_1(12)) = \tfrac{1}{2}G(12),$$
$$[-\delta/\delta u_1(12)]G_{\nu-1}(3,\cdots,2\nu) = \tfrac{1}{2}G_\nu(1,\cdots,2\nu) - \tfrac{1}{2}G_1(12)G_{\nu-1}(3,\cdots,2\nu). \quad (27)$$

We shall refer to $G_{\frac{1}{2}}$ as the condensation wavefunction, and to $\tilde{G}_1$ as the propagator.

IV. EQUATIONS FOR THE RENORMALIZED CONDENSATE WAVEFUNCTION AND PROPAGATOR

A. Condensate

From the relations listed above, we may express the equation of motion for the condensate in terms of derivatives applied to the function $G_{\frac{1}{2}}$. In particular, we have

$$-[G_1^0]^{-1}(12)G(2) - \tfrac{1}{2}v(123)\left[\frac{-\delta}{\delta u(3)} + G(3)\right]G(2) - \frac{1}{3!}v(1234)\left[\frac{-\delta}{\delta u(4)} + G(4)\right]\left[\frac{-\delta}{\delta u(3)} + G(3)\right]G(2) = v(1) + u(1), \quad (28)$$

or equivalently, from (27),

$$-[G_1^0]^{-1}(12)G(2) - \tfrac{1}{2}v(123)\left[\frac{-\delta}{\delta u(3)} + G(3)\right]G(2) - \frac{1}{3!}v(1234)\left[-2\frac{\delta}{\delta u(34)} + G(34)\right]G(2) = u(1) + v(1), \quad (29)$$

where we have introduced the symbol

$$[G_1^0]^{-1}(12) = \sigma^{\epsilon}\delta(12)(\partial/\partial z_2) + u(12) + v(12) \quad (30)$$

whose inverse G_1^0 will never really be required.

Equation (29) suggests that we introduce an effective source function, $K_{\frac{1}{2}}(1)$ such that

$$-[G_1^0]^{-1}(12)G(2) = u(1) + v(1) - K(1). \quad (31)$$

It is apparent from Eqs. (30) and (31) that the potential $u(1)$ and $u(12)$ act like "time"-dependent parts of the potentials $v(1)$ and $v(12)$. In the subsequent equations we shall therefore incorporate the potentials u into time-dependent potentials v, which revert to the original potentials v when u vanishes.

From (31) and (29) it follows that the source function is given by

$$-K(1) = \tfrac{1}{2}v(123)\left[\frac{-\delta}{\delta v(3)} + G(3)\right]G(2) + \frac{1}{3!}v(1234)\left[-2\frac{\delta}{\delta v(34)} + G(34)\right]G(2),$$

or in terms of the cumulants,

$$-K(1) = \frac{1}{2!}v(123)[\tilde{G}(23) + G(2)G(3)] + \frac{1}{3!}v(1234)[\tilde{G}(234) + 3G(2)\tilde{G}(34) + G(2)G(3)G(4)]. \quad (32)$$

B. Propagator Renormalization

The equation for the source may be utilized to derive an expression for $[G_1^0]^{-1}$ as a functional of the propagator. In particular, by differentiating (31) with respect to $v(1')$ and employing the chain rule of differentiation, we obtain

$$[G_1^0]^{-1}(12)\tilde{G}(21') = \delta(11') - \left(\frac{\delta K(1)}{\delta G(2)}\right)\left(\frac{\delta G(2)}{\delta v(1')}\right). \quad (33)$$

We may therefore write

$$[G_1^0]^{-1}(11') - K_1(11') = \tilde{G}_1^{-1}(11'), \quad (34)$$

where

$$K_1(11') = \left(\frac{\delta K_{\frac{1}{2}}(1)}{\delta G_{\frac{1}{2}}(1')}\right)_{v_1, v_2/2, v_3} \quad (35)$$

By substituting (32) into (35) and using (24), or directly from (26) and the equation of motion for G_1, we obtain

$$-K(11') = \tfrac{1}{2}v(123)[\tilde{G}(234)\tilde{G}^{-1}(41') + 2G(2)\delta(31')] + \frac{1}{3!}v(1234)[\tilde{G}(2345)\tilde{G}^{-1}(51') + 3\tilde{G}(345)\tilde{G}^{-1}(51')G(2) + 3\tilde{G}(34)\delta(21') + 3G(3)G(4)\delta(21')]. \quad (36)$$

In accordance with standard usage, we shall call $-K_1$ the mass operator. In order to write this equation in a form appropriate for expressing K_1 as a functional of $G_{\frac{1}{2}}$ it is useful to introduce the functions, which we call C_ν, and which are generated by the equation

$$C_\nu(1 \cdots 2\nu) = [\delta/\delta G_{\frac{1}{2}}(2\nu)] \cdots [\delta/\delta G_{\frac{1}{2}}(3)]K_1(21). \quad (37)$$

In particular, taking into account that

$$-\tilde{G}^{-1}(1\bar{1})\frac{\delta}{\delta u_{\frac{1}{2}}(\bar{1})} = \frac{\delta}{\delta G_{\frac{1}{2}}(1)},$$

and that $A^{-1}\delta A + (\delta A^{-1})A = 0$, so that

$$-\frac{\delta}{\delta u_{\frac{1}{2}}(\bar{3})}K(21) = -\frac{\delta}{\delta u_{\frac{1}{2}}(\bar{3})}\tilde{G}^{-1}(21) = \tilde{G}^{-1}(2\bar{2})\left[\frac{\delta}{\delta u_{\frac{1}{2}}(\bar{3})}\tilde{G}(\bar{2}\,\bar{1})\right]\tilde{G}^{-1}(\bar{1}1),$$

we obtain from (24)

$$\begin{aligned} C_1(12) &= K(21) = [\delta/\delta G(1)]K(2), \\ C_{\frac{1}{2}}(123) &= \tilde{G}(\bar{1}\,\bar{2}\,\bar{3})\tilde{G}^{-1}(\bar{1}1)\tilde{G}^{-1}(\bar{2}2)\tilde{G}^{-1}(\bar{3}3), \\ C_2(1234) &= \tilde{G}(\bar{1}\,\bar{2}\,\bar{3}\,\bar{4})\tilde{G}^{-1}(\bar{1}1)\tilde{G}^{-1}(\bar{2}2)\tilde{G}^{-1}(\bar{3}3)\tilde{G}^{-1}(\bar{4}4) \\ &\quad - C_{\frac{1}{2}}(523)\tilde{G}(56)C_{\frac{1}{2}}(164) \\ &\quad - C_{\frac{1}{2}}(153)\tilde{G}(56)C_{\frac{1}{2}}(264) \\ &\quad - C_{\frac{1}{2}}(125)\tilde{G}(56)C_{\frac{1}{2}}(364). \end{aligned} \quad (38)$$

Loosely speaking, we may say that the functions $C_{\frac{1}{2}n}$ represent the interactions between n real excitations. The connection is only a loose one because the propagators we are using are analytically continued, time-ordered ones rather than the related causal ones, which would describe scattering. In this imprecise manner of speaking, the function $C_{\frac{1}{2}}$ describes the matrix element for a process in which one excitation divides into two, or two combine into one.

C_2 describes the way two excitations scatter into two others, apart from these processes in which a single excitation occurs in an intermediate state. The functions $C_{\frac{1}{2}n}$, which are the intrinsic parts of the cumulant functions $\tilde{G}_{\frac{1}{2}n}$, will be called the renormalized n-point vertex functions.[21]

There exists a sequence of approximations in which successively larger correlation functions (that is, higher derivatives of the source $K_{\frac{1}{2}}$ with respect to the condensate function $G_{\frac{1}{2}}$) are determined self-consistently. In each approximation, the remaining

[21] The symbol C stands for correlation, but as we shall see in the following paper, it also represents the contribution of completely connected diagrams.

correlation functions are then expanded in terms of the self-consistently determined ones. There are parallel sequences of approximations for all systems and, in particular, for classical systems and ferromagnets for which the correlation functions involve operators with more complicated commutation relations.[22]

In terms of the functions $C_{\frac{1}{2}n}$, we have the two identities

$$\begin{aligned}-K(11') &= \frac{1}{2!}v(123)[2\delta(31')G(2)\\ &+ C(\bar{2}\,\bar{3}\,1')\tilde{G}(\bar{3}\bar{3})\tilde{G}(\bar{2}\bar{2})]\\ &+ \frac{1}{3!}v(1234)[3\delta(21')G(3)G(4) + 3\delta(21')\tilde{G}(34)\\ &+ 3G(4)C(\bar{2}\bar{3}\,1')\tilde{G}(\bar{2}\bar{2})\tilde{G}(\bar{3}\bar{3})\\ &+ C(\bar{2}\,\bar{3}\,\bar{4}\,1')\tilde{G}(\bar{2}\bar{2})\tilde{G}(\bar{3}\bar{3})\tilde{G}(\bar{4}\bar{4})\\ &+ 3C(5\,\bar{3}\,\bar{4})\tilde{G}(56)C(61'\bar{2})\tilde{G}(\bar{2}\bar{2})\tilde{G}(\bar{3}\bar{3})\tilde{G}(\bar{4}\bar{4})], \qquad (39)\end{aligned}$$

$$\begin{aligned}-K(1) &= \frac{1}{2!}v(123)[G(2)G(3) + \tilde{G}(23)]\\ &+ \frac{1}{3!}v(1234)[G(2)G(3)G(4) + 3G(2)\tilde{G}(34)\\ &+ C(\bar{2}\,\bar{3}\,\bar{4})\tilde{G}(\bar{2}\bar{2})\tilde{G}(\bar{3}\bar{3})\tilde{G}(\bar{4}\bar{4})]. \qquad (40)\end{aligned}$$

These equations together with (31) and (34) may be used to generate an expansion in which $G_{\frac{1}{2}}$ is determined self-consistently, and all the functions C are expanded in powers of $G_{\frac{1}{2}}$, v_1, $v_{\frac{1}{2}}$, and v_2.

C. Renormalized Equations for the Source Function $K_{\frac{1}{2}}$ and the Mass Operator K_1

In order to see that these equations generate expressions for $K_{\frac{1}{2}}$ and K_1 which involve only the potentials $v_{\frac{1}{2}}$ and v_2 and the functions $G_{\frac{1}{2}}$ and $\tilde{G}_1$, and to investigate the structure of these expressions, we shall now prove two identities which relate the functions $C_{\frac{1}{2}}$ and C_2 to derivatives of $K_1\{12; G_{\frac{1}{2}}, G_1, v_{\frac{1}{2}}, v_2\}$ with respect to $G_{\frac{1}{2}}$ and $\tilde{G}_1$. From (37) and the partial derivative identity

$$\begin{aligned}C_{\frac{1}{2}}(321) &= \left(\frac{\delta K_1(12)}{\delta G_{\frac{1}{2}}(3)}\right)_{v_1}\\ &= \left(\frac{\delta K_1(12)}{\delta G_{\frac{1}{2}}(3)}\right)_{\tilde{G}_1} + \left(\frac{\delta K_1(12)}{\delta \tilde{G}(45)}\right)_{G_{1/2}}\left(\frac{\delta \tilde{G}(45)}{\delta G_{\frac{1}{2}}(3)}\right)_{v_1},\end{aligned}$$

and the formula for the derivative of an inverse,

$$\left(\frac{\delta \tilde{G}(45)}{\delta G_{\frac{1}{2}}(3)}\right)_{v_1} = \tilde{G}(46)\left(\frac{\delta K(67)}{\delta G_{\frac{1}{2}}(3)}\right)_{v_1}\tilde{G}(75),$$

we obtain the first of the desired identities,

$$\begin{aligned}C(321) &= \left(\frac{\delta K(12)}{\delta G(3)}\right)_{\tilde{G}_1} + \left(\frac{\delta K(12)}{\delta \tilde{G}(45)}\right)_{G_{1/2}}\\ &\times \tilde{G}(46)C(376)\tilde{G}(75). \qquad (41)\end{aligned}$$

To obtain the corresponding identity for C_2, we observe that

$$\begin{aligned}\left(\frac{\delta \tilde{G}_1(12)}{\delta v_1(34)}\right)_{G_{1/2}} &= \left(\frac{\delta \tilde{G}_1(12)}{\delta v_1(34)}\right)_{v_{1/2}}\\ &- \left(\frac{\delta \tilde{G}_1(12)}{\delta G_{\frac{1}{2}}(5)}\right)_{v_1}\left(\frac{\delta G_{\frac{1}{2}}(5)}{\delta v_1(34)}\right)_{v_{1/2}}. \qquad (42)\end{aligned}$$

In the right-hand side we may substitute the relation

$$\begin{aligned}-\left(\frac{\delta \tilde{G}_1(12)}{\delta v_1(34)}\right)_{v_{1/2}} &\equiv -\frac{\delta}{\delta v_{\frac{1}{2}}(2)}\frac{\delta}{\delta v_{\frac{1}{2}}(1)}\frac{\delta W}{\delta v_1(34)}\\ &= \tfrac{1}{2}[\tilde{G}_2(3\ 4\ 1\ 2) + G_{\frac{1}{2}}(3)\tilde{G}_{\frac{1}{2}}(412) + G_{\frac{1}{2}}(4)\tilde{G}_{\frac{1}{2}}(321)\\ &+ \tilde{G}_1(13)\tilde{G}_1(42) + \tilde{G}_1(32)\tilde{G}_1(41)],\end{aligned}$$

which follows from Eqs. (25) through (27), and the relation

$$\begin{aligned}\left(\frac{\delta \tilde{G}_1(12)}{\delta G_{\frac{1}{2}}(5)}\right)_{v_1}\left(\frac{\delta G_{\frac{1}{2}}(5)}{\delta v_1(34)}\right)_{v_{1/2}} &= -C(\bar{1}\,\bar{2}\,5)\tilde{G}(\bar{2}\bar{2})\tilde{G}(\bar{1}\bar{1})\\ &\times \tfrac{1}{2}[\tilde{G}(345) + G(3)\tilde{G}(54) + G(4)\tilde{G}(35)],\end{aligned}$$

which follows from (38). We then obtain for the left-hand side of (42)

$$\begin{aligned}\left(-\frac{\delta \tilde{G}_1(12)}{\delta v_1(34)}\right)_{G_{1/2}} &= \tfrac{1}{2}[\tilde{G}(13)\tilde{G}(42) + \tilde{G}(32)\tilde{G}(41)\\ &+ D(\bar{1}\,\bar{2};\ \bar{3}\,\bar{4})\tilde{G}(\bar{4}\bar{4})\tilde{G}(\bar{3}\bar{3})\tilde{G}(\bar{2}\bar{2})\tilde{G}(\bar{1}\bar{1})], \qquad (43)\end{aligned}$$

where

$$\begin{aligned}D_2(12; 34) &= C_2(1234) + C_{\frac{1}{2}}(135)\tilde{G}_1(56)C_{\frac{1}{2}}(624)\\ &+ C_{\frac{1}{2}}(164)\tilde{G}_1(56)C_{\frac{1}{2}}(523).\end{aligned}$$

Alternatively, we may write the left-hand side as

$$\begin{aligned}-\frac{\delta \tilde{G}_1(12)}{\delta v_1(34)} &= \frac{1}{2}\Big[\tilde{G}_1(13)\tilde{G}_1(42) + \tilde{G}_1(14)\tilde{G}_1(23)\\ &- \tilde{G}_1(15)\tilde{G}_1(62)\left(\frac{\delta K_1(56)}{\delta \tilde{G}_1(78)}\right)_{G_{1/2}}\left(\frac{\delta \tilde{G}_1(78)}{\delta v_1(34)}\right)_{G_{1/2}}\Big].\end{aligned}$$

Equating this expression with (43), we obtain the second of the desired identities,

$$\begin{aligned}\tfrac{1}{2}D_2(12; 34) &= \left(\frac{\delta K(12)}{\delta \tilde{G}(34)}\right)_{G_{1/2}} + \frac{1}{2}\left(\frac{\delta K(12)}{\delta \tilde{G}(78)}\right)_{G_{1/2}}\\ &\times \tilde{G}(7\bar{7})\tilde{G}(\bar{8}8)D_2(\bar{7}\,\bar{8}; 34). \qquad (44)\end{aligned}$$

These two equations determine D_2 and $C_{\frac{1}{2}}$ in terms

[22] In these cases, the diagrammatic analysis must begin with a rather complicated extension of the Wick's theorem discussed in the following paper.

of derivatives of K_1 with respect to $G_{\frac{1}{2}}$ and $\tilde{G}_1$. The result of inserting (41) and (44) into (39) and (40) are differential equations for $K_{\frac{1}{2}}$ and K_1 as functions of $G_{\frac{1}{2}}$, $\tilde{G}_1$, $v_{\frac{1}{2}}$, and v_2, which may be treated iteratively. More specifically, Eqs. (41) and (44) determine the functions $\tilde{G}_{\frac{1}{2}}$ and $\tilde{G}_2$ in terms of the functions $G_{\frac{1}{2}}$ and $\tilde{G}_1$, and two interaction kernels $(\delta K_1/\delta G_{\frac{1}{2}})_{\tilde{G}_1}$, and $(\delta K_1/\delta \tilde{G}_1)_{G_{1/2}}$. Equation (44) involves only the latter kernel and has been extensively studied in the particular case of normal Fermi systems. Equation (41) determines the three-point function $\tilde{G}_{\frac{1}{2}}$ from an integral equation with the same kernel and an inhomogeneous term which is the other interaction kernel $(\delta K_1/\delta G_{\frac{1}{2}})_{\tilde{G}_1}$. These two interaction kernels are in turn determined from the mass operator K_1 and source function $K_{\frac{1}{2}}$, which they determine through Eqs. (39) and (40). We may symbolically indicate how Eqs. (39), (40), (41), and (44) determine $\tilde{G}_1^{-1}$, and therefore the thermodynamic properties, in terms of the interaction potentials $v_{\frac{1}{2}}$ and v_2, and $\tilde{G}_1$, and $G_{\frac{1}{2}}$ by inserting (41) and (44) into (39) and (40),

$$-K_{\frac{1}{2}} = \frac{1}{2!}v_{\frac{1}{2}}[\tilde{G}_1 + G_{\frac{1}{2}}G_{\frac{1}{2}}] + \frac{1}{3!}v_2\Big[G_{\frac{1}{2}}G_{\frac{1}{2}}G_{\frac{1}{2}} + 3\tilde{G}_1G_{\frac{1}{2}} + \Lambda\Big(\frac{\delta K_1}{\delta G_{\frac{1}{2}}}\Big)_{\tilde{G}_1}\tilde{G}_1\tilde{G}_1\tilde{G}_1\Big], \tag{45}$$

and

$$-K_1 = \frac{1}{2!}v_{\frac{1}{2}}\Big[2G_{\frac{1}{2}} + \Lambda\Big(\frac{\delta K_1}{\delta G_{\frac{1}{2}}}\Big)_{\tilde{G}_1}\tilde{G}_1\tilde{G}_1\Big] + \frac{1}{3!}v_2\Big[3G_{\frac{1}{2}}G_{\frac{1}{2}} + 3\tilde{G}_1 + 3G_{\frac{1}{2}}\Lambda\Big(\frac{\delta K_1}{\delta G_{\frac{1}{2}}}\Big)_{\tilde{G}_1}\tilde{G}_1\tilde{G}_1 + \Big(\frac{\delta K_1}{\delta \tilde{G}_1}\Big)_{\frac{1}{2}}\Lambda\tilde{G}_1\tilde{G}_1 + \Lambda\Big(\frac{\delta K_1}{\delta G_{\frac{1}{2}}}\Big)_{\tilde{G}_1}\tilde{G}_1\Lambda\Big(\frac{\delta K_1}{\delta G_{\frac{1}{2}}}\Big)_{\tilde{G}_1}\tilde{G}_1\tilde{G}_1\Big], \tag{46}$$

with

$$\Lambda^{-1} \equiv 1 - \tilde{G}_1\tilde{G}_1\Big(\frac{\delta K_1}{\delta \tilde{G}_1}\Big)_{G_{\frac{1}{2}}}.$$

It is clear from this form that $K_{\frac{1}{2}}$ and K_1 are determined by $G_{\frac{1}{2}}$ and $\tilde{G}_1$, and in particular, that we may generate power series for $K_{\frac{1}{2}}$ and K_1 in $G_{\frac{1}{2}}$ and $\tilde{G}_1$. We shall show later that the power series really depend on only two functions rather than four, that is, that the potentials and propagators always occur in only two combinations.

When $v_{\frac{1}{2}} = 0$, and $G_{\frac{1}{2}} = 0$, as for example, for a single Fermi field, or a Bose field when $T > T_c$ and $v_{\frac{1}{2}} = 0$, (45) gives $K_{\frac{1}{2}} = 0$, and (46) which summarizes (39) and (44) reduces to

$$-K_1 = \tfrac{1}{2}v_2\tilde{G}_1 + \frac{1}{3!}v_2\tilde{G}_1\tilde{G}_1\Big(\frac{\delta K_1}{\delta \tilde{G}_1}\Big)\Big[1 - \tilde{G}_1\tilde{G}_1\frac{\delta K_1}{\delta \tilde{G}_1}\Big]^{-1}. \tag{47}$$

This equation determines the mass operator K_1 as a functional of $\tilde{G}_1$, (and v_2). It can, in particular, be used to generate the so-called irreducible diagram expansion of the mass operator.[23] We shall show later that the functions, v_2 and $\tilde{G}_1$ occur in essentially one combination in this expansion.

V. THERMODYNAMICAL FUNCTIONS RESULTING FROM PROPAGATOR AND CONDENSATE RENORMALIZATION

Thus far we have developed expressions which determine the relationship between $v_{\frac{1}{2}}$ and v_1, and $G_{\frac{1}{2}}$ and $\tilde{G}_1$. We wish now to use those relations to construct variational principles by the method described in the introduction.

A. Thermodynamical Functions after Condensate Renormalization

From Eq. (45) for $K_{\frac{1}{2}}$, and Eq. (46) for K_1, we can generate $K_{\frac{1}{2}}$ and consequently $v_{\frac{1}{2}}$ as a functional of $G_{\frac{1}{2}}$, v_1, $v_{\frac{1}{2}}$, and v_2. The connection between $v_{\frac{1}{2}}$ and $G_{\frac{1}{2}}$ together with the statement

$$-\delta W/\delta v_{\frac{1}{2}}(1) = G_{\frac{1}{2}}(1) \tag{48}$$

permit us to carry out a Legendre transformation to the function $F^{(\frac{1}{2})}$ whose natural variables are $G_{\frac{1}{2}}$, v_1, $v_{\frac{1}{2}}$, v_2,

$$F^{(\frac{1}{2})}\{G_{\frac{1}{2}}\} = W + v_0 + G_{\frac{1}{2}}(1)v_{\frac{1}{2}}(1). \tag{49}$$

Since we have the conjugate relation [cf. Eq. (4′)]

$$\delta F^{(\frac{1}{2})}\{G_{\frac{1}{2}}\}/\delta G_{\frac{1}{2}}(1) = v_{\frac{1}{2}}(1) \tag{50}$$

for $v_{\frac{1}{2}}$ as a functional of $G_{\frac{1}{2}}$ we may write [cf. Eq. (5′)] the stationarity principle

$$\delta W/\delta G_{\frac{1}{2}}(1) = 0, \tag{51}$$

where

$$W \equiv F^{(\frac{1}{2})}\{G_{\frac{1}{2}}\} - G_{\frac{1}{2}}(1)v_{\frac{1}{2}}(1) - v_0$$

is a functional of $G_{\frac{1}{2}}$, and $v_{\frac{1}{2}}$ is fixed.

We may compute $F^{(\frac{1}{2})}$ by integrating Eq. (50) with $v_{\frac{1}{2}}$ as a functional of $G_{\frac{1}{2}}$ determined from (31),

[23] This irreducible-diagram expansion is more precisely the expansion in terms of 1-irreducible 1-diagrams in the terminology of the following paper.

$$\frac{\delta F^{(\frac{1}{2})}}{\delta G_{\frac{1}{2}}(1)} = \frac{\delta}{\delta G_{\frac{1}{2}}(1)} \times (-\tfrac{1}{2}G_{\frac{1}{2}}(2)[G_1^0]^{-1}(23)G_{\frac{1}{2}}(3)) + K_{\frac{1}{2}}(1), \quad (52)$$

$$F^{(\frac{1}{2})}\{G_{\frac{1}{2}}, v_1, v_{\frac{1}{2}}, v_2\} = -\tfrac{1}{2}G_{\frac{1}{2}}(2)[G_1^0]^{-1}(23)G_{\frac{1}{2}}(3) + \int_{\underline{G}_{1/2}}^{G_{1/2}} K_{\frac{1}{2}}\{1; G'_{\frac{1}{2}}\}\ \delta G'_{\frac{1}{2}}(1) + \Phi^{(\frac{1}{2})}\{\underline{G}_{\frac{1}{2}}, v_1, v_{\frac{1}{2}}, v_2\}, \quad (53)$$

where $\underline{G}_{\frac{1}{2}}$ is an arbitrary lower limit of integration. If we choose $\underline{G}_{\frac{1}{2}} = 0$, we have

$$\Phi^{(\frac{1}{2})} = W\{v_{\frac{1}{2}}, v_1, v_{\frac{1}{2}}, v_2\};$$

for the value $v_{\frac{1}{2}}$,

$$v_{\frac{1}{2}} = K_{\frac{1}{2}}\{\underline{G}_{\frac{1}{2}} = 0, v_1, v_{\frac{1}{2}}, v_2\}. \quad (54)$$

From the present point of view this choice is arbitrary.

B. Thermodynamic Functions after Propagator Renormalization

There is a natural choice, however, when we also perform the Legendre transformation to eliminate v_1. In this case, the equation analogous to (3) is

$$F^{(1)}\{G_{\frac{1}{2}}, G_1\} = W + v_0 + G_{\frac{1}{2}}(1)v_{\frac{1}{2}}(1) + \tfrac{1}{2}G_1(12)v_1(12) = F^{(\frac{1}{2})}\{G_{\frac{1}{2}}, v_1\} + \tfrac{1}{2}G_1(12)v_1(12). \quad (55)$$

Corresponding to (4), we have

$$\delta F^{(1)}/\delta G_1(12) = \tfrac{1}{2}v_1(12), \quad (56)$$

and corresponding to (5) we have the statement that the functional

$$W \equiv F^{(1)}\{G_{\frac{1}{2}}, G_1\} - v_0 - G_{\frac{1}{2}}(1)v_{\frac{1}{2}}(1) - \tfrac{1}{2}G_1(12)v_1(12), \quad (57)$$

where $v_{\frac{1}{2}}$ and v_1 are fixed, is stationary with respect to changes in $G_{\frac{1}{2}}$ and G_1.

We compute $F^{(1)}$ by integrating (56), using (34),

$$\left(\frac{\delta F^{(1)}}{\delta G_1(12)}\right)_{G_{1/2}} = \left(\frac{\delta F^{(1)}}{\delta \tilde{G}_1(12)}\right)_{G_{1/2}} = \frac{1}{2}\left[-\sigma^\epsilon\delta(1\,2)\frac{\partial}{\partial z} + K_1(1\,2) + \tilde{G}_1^{-1}(1\,2)\right]$$

$$F^{(1)}\{G_{\frac{1}{2}}, \tilde{G}_1, v_{\frac{1}{2}}, v_2\} = \frac{\epsilon}{2}\operatorname{tr}[\ln \tilde{G}_1] - \frac{\epsilon}{2}\operatorname{tr}\left[\sigma^\epsilon\frac{\partial}{\partial z}\tilde{G}_1\right] + \frac{\epsilon}{2}\int_{\underline{\tilde{G}}_1}^{\tilde{G}_1} K_1\{12; \tilde{G}_1\}\ \delta\tilde{G}'_1(21) + \Phi^{(1)}\{G_{\frac{1}{2}}, \underline{\tilde{G}}_1, v_{\frac{1}{2}}\}. \quad (58)$$

The notation tr indicates summation over diagonal elements of a matrix whose indices are (i, j) or $(x_i, z_i; x_j, z_j)$. The matrix $\ln \tilde{G}_1$ is a matrix, the exponential of which has elements $\tilde{G}_1(i, j)$. It is simplest to take $\underline{\tilde{G}}_1 = 0$ since in that case the dependence of $\Phi^{(1)}$ on $v_{\frac{1}{2}}$ and v_2 is determined by using the identity

$$\left(\frac{\delta F^{(1)}}{\delta v(456)}\right)_{G_{1/2}, \tilde{G}_1, v_2} = \frac{1}{3!}G_{\frac{1}{2}}(456) = \frac{1}{2}\int_0^{\tilde{G}_1}\left(\frac{\delta K_1(12)}{\delta v_{\frac{1}{2}}(456)}\right)_{G_{1/2}, \tilde{G}_1, v_2}\delta\tilde{G}'_1(12) + \left(\frac{\delta\Phi^{(1)}}{\delta v_{\frac{1}{2}}(456)}\right)_{G_{1/2}, v_2},$$

and the equality of the cross derivatives

$$\frac{\delta}{\delta v_{\frac{1}{2}}(456)}\frac{\delta}{\delta\tilde{G}_1(12)}\left\{F^{(1)} - \frac{\epsilon}{2}\operatorname{tr}\left[\ln\tilde{G}_1 - \sigma^\epsilon\frac{\partial}{\delta z}\tilde{G}_1\right]\right\} = \frac{1}{2}\frac{\delta K_1(12)}{\delta v_{\frac{1}{2}}(456)} = \frac{1}{3!}\frac{\delta G_{\frac{1}{2}}(456)}{\delta\tilde{G}_1(12)},$$

to deduce that

$$\frac{\delta\Phi^{(1)}\{G_{\frac{1}{2}}, v_{\frac{1}{2}}, v_2\}}{\delta v_{\frac{1}{2}}(456)} = \frac{1}{3!}G_{\frac{1}{2}}\{4\ 5\ 6; G_{\frac{1}{2}}, \tilde{G}_1 = 0, v_{\frac{1}{2}}, v_2\} = \frac{1}{3!}G_{\frac{1}{2}}(4)G_{\frac{1}{2}}(5)G_{\frac{1}{2}}(6). \quad (59)$$

To obtain the last equality we have used (38) and (26). By precisely the same argument we also have

$$\frac{\delta\Phi^{(1)}}{\delta v_2(4567)} = \frac{1}{4!}G_{\frac{1}{2}}(4)G_{\frac{1}{2}}(5)G_{\frac{1}{2}}(6)G_{\frac{1}{2}}(7). \quad (60)$$

Consequently, we may write

$$F^{(1)}\{G_{\frac{1}{2}}, \tilde{G}_1, v_{\frac{1}{2}}, v_2\} = \frac{\epsilon}{2}\operatorname{tr}[\ln\tilde{G}_1] - \frac{\epsilon}{2}\operatorname{tr}\left[\sigma^\epsilon\frac{\partial}{\partial z}\tilde{G}_1\right] + \frac{1}{3!}G_{\frac{1}{2}}(1)G_{\frac{1}{2}}(2)G_{\frac{1}{2}}(3)v_{\frac{1}{2}}(123) + \frac{1}{4!}G_{\frac{1}{2}}(1)G_{\frac{1}{2}}(2)G_{\frac{1}{2}}(3)G_{\frac{1}{2}}(4)v_2(1234) + \frac{\epsilon}{2}\int_0^{\tilde{G}_1} K_1(12; \tilde{G}'_1)\ \delta\tilde{G}'_1(21) + \Phi\{G_{\frac{1}{2}}\}. \quad (61)$$

The function Φ is determined by considering a system in which $v_{\frac{1}{2}} = v_2 = 0$. In this case it follows, from (50) and (46), that

$$\delta\Phi/\delta G_{\frac{1}{2}} = 0.$$

Thus Φ is a constant whose value may be determined by taking account of the fact that, in a noncondensed, noninteracting system,

$$W = \tfrac{1}{2}\epsilon\operatorname{tr}\ln\tilde{G}_1 + f^{(0)} - v_0,$$

$$F^{(1)} = \tfrac{1}{2}\epsilon \operatorname{tr} \ln \tilde{G}_1 + \tfrac{1}{2}\epsilon \tilde{G}_1(12)v_1(21) + f^{(0)},$$

where $f^{(0)}$ is a constant evaluated in the Appendix (A8). We therefore have

$$\begin{aligned}\Phi &= \tfrac{1}{2}\epsilon[v_1(12) + \sigma^{\epsilon}\delta(12)\partial/\partial z_2]\tilde{G}_1(21) + f^{(0)} \\ &= \tfrac{1}{2}\epsilon \operatorname{tr} \delta(12) + f^{(0)}. \end{aligned} \tag{62}$$

The constant Φ, which is not relevant for the stationary properties, but is relevant for the evaluation of W, will be suppressed in the remainder of this discussion. From (61) we have the stationary expression in $\tilde{G}_1$ and $G_{\frac{1}{2}}$ for fixed $v_{\frac{1}{2}}$ and v_1,

$$\begin{aligned} W = &-v_0 - G_{\frac{1}{2}}(1)v_{\frac{1}{2}}(1) \\ &- \tfrac{1}{2}G_1(12)v_1(12) - \frac{\epsilon}{2}\operatorname{tr}\left[\sigma^{\epsilon}\frac{\partial}{\partial z}\tilde{G}_1\right] \\ &+ \frac{\epsilon}{2}\operatorname{tr}[\ln \tilde{G}] + \frac{1}{3!}G_{\frac{1}{2}}(1)G_{\frac{1}{2}}(2)G_{\frac{1}{2}}(3)v_{\frac{1}{2}}(123) \\ &+ \frac{1}{4!}G_{\frac{1}{2}}(1)G_{\frac{1}{2}}(2)G_{\frac{1}{2}}(3)G_{\frac{1}{2}}(4)v_2(1234) \\ &+ \frac{\epsilon}{2}\int_0^{\tilde{G}_1} \operatorname{tr}[K_1\{\tilde{G}_1'\}\ \delta\tilde{G}_1']. \end{aligned} \tag{63}$$

Expressions like (63) have previously been discussed for "nonanomalous" systems by several authors.[24,25] Several explicit additional terms are present in (63) because $G_{\frac{1}{2}} \neq 0$. Furthermore, the structure of K_1 is considerably more complicated for this reason. Of course, when there are "anomalous" propagators, but $v_{\frac{1}{2}} = v_{\frac{1}{2}} = 0$, the only additional terms remaining are

$$\begin{aligned} &-\tfrac{1}{2}G_{\frac{1}{2}}(1)G_{\frac{1}{2}}(2)v_1(12) \\ &\qquad + (1/4!)G_{\frac{1}{2}}(1)G_{\frac{1}{2}}(2)G_{\frac{1}{2}}(3)G_{\frac{1}{2}}(4)v_2(1234), \end{aligned}$$

and when the condensate is uniform, so that $G_{\frac{1}{2}}$ is a constant matrix whose two components are complex conjugate, the extra terms reduce to

$$\mu n_0 + (1/4!)G_{\frac{1}{2}}(1)G_{\frac{1}{2}}(2)G_{\frac{1}{2}}(3)G_{\frac{1}{2}}(4)v_2(1234),$$

where

$$|G_{\frac{1}{2}}(\alpha = 1)| = |G_{\frac{1}{2}}(\alpha = 2)| \equiv n_0^{\frac{1}{2}}$$

is the square root of the condensate density.

In Bose systems with $v_{\frac{1}{2}} = 0$ above the critical temperature, and in Fermi systems in which $v_{\frac{1}{2}}$, $v_{\frac{1}{2}}$, $G_{\frac{1}{2}}$ and $G_{\frac{1}{2}}$ vanish even if the system is superconducting, we have the stationary functional of G for fixed v_1 (and v_2),

$$\begin{aligned} W = &-v_0 - \tfrac{1}{2}\epsilon \operatorname{tr}([G_1^0]^{-1}\tilde{G}_1) + \tfrac{1}{2}\epsilon \operatorname{tr}[\ln \tilde{G}_1] \\ &+ \int_0^{\tilde{G}_1} \operatorname{tr}[K_1\{\tilde{G}_1'\}\ \delta\tilde{G}_1']. \end{aligned} \tag{63'}$$

[24] J. Luttinger and J. Ward, Phys. Rev. **118**, 1417 (1960).
[25] G. Baym, Phys. Rev. **127**, 1391 (1962).

When the system is superconducting the diagonal elements of G_1 play an essential role.[26]

VI. VARIATIONAL EXPRESSION FOR ENTROPY: DEPENDENCE ON VERTEX FUNCTIONS

In the previous section, expressions for the thermodynamical potentials (and the correlation functions) in terms of the renormalized propagator and condensate wavefunction have been derived. These expressions, (61) and (63), in which a term in $v_2G_{\frac{1}{2}}^4$ occurs explicitly, make particularly apparent the desirability of a further (vertex) renormalization. The term $v_2G_{\frac{1}{2}}^4$ is large, for example, for an interaction potential v_2 which has a (small) repulsive core, and its contribution must then be canceled by terms arising from $\int K_1\ \delta\tilde{G}_1$. Difficulties of this type are overcome, and improved convergence may be more generally anticipated when renormalized vertices are introduced by the procedure outlined in the introduction.

A. Variational Expression for the Entropy

We note that if we Legendre transform completely, we have

$$\begin{aligned} F^{(2)}\{G_\nu\} &= W \\ &+ \sum_{2\nu=0}^{4} \frac{1}{(2\nu)!}G_\nu(1, 2, \cdots, 2\nu)v_\nu(1, 2, \cdots 2\nu), \\ \delta F^{(2)} &= \sum \frac{1}{(2\nu)!}\delta G_\nu(1, 2, \cdots 2\nu) \\ &\qquad \times v_\nu(1, 2, \cdots 2\nu). \end{aligned} \tag{64}$$

As in each of the previous cases, we have a variational principle, the present one stating that for fixed values of v_ν, the functional

$$\begin{aligned} W \equiv &F^{(2)}\{G_\nu\} \\ &- \sum \frac{1}{(2\nu)!}G_\nu(1, 2, \cdots, 2\nu)v_\nu(1, 2, \cdots 2\nu) \end{aligned} \tag{65}$$

is stationary in the variables G_ν. This variational principle may be viewed alternatively as a stationarity condition on $F^{(2)}$, subject to constraint conditions on the variables G_ν. Since the quantity $\sum [(2\nu)!]^{-1}G_\nu v_\nu$ is equal to $\beta(E - \mu N)$, where E and N are the average energy and particle number, we see that *the functional, $F^{(2)}$, of the distribution functions G_ν, is stationary subject to the constraint of a prescribed energy and particle number, and its stationary value is the entropy.*

In order to construct $F^{(2)}$ we must integrate (64),

[26] These are the off-diagonal elements in the conventional representation.

as previously we integrated (52) and (56). For this purpose we must know the dependence of the potentials on the distribution functions. We proceed in two steps. First we prove the statement we previously made, namely, that $F^{(2)}$ depends (in a complicated way) on only two combinations of the four distribution functions (combinations closely related to the renormalized vertices). On the basis of that statement, we formally carry out the functional integration of $F^{(2)}$ (Eq. 70) in terms of the assumed known dependence of the renormalized vertices on the unrenormalized ones. We also indicate some implications of this intrinsic dependence on two variables for the previously constructed functional $F^{(1)}$. Then in the following section, (Sec. VII) we use this dependence on two variables to carry out the second step. We construct differential equations which explicitly determine the unrenormalized vertex functions and hence the entropy $F^{(2)}$ in terms of the renormalized vertex functions.

B. Dependence of Thermodynamical Functions on Two Vertex Functions

To demonstate that $F^{(2)}$ depends on only two variables we use Eqs. (31) and (34), which correspond to the first two equations of the hierarchy. Substituting for $v_{\frac{1}{2}}$ and v_1 in (64), we have

$$\delta F^{(2)} = \delta\left[-\tfrac{1}{2}\epsilon \operatorname{tr} \sigma^{\epsilon} \frac{\partial}{\partial z} G_1 + \tfrac{1}{2}\epsilon \operatorname{tr} \ln \tilde{G}_1 \right] + \delta X, \tag{66}$$

$$\delta X \equiv [\delta G_{\frac{1}{2}}(1)]K_{\frac{1}{2}}(1) + \tfrac{1}{2}[\delta G_1(12)]K_1(12) + \frac{1}{3!}[\delta G_{\frac{3}{2}}(123)]v_{\frac{3}{2}}(123) + \frac{1}{4!}[\delta G_2(1234)]v_2(1234).$$

Using the relations (26) and (38), between the distribution functions and the vertices, we also have

$$\begin{aligned}
\delta G_{\frac{1}{2}} &= \delta[C_{\frac{1}{2}}\tilde{G}_1\tilde{G}_1\tilde{G}_1] \\
&\quad + \tfrac{1}{2}\sum_P \epsilon^P [G_{\frac{1}{2}}G_{\frac{1}{2}}\delta G_{\frac{1}{2}} + \tilde{G}_1\delta\tilde{G}_{\frac{1}{2}} + G_{\frac{1}{2}}\delta\tilde{G}_1], \\
\delta G_2 &= \delta[(C_2 + 3C_{\frac{1}{2}}\tilde{G}_1C_{\frac{1}{2}})\tilde{G}_1\tilde{G}_1\tilde{G}_1\tilde{G}_1] \\
&\quad + \frac{1}{4!}\sum_P \epsilon^P [4G_{\frac{1}{2}}G_{\frac{1}{2}}G_{\frac{1}{2}}\delta G_{\frac{1}{2}} \\
&\quad + 4\delta G_{\frac{1}{2}}\tilde{G}_{\frac{1}{2}} + 4G_{\frac{1}{2}}\delta(C_{\frac{1}{2}}\tilde{G}_1\tilde{G}_1\tilde{G}_1) \\
&\quad + 6G_{\frac{1}{2}}G_{\frac{1}{2}}\delta\tilde{G}_1 + 6\tilde{G}_1\delta\tilde{G}_1 + 12\delta G_{\frac{1}{2}}G_{\frac{1}{2}}\tilde{G}_1].
\end{aligned} \tag{67}$$

If we substitute these expressions into (66) and use the expression for the source (40), we find that *all the terms in* (66) *involving* $\delta G_{\frac{1}{2}}$ *vanish.*[27] Furthermore, the terms from Eq. (39) for K_1 and the terms from (67) containing $\delta\tilde{G}_1$ are almost compensatory. The only terms which remain when we make these substitutions and write the resulting form of (66) in a more symmetrical fashion are the terms

$$\begin{aligned}
\delta X &= \frac{1}{3!}\,\delta\{\tilde{G}_{\frac{3}{2}}(1\,\bar{1}\,2\,\bar{2}\,3\,\bar{3})\tilde{G}^{-\frac{1}{2}}(1\,\bar{1})\tilde{G}^{-\frac{1}{2}}(2\,\bar{2})\tilde{G}^{-\frac{1}{2}}(3\,\bar{3})\} \\
&\quad \times \{\tilde{G}^{\frac{1}{2}}(\bar{1}\,1)\tilde{G}^{\frac{1}{2}}(\bar{2}\,2)\tilde{G}^{\frac{1}{2}}(\bar{3}\,3)[v_{\frac{3}{2}}(123) + v_2(1234)G_{\frac{1}{2}}(4)]\} \\
&\quad + \frac{1}{4!}\,\delta\{\tilde{G}_2(1\,\bar{1}\,2\,\bar{2}\,3\,\bar{3}\,4\,\bar{4})\tilde{G}^{-\frac{1}{2}}(1\,\bar{1})\tilde{G}^{-\frac{1}{2}}(2\,\bar{2})\tilde{G}^{-\frac{1}{2}}(3\,\bar{3})\tilde{G}^{-\frac{1}{2}}(4\,\bar{4})\} \\
&\quad \times [\tilde{G}^{\frac{1}{2}}(\bar{1}\,1)\tilde{G}^{\frac{1}{2}}(\bar{2}\,2)\tilde{G}^{\frac{1}{2}}(\bar{3}\,3)\tilde{G}^{\frac{1}{2}}(\bar{4}\,4)v_2(1234)].
\end{aligned} \tag{68}$$

This equality demonstrates the important point that, apart from the "free-particle" terms, *the thermodynamical function* $F^{(2)}$ *really depends only on two variables*; that is,

$$\begin{aligned}
&\delta F^{(2)}\{G_{\frac{1}{2}}, G_1, G_{\frac{3}{2}}, G_2\} \\
&\quad = \delta[\tfrac{1}{2}\epsilon \operatorname{tr}(-\sigma^{\epsilon}(\partial/\partial z)G_1 + \ln \tilde{G}_1)] \\
&\qquad + \delta X\{\tilde{G}_{\frac{3}{2}}(\tilde{G}_1\tilde{G}_1\tilde{G}_1)^{-\frac{1}{2}}, \tilde{G}_2(\tilde{G}_1\tilde{G}_1\tilde{G}_1\tilde{G}_1)^{-\frac{1}{2}}\}.
\end{aligned} \tag{69}$$

Loosely speaking, this feature corresponds to the statement that in a diagrammatic description there are only two "structures" from which diagrams are built, namely, vertices with three or four emerging lines. This diagrammatic description is discussed in detail in the following paper.

C. Implications for the Thermodynamical Functions

To determine the functional $F^{(2)}$ we integrate Eq. (69) using (68). The resulting expression is

$$\begin{aligned}
F^{(2)} &= -\tfrac{1}{2}\epsilon \operatorname{tr}\left[\sigma^{\epsilon}\frac{\partial}{\partial z}G_1\right] + \tfrac{1}{2}\epsilon \operatorname{tr}[\ln \tilde{G}_1] \\
&\quad + \frac{1}{3!}\int_{\bar{C}_{\frac{3}{2}}=0}^{\bar{C}_{\frac{3}{2}}} \delta\bar{C}'_{\frac{3}{2}}(123)\bar{v}_{\frac{3}{2}}\{1\,2\,3;\, \bar{C}'_{\frac{3}{2}}, \bar{E}_2 = 0\} \\
&\quad + \frac{1}{4!}\int_{\bar{E}_2=0}^{\bar{E}_2} \delta\bar{E}'_2(1234)\bar{v}_2\{1234;\, \bar{C}_{\frac{3}{2}}, \bar{E}'_2\}
\end{aligned} \tag{70}$$

[plus the numerical constant of integration given by (62)].

In this expression we have introduced an abbreviated notation for the intrinsic variables

$$\begin{aligned}
\bar{v}_{\frac{3}{2}}(123) &= \tilde{G}^{\frac{1}{2}}(1\,\bar{1})\tilde{G}^{\frac{1}{2}}(2\,\bar{2})\tilde{G}^{\frac{1}{2}}(3\,\bar{3}) \\
&\qquad \times [v_{\frac{3}{2}}(\bar{1}\,\bar{2}\,\bar{3}) + v_2(\bar{1}\,\bar{2}\,\bar{3}\,\bar{4})G(\bar{4})], \\
\bar{v}_2(1234) &= \tilde{G}^{\frac{1}{2}}(1\,\bar{1})\tilde{G}^{\frac{1}{2}}(2\,\bar{2})\tilde{G}^{\frac{1}{2}}(3\,\bar{3})\tilde{G}^{\frac{1}{2}}(4\,\bar{4})v_2(\bar{1}\,\bar{2}\,\bar{3}\,\bar{4}), \\
\bar{C}_{\frac{3}{2}}(123) &= C_{\frac{3}{2}}(\bar{1}\,\bar{2}\,\bar{3})\tilde{G}^{\frac{1}{2}}(\bar{1}\,1)\tilde{G}^{\frac{1}{2}}(\bar{2}\,2)\tilde{G}^{\frac{1}{2}}(\bar{3}\,3) \\
&= \tilde{G}_{\frac{3}{2}}(\bar{1}\,\bar{2}\,\bar{3})\tilde{G}^{-\frac{1}{2}}(\bar{1}\,1)\tilde{G}^{-\frac{1}{2}}(\bar{2}\,2)\tilde{G}^{-\frac{1}{2}}(\bar{3}\,3),
\end{aligned}$$

[27] We remark that this statement is closely connected with the fact that the superfluid flow (which is described by $\nabla G_{\frac{1}{2}}$) is reversible.

$$\begin{aligned}\bar{E}_2(1234) &= \tilde{G}_2(\bar{1}\,\bar{2}\,\bar{3}\,\bar{4})\tilde{G}^{-\frac{1}{2}}(\bar{1}\,1)\tilde{G}^{-\frac{1}{2}}(\bar{2}\,2)\tilde{G}^{-\frac{1}{2}}(\bar{3}\,3)\tilde{G}^{-\frac{1}{2}}(\bar{4}\,4)\\ &= \bar{C}_2(1234) + \bar{C}_{\frac{1}{2}}(125)\bar{C}_{\frac{1}{2}}(543)\\ &\quad + \bar{C}_{\frac{1}{2}}(135)\bar{C}_{\frac{1}{2}}(524) + \bar{C}_{\frac{1}{2}}(145)\bar{C}_{\frac{1}{2}}(532)\\ &\equiv E_2(\bar{1}\,\bar{2}\,\bar{3}\,\bar{4})\tilde{G}^{\frac{1}{2}}(\bar{1}\,1)\tilde{G}^{\frac{1}{2}}(\bar{2}\,2)\tilde{G}^{\frac{1}{2}}(\bar{3}\,3)\tilde{G}^{\frac{1}{2}}(\bar{4}\,4),\\ \bar{C}_2(1234) &= \bar{C}_2(\bar{1}\,\bar{2}\,\bar{3}\,\bar{4})\tilde{G}^{\frac{1}{2}}(\bar{1}\,1)\tilde{G}^{\frac{1}{2}}(\bar{2}\,2)\tilde{G}^{\frac{1}{2}}(\bar{3}\,3)\tilde{G}^{\frac{1}{2}}(\bar{4}\,4),\\ \bar{D}_2(12;34)&\\ &= D_2(\bar{1}\,\bar{2};\bar{3}\,\bar{4})\tilde{G}^{\frac{1}{2}}(\bar{1}\,1)\tilde{G}^{\frac{1}{2}}(\bar{2}\,2)\tilde{G}^{\frac{1}{2}}(\bar{3}\,3)\tilde{G}^{\frac{1}{2}}(\bar{4}\,4).\end{aligned} \tag{71}$$

In the following paper, Eq. (70) will be obtained by a different method in which one goes through the intermediary of a functional $F^{(\frac{1}{2})}$ from which one may also obtain a stationary principle for W in the functions $G_{\frac{1}{2}}$, G_1, $G_{\frac{3}{2}}$. From the present point of view, $F^{(\frac{1}{2})}$ is rather artificial, in the same way as $F^{(\frac{1}{2})}$ was. If we are intent on obtaining it, it is useful to take advantage of the information contained in Eq. (69), about the intrinsic dependence of $F^{(2)}$ on only two variables. That is to say, it is useful to perform an inverse Legendre transformation on $F^{(2)}$ instead of performing a Legendre transformation on $F^{(1)}$. One obtains in this manner a form depending in a complicated way on the two variables $\bar{C}_{\frac{1}{2}}$ and $\bar{v}_2$, and in addition, a few explicit terms which reflect the difference between v_2 and $\bar{v}_2$. Indeed, by performing two Legendre transformations from the correlation functions $G_{\frac{1}{2}}$ and G_2 to the potentials $v_{\frac{1}{2}}$ and v_2 we may deduce the fact that, apart from the explicit terms resulting from the difference between $\bar{v}_2$, $\bar{v}_{\frac{1}{2}}$ and v_2, $v_{\frac{1}{2}}$,

$$-\mathfrak{F}^{(1)} \equiv \frac{1}{4!}v_2G_{\frac{1}{2}}^4 + \tfrac{1}{4}v_2G_{\frac{1}{2}}^2\tilde{G}_1 + \tfrac{1}{8}v_2\tilde{G}_1\tilde{G}_1 + \frac{1}{3!}v_{\frac{1}{2}}G_{\frac{1}{2}}^3 + \tfrac{1}{2}v_{\frac{1}{2}}G_{\frac{1}{2}}\tilde{G}_1$$

and the "free-particle" terms, the functional $F^{(1)}$ exhibited in (61) is a functional of $G_{\frac{1}{2}}$, $\tilde{G}_1$ $v_{\frac{1}{2}}$ and v_2 only through their occurence in the "true" potential matrix elements $\bar{v}_{\frac{1}{2}}$ and $\bar{v}_2$. (The name "true" potential matrix element is not an inappropriate one since $\tilde{G}^{\frac{1}{2}}$ bears a similarity to the "true" wavefunction).

The fact that the remainder of $F^{(1)}$ involves only the two functions $\bar{v}_{\frac{1}{2}}$ and $\bar{v}_2$ implies conversely that

$$\left(\frac{\delta(F^{(1)} - \mathfrak{F}^{(1)})}{\delta G_{\frac{1}{2}}(1)}\right)_{\tilde{G}_1,v_{\frac{1}{2}},v_2} = \left(\frac{\delta(F^{(1)} - \mathfrak{F}^{(1)})}{\delta v_{\frac{1}{2}}(234)}\right)_{\tilde{G}_1,G_{\frac{1}{2}},v_2} \times v_2(2341) = -\frac{1}{3!}\tilde{G}_{\frac{1}{2}}(234)v_2(2341),$$

which reproduces Eq. (32).

VII. RELATIONSHIP BETWEEN RENORMALIZED AND UNRENORMALIZED VERTEX FUNCTIONS: FULLY RENORMALIZED FORMULATION

To obtain an equation for the fully renormalized functional $F^{(2)}$ we must incorporate the fact that it depends intrinsically on only two variables. For this purpose we derive the nonlinear functional differential equations which relate the pairs of variables $(\bar{v}_2, \bar{v}_{\frac{1}{2}})$, $(\bar{v}_2,\bar{C}_{\frac{1}{2}})$, and $(\bar{E}_2, \bar{C}_{\frac{1}{2}})$, to one another.

We first observe that, from the equality of the cross derivatives in (66), the identity (41), and Eq. (40) for $K_{\frac{1}{2}}$,

$$\begin{aligned}\left(\frac{\delta K_1(21)}{\delta G_{\frac{1}{2}}(3)}\right)_{\tilde{G}} &= 2\left(\frac{\delta K_{\frac{1}{2}}(3)}{\delta\tilde{G}(12)}\right)_{G_{1/2}}\\ &= C(123) - \left(\frac{\delta K(12)}{\delta\tilde{G}(45)}\right)_{G_{1/2}}\tilde{G}(44)C(\bar{4}\,\bar{5}\,3)\tilde{G}(\bar{5}5)\\ &= [-v(123) - v(1234)G(4)] - \tfrac{1}{3}v(3\,\bar{2}\,\bar{3}\,\bar{4})\\ &\quad\times\left(\frac{\delta}{\delta\tilde{G}(12)}[C(\bar{\bar{2}}\,\bar{\bar{3}}\,\bar{\bar{4}})\tilde{G}(\bar{\bar{2}}\,\bar{2})\tilde{G}(\bar{\bar{3}}\,3)\tilde{G}(\bar{\bar{4}}\,4)]\right)_{G_{1/2}}.\end{aligned}$$

at constant $v_{\frac{1}{2}}$ and v_2. In virtue of (44) we obtain

$$\begin{aligned}&-v(123) - v(1234)G(4)\\ &= [1 + \tfrac{1}{2}DGG]^{-1}(12;\bar{1}\,\bar{2})C(\bar{1}\,\bar{2}\,3)\\ &\quad + \tfrac{1}{3}v(3456)\frac{\delta}{\delta\tilde{G}(12)}[C(\bar{4}\,\bar{5}\,\bar{6})\tilde{G}(\bar{4}\,4)\tilde{G}(\bar{5}\,5)\tilde{G}(\bar{6}\,6)].\end{aligned} \tag{72}$$

Similarly, by using (39) and (44), we find the relation

$$\begin{aligned}2\left(\frac{\delta K(12)}{\delta\tilde{G}(34)}\right)_{G_{1/2}} &= \left[\frac{1}{1 + \frac{1}{2}DGG}D\right](12;34)\\ &= -v(1234) - [v(1\,\bar{2}\,\bar{3}\,\bar{4})G(\bar{4}) + v(1\,\bar{2}\,\bar{3})]\\ &\quad\times\frac{\delta}{\delta\tilde{G}(34)}[C(562)\tilde{G}(5\,\bar{3})\tilde{G}(6\,\bar{2})]\\ &\quad - \tfrac{1}{3}v(1\,\bar{2}\,\bar{3}\,\bar{4})\frac{\delta}{\delta\tilde{G}(34)}[E(5672)\tilde{G}(5\,\bar{2})\tilde{G}(6\,\bar{3})\tilde{G}(7\,\bar{4})].\end{aligned}$$

Substituting (72) into this expression, we obtain the equation

$$\begin{aligned}&[(1 + \tfrac{1}{2}DGG)^{-1}D](12;34)\\ &- (1 + \tfrac{1}{2}DGG)^{-1}(1\bar{2};56)C(56\bar{3})\\ &\times\frac{\delta}{\delta\tilde{G}(34)}[C(892)\tilde{G}(8\bar{3})\tilde{G}(9\bar{2})]_{G_{1/2}} = -v(1234)\\ &+ \tfrac{1}{3}v(\bar{3}567)\frac{\delta}{\delta\tilde{G}(1\bar{2})}[C(\bar{5}\,\bar{6}\,\bar{7})\tilde{G}(\bar{5}5)\tilde{G}(\bar{6}6)\tilde{G}(\bar{7}7)]_{G_{1/2}}\\ &\times\frac{\delta}{\delta\tilde{G}(34)}[C(892)\tilde{G}(8\bar{3})\tilde{G}(9\bar{2})]_{G_{1/2}} - \tfrac{1}{3}v(1\,\bar{2}\,\bar{3}\,\bar{4})\end{aligned}$$

$$\times \frac{\delta}{\delta\tilde{G}(34)} [E(5672)\tilde{G}(5\bar{2})\tilde{G}(6\bar{3})\tilde{G}(7\bar{4})]_{G_{1/2}}, \tag{73}$$

which determines the potential v_2 as a functional of the distribution functions. Together with (73), Eq. (72) determines $v_{\frac{1}{2}}$ as a functional of the distribution functions.

We next introduce the information that there are only two intrinsic variables that is, that $\bar{C}_{\frac{1}{2}}$ and $\bar{E}_2$ are functionals only of $\bar{v}_{\frac{1}{2}}$ and $\bar{v}_2$. Specifically we use this fact to reexpress the derivatives in (72) and (73) in terms of the functions $\bar{C}_{\frac{1}{2}}$ and $\bar{E}_{\frac{1}{2}}$ with the aid of the relation

$$\left(\frac{\delta}{\delta\tilde{G}_1(12)}\right)_{G_{1/2},v_{3/2},v_2} = \left(\frac{\delta\bar{v}_{\frac{1}{2}}(345)}{\delta\tilde{G}_1(12)}\right)_{G_{3/2},v_{2},v_{3/2}} \left(\frac{\delta}{\delta\bar{v}_{\frac{1}{2}}(345)}\right)_{\bar{v}_2} + \left(\frac{\delta\bar{v}_2(3456)}{\delta\tilde{G}_1(12)}\right)_{G_{1/2},v_2,v_{3/2}} \left(\frac{\delta}{\delta\bar{v}_2(3456)}\right)_{\bar{v}_{3/2}}. \tag{74}$$

If, therefore we express the functions C and E in terms of $\tilde{G}_1$, $\bar{C}$, and $\bar{E}$, we may rewrite (72) and (73) directly as relations between the true potential matrix elements $\bar{v}_2$, $\bar{v}_{\frac{1}{2}}$ and the correlation matrices $\bar{C}_{\frac{1}{2}}$, $\bar{E}_2$.

We first consider systems in which $v_{\frac{1}{2}}$, $G_{\frac{1}{2}}$, and $G_{\frac{3}{2}}$ vanish and (72) disappears. In an extremely symbolic form, meant only for illustrating the differential structure and nonlinearity, and which does not even indicate the difference between terms with the same structure but different indices, Eq. (73) becomes

$$\tfrac{1}{2}\bar{C}_2/(1 + \tfrac{1}{2}\bar{C}_2) = -\tfrac{1}{2}\bar{v}_2 - (1/3!)\bar{v}_2\bar{C}_2 - \tfrac{1}{3}\bar{v}_2(\delta\bar{C}_2/\delta\bar{v}_2)\bar{v}_2. \tag{75}$$

This equation illustrates a natural and physical separation of the problem. In a first step, Eq. (75) is employed to determine the correlations $\bar{C}$ produced by the matrix elements $\bar{v}$, the $\bar{v}$'s depending on the density and excitation spectrum through the factors $\tilde{G}_1^{\frac{1}{2}}$. In a second step the function $\tilde{G}_1$ is determined from Eqs. (34) and (39), that is,

$$\tilde{G}^{\frac{1}{2}}[G_1^0]^{-1}\tilde{G}^{\frac{1}{2}} = 1 - \tfrac{1}{2}\bar{v}_2 - (1/3!)\bar{v}_2\bar{C}_2, \tag{76}$$

or, more explicitly,

$$\tilde{G}_1^{\frac{1}{2}}(1\bar{1})[G_1^0]^{-1}(\bar{1}\,\bar{1}')\tilde{G}_1^{\frac{1}{2}}(\bar{1}'1') = \delta(11') - \tfrac{1}{2}\bar{v}_2(11'22) - (1/3!)\bar{v}_2(1234)\bar{C}_2(2341').$$

Finally, the matrices $\bar{v}$ must be made self-consistent with the potential v and the calculated $\tilde{G}_1$. In the self-consistency scheme, physical information can be reasonably inserted. For example, if the potential has both a hard core in relation to which the interparticle spacing is large, and a weak long-range potential, then the third term in (75) can be treated as a perturbation.

Equation (75) can finally be rewritten in the form of a differential equation for the entropy functional by invoking (70), that is, we may write

$$\frac{1}{4}\frac{\tfrac{1}{2}\bar{C}_2}{1 + \tfrac{1}{2}\bar{C}_2} = -3\frac{\partial F^{(2)}}{\partial\bar{C}_2} - \frac{\partial F^{(2)}}{\partial\bar{C}_2}\bar{C}_2 - 2\frac{\partial F^{(2)}}{\partial\bar{C}_2}\left(\frac{\partial^2 F^{(2)}}{\partial\bar{C}_2\,\partial\bar{C}_2}\right)^{-1}\frac{\partial F^{(2)}}{\partial\bar{C}_2}. \tag{77}$$

This equation may be used to generate the power series expansion for $F^{(2)}$ in terms of $\bar{C}_2$, an expansion whose nth term is diagramatically characterized in the following paper. With this substitution, Eq. (76) may be written symmetrically in the alternative form

$$\tilde{G}_1^{\frac{1}{2}}[G_1^0]^{-1}\tilde{G}_1^{\frac{1}{2}} = 1 + \frac{\tfrac{1}{2}\bar{C}}{1 + \tfrac{1}{2}\bar{C}} + 8\frac{\partial F^{(2)}}{\partial\bar{C}_2}\left(\frac{\partial^2 F^{(2)}}{\partial\bar{C}_2\,\partial\bar{C}_2}\right)^{-1}\frac{\partial F^{(2)}}{\partial\bar{C}_2}. \tag{78}$$

In general, Eqs. (72) and (73) are more complicated since the condensate $G_{\frac{1}{2}}$, the three-point distribution function $G_{\frac{3}{2}}$, and the three-point vertex ($\bar{v}_{\frac{3}{2}} = v_2 G_{\frac{3}{2}}(\tilde{G}\tilde{G}\tilde{G})^{\frac{1}{2}}$) do not vanish. Again in symbolic form,

$$\begin{aligned}
-\bar{v}_{\frac{1}{2}} &= [1 + \tfrac{1}{2}\bar{D}_2]^{-1}\bar{C}_{\frac{1}{2}} + \tfrac{1}{2}\bar{v}_2\left(\frac{\delta\bar{C}_{\frac{1}{2}}}{\delta\bar{v}_{\frac{1}{2}}}\right)\bar{v}_{\frac{1}{2}} + \tfrac{2}{3}\bar{v}_2\left(\frac{\delta\bar{C}_{\frac{1}{2}}}{\delta\bar{v}_2}\right)\bar{v}_2 + \tfrac{1}{2}\bar{v}_2\bar{C}_{\frac{1}{2}},\\
-\bar{v}_2 &= [1 + \tfrac{1}{2}\bar{D}_2]^{-1}\bar{D}_2 + \tfrac{1}{2}\bar{v}_{\frac{1}{2}}\bar{C}_{\frac{1}{2}} + \tfrac{1}{3}\bar{v}_2\bar{E}_2 + \tfrac{3}{2}\bar{v}_{\frac{1}{2}}\left(\frac{\delta\bar{C}_{\frac{1}{2}}}{\delta\bar{v}_{\frac{1}{2}}}\right)\bar{v}_{\frac{1}{2}} + 2\bar{v}_{\frac{1}{2}}\left(\frac{\delta\bar{C}_{\frac{1}{2}}}{\delta\bar{v}_2}\right)\bar{v}_2 + \tfrac{1}{2}\bar{v}_2\left(\frac{\delta\bar{E}_2}{\delta\bar{v}_{\frac{1}{2}}}\right)\bar{v}_{\frac{1}{2}} + \tfrac{2}{3}\bar{v}_2\left(\frac{\delta\bar{E}_2}{\delta\bar{v}_2}\right)\bar{v}_2,\\
\bar{E}_2 &= \bar{D}_2 + \bar{C}_{\frac{1}{2}}^2.
\end{aligned} \tag{79}$$

The counterparts of (76)–(78) take a similarly complicated form and shall not be exhibited here.

We finally remark that, while the above equations amount essentially to an expansion of the vertices $C_{5/2}$ and C_3 in powers of the vertices $C_{3/2}$ and C_2, one can similarly express all other vertices in terms of $C_{3/2}$ and C_2, which completes the renormalization procedure.

Equation (75) or (77) for noncondensed systems and (79) for condensed systems are equations which determine the unrenormalized vertices in terms of the renormalized ones and conversely. However, they are so nonlinear that their power series solution is not easily generated.

In the following paper we shall derive by an alternative method a diagrammatic characterization of all terms in every order of the power series solution. We have not attempted to obtain this characterization from the differential equations or the converse.[28] Indeed we have not even carried out this comparison to general order for uncondensed systems for which we have the differential equation (75). We have, however, as a check on the two expressions, compared the first few orders of $\bar{v}$ as a functional of $\bar{C}$, generated by (75), against the power series derived in the following paper.

To summarize, we have obtained functional differential equations for the potentials v_ν in terms of the distribution functions $G_{\nu'}(\nu' \leq \nu)$ and the potentials v_ν $(\nu' > \nu)$, and in terms of the distribution functions alone. These equations correspond to calculational schemes involving successively more renormalized interactions. We have, in particular, exhibited the structure of these renormalized equations for a superfluid system in which the Hamiltonian contains no interactions involving more than two particles. Finally, by functionally integrating, we have obtained stationary expressions for the thermodynamical potentials corresponding to these schemes.

APPENDIX

In this appendix we list the results of the transformation from the original Hamiltonian to the ϵ-symmetrized form. For this purpose it is necessary to employ the commutation relations diligently, and thereby to express products of operators in terms of symmetrized products of operators. Because the commutator (or anticommutator) is a c number, this process poses no difficulty. We also calculate the constant of integration $f^{(0)}$ used in Secs. V and VI.

(1) If the original Hamiltonian contains the potentials

$$\begin{gathered}\hat{w}_{\frac{1}{2}}(\mathbf{r}_1, \alpha_1), \\ \hat{w}_1(\mathbf{r}_1, \alpha_1; \mathbf{r}_2, \alpha_2), \\ \hat{w}_{\frac{3}{2}}(\mathbf{r}_1, \alpha_1; \mathbf{r}_2, \alpha_2; \mathbf{r}_3, \alpha_3), \\ \hat{w}_2(\mathbf{r}_1, \alpha_1; \mathbf{r}_2, \alpha_2; \mathbf{r}_3, \alpha_3; \mathbf{r}_4, \alpha_4),\end{gathered} \tag{A1}$$

then the symmetrized potentials are

$$\begin{aligned}\hat{v}_0 &= \tfrac{1}{2}\hat{w}_1(\mathbf{r}_1, \alpha_1; \mathbf{r}_1, \alpha_2)(\tau^\epsilon)_{\alpha_1\alpha_2} + \tfrac{1}{4}(\tau^\epsilon)_{\alpha_1\alpha_2}(\tau^\epsilon)_{\alpha_3\alpha_4} \\ &\quad \times [\hat{w}_2(\mathbf{r}_1, \alpha_1; \mathbf{r}_1, \alpha_2; \mathbf{r}_3, \alpha_3; \mathbf{r}_3, \alpha_4) \\ &\quad + \hat{w}_2(\mathbf{r}_1, \alpha_1; \mathbf{r}_3, \alpha_3; \mathbf{r}_3, \alpha_4; \mathbf{r}_1, \alpha_2) \\ &\quad - \hat{w}_2(\mathbf{r}_1, \alpha_1; \mathbf{r}_3, \alpha_4; \mathbf{r}_1, \alpha_2; \mathbf{r}_3, \alpha_3)], \\ \hat{v}_{\frac{1}{2}}(\mathbf{r}_1, \alpha_1) &= \hat{w}_{\frac{1}{2}}(\mathbf{r}_1, \alpha_1) + \tfrac{1}{2}(\tau^\epsilon)_{\alpha_2\alpha_3} \\ &\quad \times [\hat{w}_{\frac{3}{2}}(\mathbf{r}_1, \alpha_1; \mathbf{r}_2, \alpha_2; \mathbf{r}_2, \alpha_3) \\ &\quad + \hat{w}_{\frac{3}{2}}(\mathbf{r}_2, \alpha_2; \mathbf{r}_2, \alpha_3; \mathbf{r}_1, \alpha_1) \\ &\quad + \hat{w}_{\frac{3}{2}}(\mathbf{r}_2, \alpha_3; \mathbf{r}_1, \alpha_1; \mathbf{r}_2, \alpha_2)], \\ \hat{v}_1(\mathbf{r}_1, \alpha_1; \mathbf{r}_2, \alpha_2) &= \hat{w}_1(\mathbf{r}_1, \alpha_1; \mathbf{r}_2, \alpha_2) \\ &\quad + \epsilon\hat{w}_1(\mathbf{r}_2, \alpha_2; \mathbf{r}_1, \alpha_1) + \tfrac{1}{2}(\tau^\epsilon)_{\alpha_3\alpha_4} \\ &\quad \times [\hat{w}_2(\mathbf{r}_3, \alpha_3; \mathbf{r}_3, \alpha_4; \mathbf{r}_1, \alpha_1; \mathbf{r}_2, \alpha_2) \\ &\quad + \hat{w}_2(\mathbf{r}_1, \alpha_1; \mathbf{r}_3, \alpha_3; \mathbf{r}_3, \alpha_4; \mathbf{r}_2, \alpha_2) \\ &\quad + \hat{w}_2(\mathbf{r}_1, \alpha_1; \mathbf{r}_2, \alpha_2; \mathbf{r}_3, \alpha_3; \mathbf{r}_3, \alpha_4) \\ &\quad + \hat{w}_2(\mathbf{r}_3, \alpha_3; \mathbf{r}_2, \alpha_2; \mathbf{r}_3, \alpha_4; \mathbf{r}_1, \alpha_1) \\ &\quad + \hat{w}_2(\mathbf{r}_2, \alpha_2; \mathbf{r}_3, \alpha_3; \mathbf{r}_1, \alpha_1; \mathbf{r}_3, \alpha_4) \\ &\quad + \hat{w}_2(\mathbf{r}_3, \alpha_3; \mathbf{r}_1, \alpha_1; \mathbf{r}_2, \alpha_2; \mathbf{r}_3, \alpha_4)], \\ \hat{v}_{\frac{3}{2}}(\mathbf{r}_1, \alpha_1; \mathbf{r}_2, \alpha_2; \mathbf{r}_3, \alpha_3) &= \sum_P (\epsilon)^P \hat{w}_{\frac{3}{2}}(\mathbf{r}_1, \alpha_1; \mathbf{r}_2, \alpha_2; \mathbf{r}_3, \alpha_3), \\ \hat{v}_2(\mathbf{r}_1, \alpha_1; \mathbf{r}_2, \alpha_2; \mathbf{r}_3, \alpha_3; \mathbf{r}_4, \alpha_4) &= \sum_P (\epsilon)^P \hat{w}_2(\mathbf{r}_1, \alpha_1; \mathbf{r}_2, \alpha_2; \mathbf{r}_3, \alpha_3; \mathbf{r}_4, \alpha_4).\end{aligned} \tag{A2}$$

We may substitute into these formulas, in particular, the potentials that correspond to a system of bosons or fermions interacting by a two-body force. In that case we have

$$\begin{aligned}\hat{w}_{\frac{1}{2}} &= 0, \\ \hat{w}_1(\mathbf{r}_1, \alpha_1; \mathbf{r}_2, \alpha_2) &= \delta_{\alpha_1,2}\delta_{\alpha_2,1}t(\mathbf{r}_1, \mathbf{r}_2), \\ \hat{w}_{\frac{3}{2}} &= 0, \\ \hat{w}_2(\mathbf{r}_1, \alpha_1; \mathbf{r}_2, \alpha_2; \mathbf{r}_3, \alpha_3; \mathbf{r}_4, \alpha_4) &= \tfrac{1}{2}\delta_{\alpha_1,2}\delta_{\alpha_2,2}\delta_{\alpha_3,1}\delta_{\alpha_4,1}V(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4),\end{aligned} \tag{A3}$$

where the potential V satisfies $V(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) = V(\mathbf{r}_2, \mathbf{r}_1, \mathbf{r}_4, \mathbf{r}_3)$ and where, without any loss in generality, it can also be taken to satisfy

$$V(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) = \epsilon V(\mathbf{r}_2, \mathbf{r}_1, \mathbf{r}_3, \mathbf{r}_4).$$

If we have nonrelativistic particles, a local potential, and no external force, we have

[28] We wish to point out that the substitution of (II. 61) (the following paper is referred to as II) into (77) does give a functional differential equation for the 2-irreducible-diagram contribution $\mathcal{K}^{(2)}$ to $F^{(2)}$ for normal systems. Likewise, the substitution of (II. 75) into the corresponding equation obtained from (79) gives a partial differential equation for $\mathcal{K}^{(2)}$ for superfluid systems. There is no compelling argument for making this subdivision of $F^{(2)}$ from the point of view of the equations here derived, although the first terms on either side of (77) do suggest the logarithmic structure present in (II. 61).

$$t(\mathbf{r}_1, \mathbf{r}_2) = -[\nabla_1^2 + \mu]\delta(\mathbf{r}_1 - \mathbf{r}_2) , \tag{A4}$$

$$V(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) = \tfrac{1}{2}V(\mathbf{r}_1 - \mathbf{r}_2)$$

$$\times [\delta(\mathbf{r}_2 - \mathbf{r}_3)\delta(\mathbf{r}_1 - \mathbf{r}_4) + \epsilon\delta(\mathbf{r}_1 - \mathbf{r}_3)\delta(\mathbf{r}_2 - \mathbf{r}_4)],$$

in units in which $2m = 1$, and $\hbar = 1$. For the potentials in (A3), the nonvanishing components of the potentials are

$$\vartheta_0 = -\tfrac{1}{2}\epsilon t(\mathbf{r}_1, \mathbf{r}_1) + \tfrac{1}{4}V(\mathbf{r}_1, \mathbf{r}_3, \mathbf{r}_3, \mathbf{r}_1), \tag{A5}$$

$$\vartheta_1(\mathbf{r}_1, \alpha_1; \mathbf{r}_2, \alpha_2)$$

$$= [t(\mathbf{r}_1, \mathbf{r}_2) - V(\mathbf{r}_3, \mathbf{r}_1, \mathbf{r}_3, \mathbf{r}_2)]\delta_{\alpha_1,2}\delta_{\alpha_2,1}$$

$$+ \epsilon[t(\mathbf{r}_2, \mathbf{r}_1) - V(\mathbf{r}_3, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_1)]\delta_{\alpha_2,2}\delta_{\alpha_1,1}, \tag{A6}$$

$$\begin{aligned}
\vartheta_2(\mathbf{r}_1, 1; \mathbf{r}_2, 1; \mathbf{r}_3, 2; \mathbf{r}_4, 2) &= V(\mathbf{r}_4, \mathbf{r}_3, \mathbf{r}_2, \mathbf{r}_1),\\
\vartheta_2(\mathbf{r}_1, 1; \mathbf{r}_2, 2; \mathbf{r}_3, 1; \mathbf{r}_4, 2) &= V(\mathbf{r}_2, \mathbf{r}_4, \mathbf{r}_3, \mathbf{r}_1),\\
\vartheta_2(\mathbf{r}_1, 1; \mathbf{r}_2, 2; \mathbf{r}_3, 2; \mathbf{r}_4, 1) &= V(\mathbf{r}_3, \mathbf{r}_2, \mathbf{r}_4, \mathbf{r}_1),\\
\vartheta_2(\mathbf{r}_1, 2; \mathbf{r}_2, 2; \mathbf{r}_3, 1; \mathbf{r}_4, 1) &= V(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4),\\
\vartheta_2(\mathbf{r}_1, 2; \mathbf{r}_2, 1; \mathbf{r}_3, 2; \mathbf{r}_4, 1) &= V(\mathbf{r}_1, \mathbf{r}_3, \mathbf{r}_4, \mathbf{r}_2),\\
\vartheta_2(\mathbf{r}_1, 2; \mathbf{r}_2, 1; \mathbf{r}_3, 1; \mathbf{r}_4, 2) &= V(\mathbf{r}_1, \mathbf{r}_4, \mathbf{r}_2, \mathbf{r}_3).
\end{aligned} \tag{A7}$$

(2) In the text we refer to the value of

$$W = \ln \mathrm{Tr}\, e^{-H}$$

for a noninteracting, noncondensed system. To determine this number, we consider

$$F^{(0)} = W + v_0$$

for such a system and define

$$F^{(0)}(\lambda) = \ln \mathrm{Tr} \exp\left[-\frac{\lambda}{2}\int \Psi(1)\Psi(2)v_1(12) - \lambda \int \Psi(1)v_{\frac{1}{2}}(1)\right].$$

For $F^{(0)}(\lambda)$ we have the differential equation

$$\lambda \frac{dF^{(0)}(\lambda)}{d\lambda} = \tfrac{1}{2}\epsilon \int v_{1\lambda}(12)G_{1\lambda}(21) + \int v_{\frac{1}{2}\lambda}(1)G_{\frac{1}{2}\lambda}(1),$$

where $v_{i\lambda} \equiv \lambda v_i$ and $G_{i\lambda}$ are the propagators for the system whose Hamiltonian consists of λv_1 and λv_2. Inserting the known values of $G_{\frac{1}{2}\lambda}$ and $G_{1\lambda}$ for such a system, and integrating from $\lambda = 0$ to $\lambda = 1$, we obtain

$$F^{(0)}(1) - F^{(0)}(0) = \tfrac{1}{2}\epsilon \,\mathrm{tr} \ln \tilde{G}_1(\lambda = 1) - \tfrac{1}{2}\epsilon \,\mathrm{tr} \ln \tilde{G}_1(\lambda = 0),$$

or

$$\begin{aligned}
F^{(0)} &= \tfrac{1}{2}\epsilon \,\mathrm{tr} \ln \tilde{G}_1 + f^{(0)},\\
f^{(0)} &= \tfrac{1}{2}\epsilon \,\mathrm{tr} \ln [\sigma(\epsilon\partial/\partial z)\delta(z - z')] - \ln \mathrm{Tr}\, 1.
\end{aligned} \tag{A8}$$

By appropriately inserting limiting operations to avoid the singular part of the equal time functions, we could eliminate the need for this explicit subtraction.

Stationary Entropy Principle and Renormalization in Normal and Superfluid Systems. II. Diagrammatic Formulation*

CYRANO DE DOMINICIS AND PAUL C. MARTIN†
Service de Physique Théorique, Centre d'Etudes Nucléaires de Saclay (S.O.) France and
Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts
(Received 2 August 1963)

Quantum statistical mechanics is renormalized, that is, the thermodynamical functions and the "bare" ν-body potentials v_ν occurring in the Hamiltonian are expressed as functionals of the ν-body distribution functions G_ν (the expectation values of 2ν field operators). The field operators are considered as having two components (creation and annihilation), and G_ν thus has $(2)^{2\nu}$ components. The renormalization is carried out for superfluid Bose systems for which it is necessary to consider the expectation values of $2\nu = 1, 2, 3, 4$ operators. Superfluid Fermi systems and normal systems appear as particular cases, described by expectation values of $2\nu = 2, 4$ operators.

This problem, dealt with by algebraic methods in part I, is attacked here by diagrammatic methods. These methods are more dependent on convergence properties but the resulting functionals are obtained explicitly as power series of the G_ν's, the general term being represented by a class of diagrams characterized by their topological structure.

In the final result, the entropy is exhibited as an explicit functional of the distribution functions G_ν, $(2\nu = 1, 2, 3, 4)$, or more precisely, of functions directly related to them. This functional no longer involves the potentials (nor the equilibrium parameters). It is stationary under independent variations of the G_ν's, subject to the constraints of constant energy and particle number. The four equations of stationarity exhibit each function v_ν as a functional of the G_ν's, self-consistently defining these distribution functions.

INTRODUCTION

IN the first part of this work,[1] algebraic methods have been used to renormalize quantum statistical mechanics, that is to express the thermodynamical functions, and the bare potentials occurring in the Hamiltonian, as functionals of a few distribution functions. These functionals appeared as solutions of functional differential equations. The renormalization procedure was applied to "superfluid" systems, "normal" (nonsuperfluid) systems being considered as a particular case.

In this second part, we use diagrammatic methods to attack the same problem. These methods rely on perturbation expansions and resummations. They are therefore more dependent on convergence properties; they are also more *explicit* in the sense that the resultant functionals in terms of the distribution functions are exhibited as power series whose general term is represented by a class of diagrams characterized by their topological structure.

We consider the grand partition function

$$e^W = \operatorname{tr} e^{-H}$$

for a Bose or a Fermi system described by the Hamiltonian (in dimensionless units)

$$H = v_0 + v_{\frac{1}{2}}(x_1)\Psi(x_1) + \frac{1}{2!}v_1(x_1, x_2)\Psi(x_1)\Psi(x_2) + \frac{1}{3!}v_{\frac{3}{2}}(x_1, x_2, x_3)\Psi(x_1)\Psi(x_2)\Psi(x_3) + \frac{1}{4!}v_2(x_1, x_2, x_3, x_4)\Psi(x_1)\Psi(x_2)\Psi(x_3)\Psi(x_4). \quad (1)$$

A summation convention is employed for the repeated variables x_i. We recall that the variable x_i stands for a spatial variable $\mathbf{r}_i$ (or a momentum variable $\mathbf{k}_i$), and a "spinor" index $\alpha_i(\alpha_i = 1, 2)$. The field operators $\Psi(x)$ are "spinors"[2] with two components $\Psi(\mathbf{r}, \alpha=1)=\psi(\mathbf{r})$ and $\Psi(\mathbf{r}, \alpha=2)=\psi^+(\mathbf{r})$ (where ψ^+, ψ are the usual creation and annihilation operators). Spatial (or momentum) variables are also assumed to incorporate spinlike internal degrees of freedom. The ν-body potentials v_ν have $(2)^{2\nu}$ components in α space. The reasons for considering a Hamiltonian containing $\frac{1}{2}$-body and $\frac{3}{2}$-body potentials to describe superfluid systems have been examined in I (Sec. I).[3]

For such a system we consider the ν-body "time"-dependent distribution functions (Green's functions) G_ν with $\nu = \frac{1}{2}, 1, \frac{3}{2}, 2$, defined by the T-ordered product [Eq. (I.13)].

* A preliminary account of this work was presented at the Brown Conference on Phase Transitions and Irreversible Thermodynamics, June 11–15, 1962; J. Math. Phys. 4, 208, 255 (1963).

† Sloan Foundation Fellow.

[1] C. De Dominicis and P. C. Martin, J. Math. Phys. 5, 14, (1964) (previous paper, hereafter denoted by I).

[2] Y. Nambu, Phys. Rev. 117, 648 (1960).

[3] P. C. Martin and P. Hohenberg (to be published).

$$G_\nu(x_1, z_1; x_2, z_2; \cdots x_{2\nu}, z_{2\nu})$$
$$= e^{-W} \operatorname{tr} \{e^{-H} T(\Psi(x_1, z_1)\Psi(x_2, z_2) \cdots \Psi(x_{2\nu}, z_{2\nu}))\}, \quad (2)$$

where the "time", z_i, varies between 0 and 1, and

$$\Psi(x_i, z_i) = e^{z_i H}\Psi(x_i)e^{-z_i H}.$$

Section I begins with the perturbation expansions of W and G_ν in powers of the potentials $v_{\nu'}$. Through a set of rules derived in appendix A, the terms of the expansion are put in a one-to-one correspondence with a class of diagrams characterized by specified topological properties. Whole classes of diagrams are then resummed. Through this resummation process, the expansions of G_ν in powers of the potentials are inverted; each v_ν, ($\nu = \frac{1}{2}, 1, \frac{3}{2}, 2$) is successively expressed in powers of the distribution functions $G_{\nu'}$ ($\nu' \leq \nu$) and, if $\nu \neq 2$, in powers of higher-order potentials $v_{\nu''}$ ($\nu'' > \nu$). These expansions are represented by new diagrams built with renormalized propagators or vertices, and are derived in the first part of Secs. II to V, respectively. The details of the resummation procedure are considered in appendix A, whose last paragraph contains a summary of all diagrammatic rules used throughout the paper.

The expression of the thermodynamical functions (W, free energy, entropy) in terms of the distribution functions is slightly more complicated and requires the use of some topological theorems derived in Appendix B. The thermodynamical functions are successively constructed as stationary functionals of the G_ν's in the second part of Secs. II to V. Eventually the entropy is explicitly exhibited (Eq. 75) as a functional of $G_{\frac{1}{2}}$, G_1, $G_{\frac{3}{2}}$, G_2, or rather of functions directly related to them. This entropy functional is *fourfold stationary* under independent variations of the G_ν's. The four coupled equations [Eqs. (40), (76), (72), (77)] expressing the stationarity conditions (subject to the constraints of constant energy and particle number) exhibit the "bare" potentials v_ν as functionals of the four distribution functions $G_{\nu'}$. The entropy functional no longer contains any dependence upon the potentials v_ν (or the equilibrium parameters) which will only appear through the constraints of the problem.

The notations used in this part are identical with those of I; however, Appendix A excepted, we shall speak in terms of "energy" rather than "time" variables. That is, we shall incorporate the property[4,5] that the functions G_ν are periodic (bosons) or antiperiodic (fermions) in the "time" interval $0 < z < 1$, by using Fourier series transforms. The transform variable ω, which will be called an energy, takes the values $2ni\pi$ (bosons) or $(2n+1)i\pi$ (fermions), n being any integer. We shall write

$$G_\nu(x_1, \omega_1; \cdots; x_{2\nu}, \omega_{2\nu}) = \int_0^1 dz_1 \cdots \int_0^1 dz_{2\nu}$$
$$\times \exp\left(\sum \omega_i z_i\right) G_\nu(x_1, z_1; \cdots; x_{2\nu}, z_{2\nu}),$$

$$G_\nu(x_1, z_1; \cdots; x_{2\nu}, z_{2\nu})$$
$$= \sum_{\{\omega_i\}} \exp\left(-\sum \omega_i z_i\right) G_\nu(x_1, \omega_1; \cdots; x_{2\nu}, \omega_{2\nu}).$$

The translational invariance in the "time" variables guarantees that $G_\nu(\{\omega_i\})$ is only nonvanishing when $\sum_{i=1}^{2\nu} \omega_i = 0$. The pair of variables x_i and ω_i will be denoted throughout the text by the variable j. We write

$$G_\nu(1, 2, \cdots 2\nu) \equiv G_\nu(x_1, \omega_1; x_2, \omega_2; \cdots x_{2\nu}, \omega_{2\nu}).$$

A summation convention is always implied for repeated *variables j*.

I. UNRENORMALIZED PERTURBATION THEORY

A. Rules of Perturbation Theory

Perturbation expansions for W and G_ν are most conveniently described through a Wick theorem which is easily proven by induction (Appendix A, Sec. 1). It is this Wick theorem which establishes a one-to-one correspondance between algebraic terms of the expansion and a set of diagrams in which each element is associated with an algebraic factor by well-defined rules.

Diagrams are build with ν-vertices, represented by a black dot where 2ν lines intersect, and with solid lines connecting vertices (internal lines) or going from a vertex to a specified point (external point). Each external point is attached to one and only one line (external line). They are Feynman diagrams, that is, only the *topology* of the diagrams is relevant; the relative position of the vertices is not. A *ν-diagram* is a diagram with *2ν external lines*; the algebraic contribution associated with any diagram is calculated by the following Rule (A) (See also Appendix A, Secs. 2, 3, 5).

(Ai) With each ν-vertex ($\nu = \frac{1}{2}, \frac{3}{2}, 2$) we associate a potential matrix[6] with $(2)^{2\nu}$ components

[4] E. Montroll and J. Ward, Phys. Fluids **1**, 55 (1958).
[5] A. Abrikosov, L. Gorkov, and I. Dzyaloshinsky, Zh. Eksperim. i Teor. Fiz. **36**, 900 (1959) [English transl.: Soviet Phys.—JETP **9**, 636 (1959)]; E. Fradkin, Nucl. Phys. **12**, 465 (1959); P. Martin and J. Schwinger, Phys. Rev. **115**, 1342 (1959).

[6] These potential matrices are just the Fourier series transforms of instantaneous potentials, that is of the potentials

$$-v_{\frac{1}{2}}(j) = -v_{\frac{1}{2}}(x_j)\delta_{\omega_j,0},$$
$$-v_{\frac{1}{2}}(j, k, l) = -v_{\frac{1}{2}}(x_j, x_k, x_l)\delta_{\omega_j+\omega_k+\omega_l,0}, \quad (3)$$
$$-v_2(j, k, l, m) = -v_2(x_j, x_k, x_l, x_m)\delta_{\omega_j+\omega_k+\omega_l+\omega_m,0}.$$

(Aii) With an internal line connecting the vertex arguments j and j', we associate the "unperturbed" propagator $G_1^0(j, j')$ defined by

$$[G_1^0]^{-1}(j, j') = -\omega_j\sigma^\epsilon_{\alpha_j\alpha_{j'}} \times \delta(\mathbf{r}_j - \mathbf{r}'_j)\delta_{\omega_j+\omega_{j'},0} + v_1(j, j'), \quad (4)$$

where $\sigma^\epsilon_{\alpha_j\alpha_{j'}}$ are elements of

$$\sigma^\epsilon = \begin{bmatrix} 0 & -\epsilon \\ 1 & 0 \end{bmatrix}.$$

With an external line connecting a vertex with argument j to an external point with a prescribed argument j' (that is, with a prescribed value of $x_{j'}$ and $\omega_{j'}$) we associate an unperturbed propagator $G_1^0(j, j')$.

(Aiii) The prescribed arguments of the 2ν external points are fixed; the arguments of the vertices are summed over.[7]

(Aiv) When a line has both ends at the same vertex, we associate with the line a factor $\frac{1}{2}$.

When two vertices are *directly* connected by m lines[8] we associate with them a factor $(m!)^{-1}$.

(Av) With each diagram, we associate a factor $(S_V)^{-1}$, where S_V is a symmetry number for the vertices of the diagram.[9]

Rule (A) would be incomplete for Fermi systems without a sign rule which is derived in Appendix A. These rules and those encountered in the rest of the text are fully contained in the general rule displayed at the end of Appendix A, Sec. 6.

When the Hamiltonian reduces to the one-body potential v_1 (plus the constant v_0), the system is said to be unperturbed. For example, in an unperturbed uniform system in which v_1 reduces to the kinetic energy minus the chemical potential μ, the ϵ-symmetrized v_1 may be written as

$$v_1(\mathbf{r}, \alpha; \mathbf{r}', \alpha') = \int \frac{d\mathbf{k}}{(2\pi)^3} \exp[i\mathbf{k}\cdot(\mathbf{r} - \mathbf{r}')]\beta(k^2 - \mu)(\sigma^\epsilon\tau^3)_{\alpha\alpha'},$$

where we have taken $2m = 1$, $\hbar = 1$, and

$$\tau^3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \qquad \sigma^\epsilon\tau^3 = \begin{bmatrix} 0 & \epsilon \\ 1 & 0 \end{bmatrix}.$$

In this case we may also write

$$G_1^0(\mathbf{r}, \alpha, z; \mathbf{r}', \alpha', z') = \int \frac{d\mathbf{k}}{(2\pi)^3} \times \sum_\omega \exp[i\mathbf{k}\cdot(\mathbf{r} - \mathbf{r}') - \omega(z - z')]G^0_{1\alpha\alpha'}(k, \omega). \quad (5)$$

From (4) we see that

$$[G_1^0(k, \omega)]^{-1} = -\omega\sigma^\epsilon + \beta(k^2 - \mu)\sigma^\epsilon\tau^3,$$

or, in an explicit form,

$$G_1^0(k, \omega) = \begin{bmatrix} 0 & [-\omega + \beta(k^2 - \mu)]^{-1} \\ \epsilon[\omega + \beta(k^2 - \mu)]^{-1} & 0 \end{bmatrix}.$$

B. Perturbation Expansions

A ν-diagram is *connected* if there exists at least one continuous path joining any two internal (vertex) or external points.

A ν-diagram is *linked* if any internal line is directly or indirectly connected to at least one external point.

The following diagrammatic representations of perturbation expansions are immediately obtained:

$$e^{W-W_0} = \sum \text{(all distinct 0-diagrams)},$$
$$W - W_0 = \sum \text{(all distinct, connected 0-diagrams)}, \quad (6)$$

where e^{W_0} is the unperturbed grand partition function. For the ν-body Green's function we have

$$G_\nu(1, 2, \cdots 2\nu) = \sum \text{(all distinct, linked, } \nu\text{-diagrams)}. \quad (7)$$

---

introduced in I, Eq. (21),

$$v_\nu(x'_1, z_1; x'_2, z_2; \cdots; x'_{2\nu}, z_{2\nu}) = v_\nu(x'_1, x'_2, \cdots, x'_{2\nu})\delta(z_2 - z_1)\cdots\delta(z_{2\nu} - z_1).$$

We recall that the potential matrices v_ν are ϵ-symmetrized, i.e.,

$$v_\nu(\cdots x \cdots x' \cdots) = \epsilon v_\nu(\cdots x' \cdots x \cdots),$$

with $\epsilon = -1$ for Fermi systems, $\epsilon = +1$ for Bose systems. These potential matrices are exhibited in terms of the usual potentials in the Appendix of I.

The one-body potential v_1 is incorporated in the propagator (see also footnote 16).

[7] Notice that at *each vertex* we have energy conservation (as a consequence, each $\frac{1}{2}$-vertex only admits the energy component $\omega_j = 0$).

For translationally invariant potentials we also have, at each ν-vertex, momentum conservation. Even if the potentials in the Hamiltonian are translationally invariant, it is useful to consider additional potentials which do not have this property and will break the translational invariance. At the associated vertices there is no momentum conservation.

[8] The only possible nontrivial values of m are $m = 2, 3, 4$. The last occurs only in the single 0-diagram built with two 2-vertices connected by four lines.

[9] To each vertex in a diagram, assign a label i $(1 \leq i \leq p)$. Two labels are connected if there is a line joining directly the corresponding vertices. The symmetry number S_V is the order of the permutation group of the p labels which leaves the set of label connections invariant.

By direct inspection, we may verify that the ν-diagrams of (7) can be generated by cutting one ν-vertex out of the 0-diagrams of (6) and assigning the 2ν fixed arguments to the ends of the cut lines, these operations being performed in all possible distinct ways. We note that this operation is represented algebraically by the functional differentiation

$$G_\nu(1, 2, \cdots 2\nu) = -(2\nu)! \times [\delta/\delta v_\nu(1, \cdots 2\nu)]W\{v_\nu\}, \quad (8)$$

where W is considered as a functional of the v_ν introduced in Rule (Ai).

Furthermore for the cumulant part $\tilde{G}_\nu$ of G_ν [for an algebraic definition see Eq. (I.24)], we also have

$$\tilde{G}_\nu(1, 2, \cdots 2\nu) = \sum (\text{all distinct, connected, } \nu\text{-diagrams}). \quad (9)$$

By means of Eqs. (9) and (7), each distribution function is easily expressed in terms of the cumulants $\tilde{G}_{\nu'}$ $(\nu' \leq \nu)$ [cf. Eq. (I.26)] or conversely.

These expansions give the explicit form for the functional dependence of W, G_ν, and $\tilde{G}_\nu$ in powers of the potentials.

The renormalization process inverts Eqs. (7) or (9) step by step, and successively expresses the potentials v_ν $(2\nu = 1, 2, 3, 4)$ in terms of the distribution functions $G_{\nu'}$ or the cumulants $\tilde{G}_{\nu'}$ $(\nu' \leq \nu)$ and higher-order potentials $v_{\nu''}$ $(\nu'' > \nu)$. At each step, as generally discussed in I, Sec. I, a new thermodynamical function $F^{(\nu)}$ is conveniently introduced. $F^{(\nu)}$ is defined by

$$F^{(\nu)} = W + v_0 + v_{\frac{1}{2}}(1)G_{\frac{1}{2}}(1) + \cdots + [1/(2\nu)!]v_\nu(1, 2, \cdots 2\nu)G_\nu(1, 2, \cdots 2\nu).$$

$F^{(1)}$ is close to the free energy; $F^{(2)}$ is the entropy. $F^{(\nu)}$ has for its natural variables $G_{\nu'}(\nu' \leq \nu)$ and $v_{\nu''}(\nu'' > \nu)$, just as W has all the potentials for its natural variables. From (8) it follows that $F^{(\nu)}$ satisfies the relations

$$(2\nu')! \, [\delta F^{(\nu)}/\delta G_{\nu'}(1, 2, \cdots 2\nu')] = v_{\nu'}(1, 2, \cdots 2\nu'),$$

$$-(2\nu'')! \, [\delta F^{(\nu)}/\delta v_{\nu''}(1, 2, \cdots 2\nu'')] = G_{\nu''}(1, 2, \cdots 2\nu'')$$

II. CONDENSATE RENORMALIZATION: ELIMINATION OF $v_{\frac{1}{2}}$

In this first step of the renormalization procedure, we express $v_{\frac{1}{2}}$ and the quantity $F^{(\frac{1}{2})}$,

$$F^{(\frac{1}{2})} = W + v_0 + v_{\frac{1}{2}}(1)G_{\frac{1}{2}}(1),$$

as functionals of the condensate wavefunction $G_{\frac{1}{2}}$ (and v_1 through G_1^0, $v_{\frac{1}{2}}$, v_2).

(1) To analyze the structure of a ν-diagram, we introduce the following definitions:

Consider a line in a connected 0-diagram. It is called an *articulation line* if the diagram separates into two disconnected parts when it is cut. If one and only one of those parts is a single $\frac{1}{2}$-vertex, the articulation line is said to be *trivial*. A *$\frac{1}{2}$-irreducible 0-diagram*, is defined as containing *only trivial* articulation lines.

Consider the class of all $\frac{1}{2}$-irreducible 0-diagrams. Such 0-diagrams may contain $\frac{1}{2}$-vertices or they may not. Their contribution is, in particular, a functional of $-v_{\frac{1}{2}}$ or of $-G_1^0 v_{\frac{1}{2}} \equiv -G_1^0(1, \bar{1})v_{\frac{1}{2}}(\bar{1})$, since each $\frac{1}{2}$-vertex is necessarily connected to the rest of the diagram through a line. We call the contribution of this class of 0-diagrams $\mathcal{K}^{(\frac{1}{2})}\{-G_1^0 v_{\frac{1}{2}}\}$.

The class of $\frac{1}{2}$-irreducible $\frac{1}{2}$-diagrams is defined correspondingly, that is, by cutting in all possible ways one $\frac{1}{2}$-vertex out of the 0-diagrams of $K^{(\frac{1}{2})}$. The contribution of this class is given by

$$G_1^0(1\bar{1})K_{\frac{1}{2}}\{\bar{1}; -G_1^0 v_{\frac{1}{2}}\}, \quad (10)$$

where $K_{\frac{1}{2}}$ is the functional derivative,

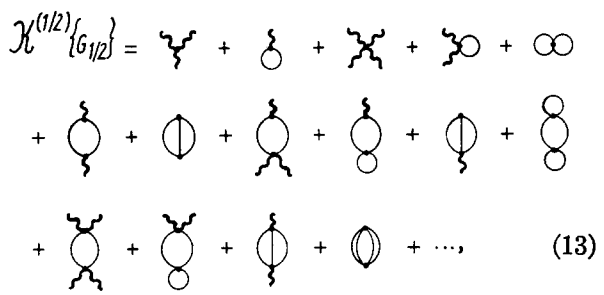
$$K_{\frac{1}{2}}\{1; -G_1^0 v_{\frac{1}{2}}\} = -[\delta/\delta G_1^0(1, \bar{1})v_{\frac{1}{2}}(\bar{1})]\mathcal{K}^{(\frac{1}{2})}\{-G_1^0 v_{\frac{1}{2}}\}. \quad (11)$$

The general $\frac{1}{2}$-diagram (except for the first-order diagram containing a single $\frac{1}{2}$-vertex) is then generated from the $\frac{1}{2}$-irreducible $\frac{1}{2}$-diagram by replacing each $\frac{1}{2}$-vertex (and the line attached to it) by the complete class of $\frac{1}{2}$-diagrams[10]; that is, by replacing in Eq. (10), $-G_1^0(1\bar{1})v_{\frac{1}{2}}(\bar{1})$ by $G_{\frac{1}{2}}(1)$. This operation yields

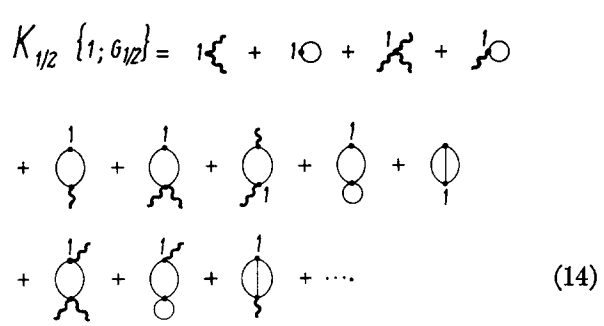
$$G_{\frac{1}{2}}(1) = G_1^0(1\bar{1})[-v_{\frac{1}{2}}(\bar{1}) + K_{\frac{1}{2}}\{\bar{1}; G_{\frac{1}{2}}\}]. \quad (12)$$

After this substitution, $K_{\frac{1}{2}}$ and correspondingly $\mathcal{K}^{(\frac{1}{2})}$, become functionals of $G_{\frac{1}{2}}$ (and of v_1 through G_1^0, $v_{\frac{1}{2}}$, v_2). They are represented by all distinct, connected, $\frac{1}{2}$-irreducible $\frac{1}{2}$-diagrams (stripped of their external line), and 0-diagrams, respectively. These diagrams are built with ν-vertices $(\nu = \frac{3}{2}, 2)$, lines (unperturbed propagators) and heavy wavy $\frac{1}{2}$-lines ($\frac{1}{2}$-propagators); these $\frac{1}{2}$-lines have a single end and are attached to a single vertex in the diagram. To second order in v_ν we have

[10] Or, equivalently, by replacing each $\frac{1}{2}$ vertex by the complete class of $\frac{1}{2}$ diagrams stripped of the external line, i.e., replacing $-v_{\frac{1}{2}}(1)$ by $[G_1{}^0]^{-1}(1, \bar{1})G_{\frac{1}{2}}(\bar{1})$.



(13)



(14)

After removal of the external line, the $\frac{1}{2}$-diagrams of $K_{\frac{1}{2}}$ contain one ν-vertex with only $(2\nu - 1)$ lines (or half lines) attached to it and a (2ν)th argument with a prescribed value. The prescribed argument [1 in Eq. (14)] is always written next to the vertex. Associated contributions are calculated with Rule (B) differing from Rule (A) only through

(Bi) With each ν-vertex ($\nu = \frac{3}{2}$ or 2), associate $-v_\nu$.

(Bii) With each heavy (wavy) $\frac{1}{2}$-line j, associate the $\frac{1}{2}$-propagator $G_{\frac{1}{2}}(j)$.

With each line j, j', associate the unperturbed propagator $G_1^0(j, j')$.

(Biv) Associate a factor $\frac{1}{2}$ with each line with both ends at the same vertex, a factor $(m!)^{-1}$ with each set of m lines directly connecting two vertices, and a factor $(n!)^{-1}$ with each set of n $\frac{1}{2}$-lines attached to the same vertex.

Naturally, instead of (11), we now have

$$K_{\frac{1}{2}}\{1; G_{\frac{1}{2}}\} = [\delta/\delta G_{\frac{1}{2}}(1)]\mathcal{K}^{(\frac{1}{2})}\{G_{\frac{1}{2}}\}. \tag{15}$$

The inversion of relation (7) which gives $G_{\frac{1}{2}}$ as a functional of $v_{\frac{1}{2}}$ is thus achieved in (12). The latter may be rewritten as

$$v_{\frac{1}{2}}(1) = -[G_1^0]^{-1}(1, \bar{1})G_{\frac{1}{2}}(\bar{1}) + K_{\frac{1}{2}}\{1; G_{\frac{1}{2}}\}, \tag{16}$$

or since $G_{\frac{1}{2}}(1)$ has only the component $\omega_1 = 0$,

$$v_{\frac{1}{2}}(1) = -v_1(1\bar{1})G_{\frac{1}{2}}(\bar{1}) + K_{\frac{1}{2}}\{1; G_{\frac{1}{2}}\}.$$

(2) We now express $F^{(\frac{1}{2})}$ (or W) as a functional of $G_{\frac{1}{2}}$. In I, this purpose was achieved by direct integration of Eq. (16) We follow here a method[11] already used elsewhere,[12,13] which analyzes in detail the topological structure of the diagrams and gives another insight into the stationarity property. This method will be used through all stages of the renormalization procedure.

It is no longer possible to resum directly families of diagrams in order to express $W - W_0$ in terms of $G_{\frac{1}{2}}$. The direct resummation procedure works for $\frac{1}{2}$-diagrams because those diagrams have some element distinguished (in this case, one external point). For a 0-diagram, however, there is no unique way to decide what is inserted in what, and this procedure fails. To remove this ambiguity we use the following topological relation, which is valid for any 0-diagram, and may be proven immediately by induction:

$$-N(\text{art 1}) + N(\mathcal{K}^{(\frac{1}{2})}) + N(v_{\frac{1}{2}}) = 1. \tag{17}$$

In this expression N (art 1) is the number of articulation lines, and $N(v_{\frac{1}{2}})$ the number of $\frac{1}{2}$-vertices in a 0-diagram. $N(\mathcal{K}^{(\frac{1}{2})})$ is the number of $\frac{1}{2}$-irreducible skeletons, that is, the number of ways in which one can generate the 0-diagram out of a $\frac{1}{2}$-irreducible 0-diagram (the skeleton) by inserting, in place of any of its $\frac{1}{2}$-vertices, any suitable $\frac{1}{2}$-diagram (stripped of its external line).

By using Eq. (17) we can replace 0-diagrams, which do not lend themselves to resummations, by a linear combination of three families of 0-diagrams, each having some element distinguished, and therefore easily amenable to resummations. We now carry out this procedure.

First we consider the ensemble of all distinct 0-diagrams (calculated with Rule A), where we *distinguish* in all possible ways *one articulation line*. The contribution of such an ensemble of 0-diagrams is equal to the sum of the contributions of all distinct 0 diagrams but now weighted by N(art 1). Since this ensemble can be generated by starting from a line $G_1^0(\bar{1}, \bar{2})$ and inserting at each end the complete class of $\frac{1}{2}$ diagrams (each one stripped of its external line), we get for this ensemble the contribution

$$\tfrac{1}{2}G_{\frac{1}{2}}(1)[G_1^0]^{-1}(1, \bar{1})G_1^0(\bar{1}, \bar{2})[G_1^0]^{-1}(\bar{2}, 2)G_{\frac{1}{2}}(2) = \tfrac{1}{2}G_{\frac{1}{2}}(1)[G_1^0]^{-1}(1, 2)G_{\frac{1}{2}}(2).$$

Next, we *distinguish* one $\frac{1}{2}$-*irreducible skeleton* or one $\frac{1}{2}$-*vertex* in all possible ways, obtaining two ensembles of 0 diagrams weighted, respectively, by $N(\mathcal{K}^{(\frac{1}{2})})$ or $N(v_{\frac{1}{2}})$. These distinguished 0-diagrams can be generated by starting from a $\frac{1}{2}$-irreducible

[11] C. Bloch, Physica Suppl. **26**, 62 (1960).

[12] R. Balian, C. Bloch, and C. De Dominicis, Nucl. Phys. **25**, 529 (1961).

[13] C. De Dominicis, J. Math. Phys. **3**, 983 (1962).

skeleton or a $\frac{1}{2}$-vertex, and inserting at each possible place the complete class of $\frac{1}{2}$-diagrams. They contribute, respectively,

$$\mathcal{K}^{(\frac{1}{2})}\{G_{\frac{1}{2}}\}, \quad \text{and} \quad -v_{\frac{1}{2}}(1)G_{\frac{1}{2}}(1).$$

Finally we use relation (17) which enables us to provide the proper weight for each 0-diagram by taking a suitable combination of the distinguished diagrams introduced above. We thus obtain[14] the functional

$$W - W_0 = -\tfrac{1}{2}G_{\frac{1}{2}}(1)[G_1^0]^{-1}(1, 2)G_{\frac{1}{2}}(2) + \mathcal{K}^{(\frac{1}{2})}\{G_{\frac{1}{2}}\} - v_{\frac{1}{2}}(1)G_{\frac{1}{2}}(1). \quad (18)$$

This expression is verified to be stationary under variations of $G_{\frac{1}{2}}$, the stationarity condition being identical with (16). This stationarity property could be seen to have its origin in the general structure of the diagrams.[11,12]

Equation (18) expresses $F^{(\frac{1}{2})}$ as a functional of $G_{\frac{1}{2}}$ only,

$$F^{(\frac{1}{2})}\{G_{\frac{1}{2}}\} = W_0 - \tfrac{1}{2}G_{\frac{1}{2}}(1)[G_1^0]^{-1}(1, 2)G_{\frac{1}{2}}(2) + \mathcal{K}^{(\frac{1}{2})}\{G_{\frac{1}{2}}\}, \quad (19)$$

and thus completes the first step of the renormalization procedure.

The second variation of W (or $F^{(\frac{1}{2})}$) with respect to $G_{\frac{1}{2}}$ is easily shown to be *negative definite.*

III. PROPAGATOR RENORMALIZATION: ELIMINATION OF v_1

In Sec. II we have obtained explicit expressions for $v_{\frac{1}{2}}$ and $F^{(\frac{1}{2})}$, and hence for the distribution functions as functionals of $G_{\frac{1}{2}}$ (and G_1^0, $v_{\frac{3}{2}}$, v_2), that is, a representation in terms of $\frac{1}{2}$-irreducible diagrams of the thermodynamical functions and the distribution functions. Here we want to express v_1 and

$$F^{(1)} = W + v_0 + v_{\frac{1}{2}}(1)G_{\frac{1}{2}}(1) + \frac{1}{2!}v_1(1, 2)G_1(1, 2) \quad (20)$$

as functionals of $\tilde{G}_1$ (and $G_{\frac{1}{2}}$, $v_{\frac{3}{2}}$, v_2).

(1) In this section, we consider only $\frac{1}{2}$-irreducible ν-diagrams (calculated with Rule B), and to analyze them further, we introduce the following definitions:

Consider a set of m internal lines in a ($\frac{1}{2}$-irreducible) 0-diagram. This set is called a *cycle*[15] *of lines* (of multiplicity m) if the diagram is separated into two disconnected parts by cutting any two lines of the set. Removal of the m lines of the cycle separates the diagram into m disconnected parts. *A 1-irreducible 0-diagram* is a 0-diagram with no cycle of multiplicity $m > 1$.

Consider the class of all 1-irreducible 0-diagrams and call their contribution

$$\mathcal{K}^{(1)}\{G_1^0\}.$$

The class of 1-irreducible 1 diagrams is defined correspondingly. That is, it is obtained by opening one line[16] in all possible ways in the 1-irreducible 0-diagrams of $\mathcal{K}^{(1)}$. Its contribution is expressed as

$$G_1^0(1, \bar{1})K_1\{\bar{1}, \bar{2}; G_1^0\}G_1^0(\bar{2}, 2),$$

in terms of the functional derivative

$$K_1\{1, 2; G_1^0\} = 2![\delta/\delta G_1^0(1, 2)]\mathcal{K}^{(1)}\{G_1^0\}. \quad (21)$$

The functional K_1 is usually known as a "mass operator."

The general connected 1-diagram is then generated by

(i) replacing each internal line by the complete class of connected 1-diagrams or in other words replacing the unperturbed Green's function G_1^0 in the argument of K_1 by the cumulant part $\tilde{G}_1$ of the exact one-body Green's function;

(ii) iterating this new "mass operator;" this construction yields a generalized Dyson–Belyaev equation

$$\tilde{G}_1(1, 2) = G_1^0(1, 2) + G_1^0(1, \bar{1})K_1\{\bar{1}, \bar{2}; \tilde{G}_1\}\tilde{G}_1(\bar{2}, 2). \quad (22)$$

After the substitution, $G_1^0 \to \tilde{G}_1$, K_1 and the corresponding $\mathcal{K}^{(1)}$ are functionals of $\tilde{G}_1$ (as well as $G_{\frac{1}{2}}$, $v_{\frac{3}{2}}$, v_2). They are represented by all distinct, connected, 1-irreducible 1-diagrams (stripped of their external lines), and 0-diagrams, respectively. These diagrams are built with ν-vertices ($\nu = \frac{3}{2}, 2$), heavy (wavy) $\frac{1}{2}$-lines (that is, $\frac{1}{2}$-propagators), heavy lines (propagators). To second order in v, we have

[14] The contributions of the three ensembles considered, as well as the contribution of the complete class of $\frac{1}{2}$-diagrams given by Eq. (12) are obtained here on a rather intuitive basis. More precisely, it is assumed (i) that the insertions satisfy a factorization property; (ii) that, in the process of insertion, the diagrams of the original ensemble enter once and only once. These properties, which are assumed throughout the text, are established in Appendix A, Sec. 4.

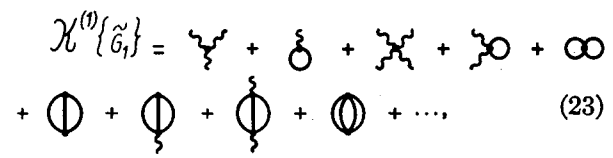
[15] A shorter name than "articulation sequence of lines" which is used in reference 13.

[16] It is possible to formulate this part in a way which is closer to the analysis of Secs. II, IV, and V by keeping the 1-body potential as a perturbation instead of including it directly into G_1^0. 1-diagrams contributing to G_1 are then generated out of 0-diagrams by removing a 1-vertex $v_1(1, 2)$ rather than opening one line. In terms of functional derivatives, this is expressed by

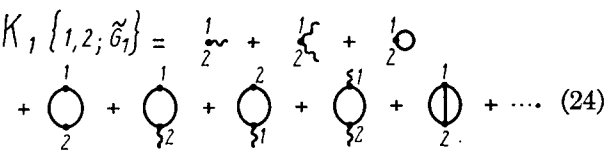
$$G_1(1, 2) = -2![\delta/\delta v_1(1, 2)]W\{v_1\},$$

with

$$v_1(1, 2) = v_1(x_1, x_2)\,\delta_{\omega_1+\omega_2,0}.$$



(23)



(24)

Associated contributions are calculated with Rule (C) differing from Rule (A) through

(Ci) With each ν vertex ($\nu = \frac{3}{2}, 2$), associate $-v_\nu$.

(Cii) With each heavy (wavy) $\frac{1}{2}$-line, associate the $\frac{1}{2}$-propagator $G_{\frac{1}{2}}$.

With each heavy line, associate the propagator $\tilde{G}_1$.

(Civ) Identical to (Biv).

Instead of (21) we now have

$$K_1\{1, 2; \tilde{G}_1\} = 2! \, [\delta/\delta\tilde{G}(1, 2)]\mathcal{K}^{(1)}\{\tilde{G}_1\}, \tag{25}$$

and the inversion of (9) considered as giving $\tilde{G}_1$ as a functional of v_1 is achieved by rewriting (22) as

$$v_1(1, 2) = \tilde{G}_1^{-1}(1, 2) + K_1\{1, 2; \tilde{G}_1\} - \sigma^{\epsilon}\omega_1\delta_{\omega_1+\omega_2,0}\delta(\mathbf{r}_1 - \mathbf{r}_2). \tag{26}$$

(2) The expression for $F^{(1)}$ (or W) as a functional of $\tilde{G}_1$ (and $G_{\frac{1}{2}}$, $v_{\frac{3}{2}}$, v_2) has been obtained in I by direct functional integration. In the same fashion as in Sec. II, we now construct the proper weight for the diagrams contributing to W or more precisely to $\mathcal{K}^{(\frac{1}{2})}$, since we deal only with $\frac{1}{2}$-irreducible diagrams. For this purpose we use the topological relation

$$N(\mathrm{c}) - N(\mathrm{l}) + N(\mathcal{K}^{(1)}) = 1, \tag{27}$$

where $N(\mathrm{c})$ is the number of cycles of lines, $N(\mathrm{l})$ the number of lines and $N(\mathcal{K}^{(1)})$ the number of 1-irreducible skeletons in a 0-diagram.

Consider now the sum of all distinct 0-diagrams (calculated with Rule B) where we *distinguish* in all possible ways *one cycle of lines*. The contribution of this ensemble of 0-diagrams is equal to the contribution of the usual ensemble of 0-diagrams weighted now by $N(\mathrm{c})$. Such an ensemble of 0-diagrams is clearly generated by iterating cyclically 1-irreducible 1-diagrams; its contribution is thus

$$\frac{\epsilon}{2}\,\mathrm{tr}\sum_{m=1}^{\infty}\frac{1}{m}\,[G_1^0K_1]^m = -\frac{\epsilon}{2}\,\mathrm{tr}\,\ln\,[1 - G_1^0K_1],$$

where the factor $(m)^{-1}$ is the symmetry factor $(S_\nu)^{-1}$, and the mth power is understood as a product of m matrices

$$G_1^0(1, \bar{1})K_1\{\bar{1}, 1'; \tilde{G}_1\}.$$

If we *distinguish one line*, the 0-diagrams are weighted by $N(\mathrm{l})$. They are generated by starting from a line and inserting at its ends the complete class of 1-diagrams (each stripped of its external lines). Thus this ensemble of 0-diagrams contributes

$$\tfrac{1}{2}\epsilon[G_1^0]^{-1}(1, \bar{1})\tilde{G}_1(\bar{1}, \bar{2})[G_1^0]^{-1}(\bar{2}, 2)G_1^0(2, 1) = \tfrac{1}{2}\epsilon[G_1^0]^{-1}(1, \bar{1})\tilde{G}_1(\bar{1}, 1).$$

If we *distinguish one 1-irreducible skeleton*, we introduce a weight $N(\mathcal{K}^{(1)})$, and we get the contribution

$$\mathcal{K}^{(1)}\{\tilde{G}_1\}.$$

Finally we use relation (27) which permits us to construct the proper weight for a 0-diagram by taking the proper combination of distinguished diagrams. We obtain

$$\mathcal{K}^{(\frac{1}{2})} = \tfrac{1}{2}\epsilon\,\mathrm{tr}\,(\ln\tilde{G}_1 - \ln G_1^0) - \tfrac{1}{2}\epsilon\,\mathrm{tr}\,([G_1^0]^{-1}\tilde{G}_1) + \mathcal{K}^{(1)}\{\tilde{G}_1\}, \tag{28}$$

which, with (18) and (20), gives

$$F^{(1)}\{\tilde{G}_1\} = \tfrac{1}{2}\epsilon\,\mathrm{tr}\,[\ln\tilde{G}_1 + \sigma^{\epsilon}\omega'\tilde{G}_1(1', 1)] + \mathcal{K}^{(1)}\{\tilde{G}_1\}. \tag{29}$$

The statement that W is stationary under variations of $\tilde{G}_1$ reproduces Eq. (26). Stationarity under variations of $G_{\frac{1}{2}}$ leads to a new form of (16),

$$v_{\frac{1}{2}}(1) = -v_1(1, \bar{1})G_{\frac{1}{2}}(\bar{1}) + [\delta/\delta G_{\frac{1}{2}}(1)]\mathcal{K}^{(1)}\{G_{\frac{1}{2}}, \tilde{G}_1\}. \tag{30}$$

However the quadratic form of the second variations in $G_{\frac{1}{2}}$ and $\tilde{G}_1$ no longer has a definite sign.

IV. FIRST VERTEX RENORMALIZATION: ELIMINATION OF $v_{\frac{3}{2}}$

Having expressed $v_{\frac{1}{2}}$, v_1, and the thermodynamical functions as functionals of $G_{\frac{1}{2}}$, $\tilde{G}_1$, we now take a further step and express $v_{\frac{3}{2}}$ and

$$F^{(\frac{3}{2})} = W + v_0 + v_{\frac{1}{2}}(1)G_{\frac{1}{2}}(1) + (1/2!)v_1(1, 2)G_1(1, 2) + (1/3!)v_{\frac{3}{2}}(1, 2, 3)G_{\frac{3}{2}}(1, 2, 3) \tag{31}$$

as functionals of $\tilde{G}_{\frac{3}{2}}$ (and $G_{\frac{1}{2}}$, $\tilde{G}_1$, v_2), or rather of the stripped $\frac{3}{2}$-correlation function $C_{\frac{3}{2}}$ defined by

$$\tilde{G}_{\frac{3}{2}}(1, 2, 3) = C_{\frac{3}{2}}(\bar{1}, \bar{2}, \bar{3})\tilde{G}_1(\bar{1}, 1)\tilde{G}_1(\bar{2}, 2)\tilde{G}_1(\bar{3}, 3). \tag{32}$$

In this section, we consider only *1-irreducible* ν-

diagrams (calculated with Rule C). In terms of these diagrams, $C_{\frac{3}{2}}(1, 2, 3)$ is represented by all distinct $\frac{3}{2}$-diagrams where the external lines have been removed.

(1) To pursue the analysis we introduce the following definitions.

Consider a set of three lines in a (1-irreducible) 0-diagram. This set is called an *articulation triplet* if the diagram separates into two disconnected parts when all three lines are cut. If one and only one of those parts is a single $\frac{3}{2}$-vertex, the articulation triplet is said to be *trivial*. A *$\frac{3}{2}$-irreducible 0-diagram* is defined as one which contains only trivial articulation triplets.

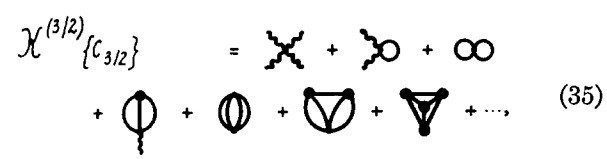
We consider again the class of all $\frac{3}{2}$-irreducible 0-diagrams; its contribution is a functional of $v_{\frac{3}{2}}$, $\mathcal{K}^{(\frac{3}{2})}\{-v_{\frac{3}{2}}\}$. [By convention $\mathcal{K}^{(\nu)}$ does not include the contribution of 0-diagrams built out of a single ν-vertex. In particular, $\mathcal{K}^{(\frac{3}{2})}$ does not contain the first two diagrams of $\mathcal{K}^{(1)}$ as given by Eq. (23)]. The class of $\frac{3}{2}$-irreducible $\frac{3}{2}$-diagrams is defined correspondingly. Its contribution is given by the functional derivative

$$K_{\frac{3}{2}}(\bar{1}, \bar{2}, \bar{3}; -v_{\frac{3}{2}})\tilde{G}_1(\bar{1}, 1)\tilde{G}_1(\bar{2}, 2)\tilde{G}_1(\bar{3}, 3) = -3!\,[\delta/\delta v_{\frac{3}{2}}(1, 2, 3)]\mathcal{K}^{(\frac{3}{2})}\{-v_{\frac{3}{2}}\}. \quad (33)$$

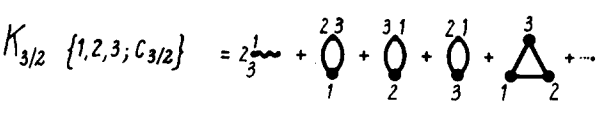
The general $\frac{3}{2}$-diagram—except for the first-order diagram containing a single vertex $v_{\frac{3}{2}}$—is then generated by replacing each $\frac{3}{2}$-vertex inside the $\frac{3}{2}$-irreducible $\frac{3}{2}$-diagram by the complete class of $\frac{3}{2}$-diagrams, each stripped of its three external lines, that is, by replacing $-v_{\frac{3}{2}}$ by $C_{\frac{3}{2}}$. We thus have

$$C_{\frac{3}{2}}(1, 2, 3) = -v_{\frac{3}{2}}(1, 2, 3) + K_{\frac{3}{2}}\{1, 2, 3; C_{\frac{3}{2}}\}. \quad (34)$$

This equation inverts relation (9) considered as giving $\tilde{G}_{\frac{3}{2}}$ (or $C_{\frac{3}{2}}$) as a functional of $v_{\frac{3}{2}}$. $K_{\frac{3}{2}}$ and $\mathcal{K}^{(\frac{3}{2})}$ are now functionals of $C_{\frac{3}{2}}$ (and $G_{\frac{3}{2}}$, $\tilde{G}_1$, v_2); they are represented by all distinct, connected, $\frac{3}{2}$-irreducible $\frac{3}{2}$-diagrams (stripped of their external lines), and 0-diagrams, respectively. These diagrams are built with 2-vertices, heavy $\frac{3}{2}$-vertices ($\frac{3}{2}$-correlation functions), heavy lines (propagators) and heavy (wavy) $\frac{1}{2}$-lines ($\frac{1}{2}$ propagators).



(35)



(36)

Associated contributions are calculated with Rule (D) differing from Rule (A) through

(Di) With each $\frac{3}{2}$-vertex, associate $C_{\frac{3}{2}}$.
With each 2-vertex, associate $-v_2$.
(Dii), (Div) Identical to (Cii), (Civ).

Now (33) becomes

$$K_{\frac{3}{2}}\{\bar{1}, \bar{2}, \bar{3}; C_{\frac{3}{2}}\}\tilde{G}_1(\bar{1}, 1)\tilde{G}_1(\bar{2}, 2)\tilde{G}_1(\bar{3}, 3) = 3!\,[\delta/\delta C_{\frac{3}{2}}(1, 2, 3)]\mathcal{K}^{(\frac{3}{2})}\{C_{\frac{3}{2}}\}.$$

The only $G_{\frac{3}{2}}$ dependence in Eq. (36) occurs through the first term. $G_{\frac{3}{2}}$ cannot occur in higher orders without violating the $\frac{3}{2}$-irreducibility.

(2) To obtain $F^{(\frac{3}{2})}$ (or W) as a functional of $C_{\frac{3}{2}}$ (and $G_{\frac{3}{2}}$, $\tilde{G}_1$, v_2), we follow closely the topological method used in the previous two sections. The proper weight for the 0-diagrams belonging to $\mathcal{K}^{(1)}$ (since this section deals only with 1-irreducible diagrams) is constructed with the relation

$$-N(\text{art t}) + N(\mathcal{K}^{(\frac{3}{2})}) + N(v_{\frac{3}{2}}) = 1. \quad (37)$$

$N(\text{art t}), N(\mathcal{K}^{(\frac{3}{2})}), N(v_{\frac{3}{2}})$ are, respectively, the number of articulation triplets of lines, the number of $\frac{3}{2}$-irreducible skeletons, and the number of $\frac{3}{2}$-vertices in a 0-diagram. The proof is again immediate by induction. The sum of the contributions of all distinct 0-diagrams (calculated with Rule C) where one *distinguishes*, respectively, one *articulation triplet*, one *$\frac{3}{2}$-irreducible skeleton*, or one *$\frac{3}{2}$-vertex* is then constructed by generating such distinguished diagrams out of their distinguished elements through repeated insertions. Relation (37), which allows us to reestablish the proper weight for a 0-diagram through the proper combination of distinguished diagrams, yields

$$\mathcal{K}^{(1)} = -\tfrac{1}{2}(1/3!)C_{\frac{3}{2}}(1, 2, 3)\tilde{G}_1(1, \bar{1})\tilde{G}_1(2, \bar{2})\tilde{G}_1(3, \bar{3})C_{\frac{3}{2}}(\bar{1}, \bar{2}, \bar{3}) - (1/3!)v_{\frac{3}{2}}(1, 2, 3)G_{\frac{3}{2}}(1, 2, 3) + \mathcal{K}^{(\frac{3}{2})}\{C_{\frac{3}{2}}\}. \quad (38)$$

Combining (38) with (29), (31), we obtain

$$F^{(\frac{3}{2})}\{C_{\frac{3}{2}}\} = \tfrac{1}{2}\epsilon\,\text{tr}\,[\ln \tilde{G}_1 + \sigma'\omega'\tilde{G}_1(1', 1)] - \tfrac{1}{2}(1/3!)C_{\frac{3}{2}}(1, 2, 3)\tilde{G}_1(1, \bar{1})\tilde{G}_1(2, \bar{2})\tilde{G}_1(3, \bar{3}) \times C_{\frac{3}{2}}(\bar{1}, \bar{2}, \bar{3}) + \mathcal{K}^{(\frac{3}{2})}\{C_{\frac{3}{2}}\}. \quad (39)$$

The implied expression for W is now stationary under changes of $C_{\frac{3}{2}}$, the stationarity condition being identical with (34). Stationarity with respect to changes of $G_{\frac{3}{2}}$ leads to a form of (16) which is identical with the equation of motion, that is, the

first equation of a generalized Green's function hierarchy,

$$-v_{\frac{1}{2}}(1) = v_1(1, 2)G_{\frac{1}{2}}(2) + (1/2!)v_{\frac{1}{2}}(1, 2, 3)G_1(2, 3) + (1/3!)v_2(1, 2, 3, 4)G_{\frac{1}{2}}(2, 3, 4). \quad (40)$$

Stationarity with respect to changes in $\tilde{G}_1$ would lead also to a new form of (26).

V. SECOND VERTEX RENORMALIZATION: ELIMINATION OF v_2

We now turn to the last step which expresses v_2 and $F^{(2)}$ (the entropy),

$$F^{(2)} = W + v_0 + v_{\frac{1}{2}}(1)G_{\frac{1}{2}}(1) + (1/2!)v_1(1, 2)G_1(1, 2) + (1/3!)v_{\frac{1}{2}}(1, 2, 3)G_{\frac{1}{2}}(1, 2, 3) + (1/4!)v_2(1, 2, 3, 4)G_2(1, 2, 3, 4), \quad (41)$$

as functionals of $\tilde{G}_2$ (or rather C_2, a function directly related to $\tilde{G}_2$). As a result, the entropy of the system $F^{(2)}$ is reexpressed as an explicit functional of $G_{\frac{1}{2}}$, $\tilde{G}_1$, $C_{\frac{1}{2}}$, and C_2, to the exclusion of the potentials v_r. Thus the elimination of the potentials v_r (and equilibrium parameters) is completed.

A. Normal Case

For simplicity, we first consider the case of a system which has no half-integer distribution functions $G_{\frac{1}{2}} = C_{\frac{1}{2}} = 0$. This is the case for a "normal" system but also for a superfluid Fermi system. In the former case, this subsection amounts to translating into a matrix language the treatment given in reference 13. We start here with the results of Sec. III, and deal now with *1-irreducible diagrams* built only with 2-vertices $(-v_2)$ and heavy lines $(G_1 = \tilde{G}_1)$.

(1) the 2-body Green's function $G_2(1, 2, 3, 4)$ is represented by the complete class of 2-diagrams with labels 1, 2, 3, 4, on the external points.

Let $C_2(1, 2, 3, 4)$ be the connected part (cumulant) of the 2-body Green's function $G_2(1, 2, 3, 4)$ after removal of the external lines, i.e.,

$$G_2(1, 2, 3, 4) = G_1(1, 2)G_1(3, 4) + \epsilon G_1(1, 3)G_1(2, 4) + G_1(1, 4)G_1(2, 3) + \tilde{G}_2(1, 2, 3, 4), \quad (42)$$

$$\tilde{G}_2(1, 2, 3, 4) = C_2(\bar{1}, \bar{2}, \bar{3}, \bar{4}) \times G_1(\bar{1}, 1)G_1(\bar{2}, 2)G_1(\bar{3}, 3)G_1(\bar{4}, 4). \quad (43)$$

We arbitrarily group the external lines into two pairs (1, 2) and (3, 4). Diagrams such that the pairs (1, 2) and (3, 4) cannot be disconnected by cutting off one internal pair of lines are called *simple with respect to (1, 2) or (3, 4)*. Let S_2 (1, 2; 3, 4) be the contribution[17] of the 2-diagrams which are simple with respect to (1, 2) (after removal of the external lines). C_2 (1, 2; 3, 4) can be generated by iterating S_2 and we have

$$C_2(1, 2, 3, 4) = S_2(1, 2; 3, 4) + \tfrac{1}{2}S_2(1, 2; \bar{1}, \bar{2})G_1(\bar{1}, \bar{3})G_1(\bar{2}, \bar{4})C_2(\bar{3}, \bar{4}, 3, 4). \quad (44)$$

Solving for S_2 (1, 2; 3, 4) we get symbolically

$$S_2 = C_2[1 + \tfrac{1}{2}G_1G_1C_2]^{-1}. \quad (45)$$

For brevity and symmetry we shall employ the same notation as in I, that is, we shall introduce along with any function X_2 (1, 2; 3, 4) the function

$$\bar{X}_2(1, 2; 3, 4) = X_2(\bar{1}, \bar{2}; \bar{3}, \bar{4}) \times [G_1]^{\frac{1}{2}}(\bar{1}, 1)[G_1]^{\frac{1}{2}}(\bar{2}, 2)[G_1]^{\frac{1}{2}}(\bar{3}, 3)[G_1]^{\frac{1}{2}}(\bar{4}, 4). \quad (46)$$

This notation is rather formal; we shall not inquire into the definition of the square roots which will always occur squared, causing no ambiguity.

We also write

$$C_2(1, 2, 3, 4) = S_2(1, 2; 3, 4) + T_2(1, 2; 3, 4), \quad (47)$$

where, in the same symbolic notation,

$$\bar{T}_2 = \tfrac{1}{2}\bar{C}_2[1 + \tfrac{1}{2}\bar{C}_2]^{-1}\bar{C}_2. \quad (48)$$

Clearly, we could have done the same analysis by selecting the pairs (1, 3), (2, 4) or (1, 4), (2, 3), obtaining

$$C_2(1, 2, 3, 4) = S_2(1, 3; 2, 4) + T_2(1, 3; 2, 4), \quad C_2(1, 2, 3, 4) = S_2(1, 4; 2, 3) + T_2(1, 4; 2, 3). \quad (49)$$

There are three possible pairings of the four external lines for a given 2-diagram. A 2-diagram can be nonsimple with respect to *one and only one* of the pairings, and thus contributes unambiguously to the corresponding T_2 functional. Finally, a 2-diagram can be simple with respect to all pairings of the external lines. It is that class of diagrams which we analyze now.

Consider a 0-diagram and a pair of lines which we denote by (l, m) in this diagram. By cutting (l, m) open, we obtain a 2-diagram with external lines denoted by (l', l'', m', m''). We have seen that a 2-diagram is necessarily simple with respect to two pairings of the external lines, and possibly nonsimple with respect to the third pairing. Consequently, if there exists a set of internal pairs of lines (j_1, k_1), (j_2, k_2) $\cdots$ $(j_{\mu-1}, k_{\mu-1})$ such that when any one of those pairs is cut open, two external

[17] Notice that $S_2(1, 2; 3, 4)$ is ϵ-symmetrical under the exchanges $(1 \leftrightarrow 2, 3 \leftrightarrow 4)$ but [contrary to $C_2(1, 2, 3, 4)$] this function is *not* ϵ-symmetrical under $1 \leftrightarrow 3$, etc.

pairs of lines become disconnected, this set is unique. The complete set of such pairs $(j_1, k_1) \cdots (j_{\mu-1}, k_{\mu-1})$ and (l, m) are said to constitute a *cycle of pairs of multiplicity* μ. In this way, all possible pairs of lines of a 0-diagram can be *unambiguously* distributed into cycles of pairs of multiplicity μ ($\mu \geq 1$). Removal of the μ pairs of lines of a cycle separates the 0-diagram into μ disconnected parts. When $\mu = 2$, and one and only one of the two disconnected parts is a single 2-vertex, the cycle is said to be *trivial*.

A 2-irreducible 0-diagram is defined as containing only cycles with $\mu = 1$ and trivial cycles with $\mu = 2$. The contribution of such a class of 0-diagrams (calculated with Rule C, where $G_{\frac{1}{2}} = 0$) is called $\mathcal{K}^{(2)}\{-v_2\}$, stressing the functional dependence upon the 2-body potential.

The class of 2-irreducible 2-diagrams is obtained by suppressing one 2-vertex in all possible ways in a 2-irreducible 0-diagram and its contribution properly given by

$$K_2\{\bar{1}, \bar{2}, \bar{3}, \bar{4}; -v_2\}G_1(\bar{1}, 1)G_1(\bar{2}, 2)G_1(\bar{3}, 3)G_1(\bar{4}, 4) = -4![\delta/\delta v_2(1, 2, 3, 4)]\mathcal{K}^{(2)}\{-v_2\}. \quad (50)$$

Clearly a 2-irreducible 2-diagram is simple with respect to all pairings of the external lines, and remains so when each one of its 2-vertices $(-v_2)$ is replaced by the complete class of (connected) 2 diagrams, each stripped of its external lines (C_2). Indeed, in this fashion, and except for the diagram of first order in v_2, we generate all 2-diagrams which are *simple with respect to all pairings of the external lines,* and their contribution (after removal of the external lines) is therefore

$$-v_2(1, 2, 3, 4) + K_2\{1, 2, 3, 4; C_2\}.$$

We can now express S_2 (1, 2; 3, 4) as

$$S_2(1, 2; 3, 4) = T_2(1, 3; 2, 4) + T_2(1, 4; 2, 3) - v_2(1, 2, \bar{3}, 4) + K_2\{1, 2, 3, 4; C_2\}. \quad (51)$$

T_2 is expressed as a functional of C_2 with (48) [and (46)]. Equation (51) is thus immediately solved for v_2 as a functional of C_2 (and G_1),

$$v_2(1, 2, 3, 4) = T_2(1, 2; 3, 4) + T_2(1, 3; 2, 4) + T_2(1, 4; 2, 3) - C_2(1, 2, 3, 4) + K_2\{1, 2, 3, 4; C_2\}, \quad (52)$$

thus inverting relation (9) considered as giving $\tilde{G}_2$ (or rather C_2) as a functional of v_2. K_2 and $\mathcal{K}^{(2)}$ are now represented by all distinct 2-irreducible 2-diagrams (stripped of their external lines), and 0-diagrams, respectively. These diagrams are built with heavy 2 vertices (2-correlation functions) and heavy lines (propagators)

$\mathcal{K}^{(2)}\{C_2\}$ =  +  + ⋯, (53)

$K_2\{1,2,3,4;C_2\}$ =  +  + ⋯ (54)

and

$$K_2\{\bar{1}, \bar{2}, \bar{3}, \bar{4}; C_2\}G_1(\bar{1}, 1)G_1(\bar{2}, 2)G_1(\bar{3}, 3)G_1(\bar{4}, 4) = 4![\delta/\delta C_2(1, 2, 3, 4)]\mathcal{K}^{(2)}\{C_2\}. \quad (55)$$

Associated contributions are calculated with Rule $(\bar{E})$ differing from (A) by

($\bar{E}$i) With each heavy 2-vertex, associate C_2.
($\bar{E}$ii) With each heavy line, associate $G_1(= \tilde{G}_1)$.

(2) In Appendix B [Eq. (B11)] we show that, for any 1-irreducible 0-diagram, the following relation holds:

$$\sum_\gamma [1 - N_\gamma(\pi) + N_\gamma((\pi)^2)] - N(\text{art q}) + N(\mathcal{K}^{(2)}) + N(v_2) = +1. \quad (56)$$

$N_\gamma(\pi)$ is the number of pairs, $N_\gamma((\pi)^2)$ the number of pairs of pairs in a cycle γ_μ, that is, μ and $\frac{1}{2}[\mu(\mu-1)]$, respectively. The summation extends over all cycles of the diagrams; thus $\sum_\gamma 1$ and $\sum_\gamma N_\gamma(\pi)$ are, respectively, the number of cycles and the number of pairs of lines in the diagram. $N(\text{art q})$ is the number of articulation quartets of lines (sets of four lines which when cut leave the diagram separated into two disconnected parts), $N(\mathcal{K}^{(2)})$ the number of 2-irreducible skeletons, $N(v_2)$ the number of 2-vertices. The proper weight for the class of 0-diagrams is then constructed by considering ensembles of 0-diagrams with one element distinguished (one cycle, one pair, one pair of pairs in a cycle, one articulation quartet, one 2-irreducible skeleton, one 2-vertex; that is, ensembles of 0-diagrams weighted by the various weights introduced above), and by combining such ensembles as indicated by relation (56). In turn, the contribution of such distinguished 0-diagrams is obtained by generating them out of their distinguished element through repeated insertions. The only delicate point concerns the contributions corresponding to the three terms inside the bracket of Eq. (56), which we now examine. Notice that the bracket vanishes for cycles of multiplicity $\mu = 1, 2$.

The ensemble of 0-diagrams with one cycle

distinguished contributes

$$\tfrac{1}{2}\,\mathrm{tr}\sum_{\mu=1}^{\infty}\frac{1}{\mu}\left[\frac{g}{2}\,S_2G_1G_1\right]^{\mu}\Big|_{g=1} = A(g)|_{g=1}, \qquad (57)$$

where the bracketed term is to be considered as a matrix

$$S_2(1, 2; \bar{1}, \bar{2})G_1(\bar{1}, 1')G_1(\bar{2}, 2').$$

The ensembles of 0-diagrams with one pair or one pair of pairs distinguished in a cycle contribute, respectively,

$$(d/dg)A(g)|_{g=1}, \qquad \tfrac{1}{2}(d/dg)^2A(g)|_{g=1}. \qquad (58)$$

Strictly speaking, the numerical factor in Eq. (57) is only correct for $\mu \geq 3$, but in view of the remark made above, the factors of the terms in $\mu = 1, 2$ are irrelevant if we use the same $A(g)$ for (57) and (58). Using Eq. (45), we have

$$[1 - d/dg + \tfrac{1}{2}(d/dg)^2]A(g)|_{g=1}$$
$$= \tfrac{1}{2}\,\mathrm{tr}\,\{\ln[1 + \tfrac{1}{2}\bar{C}_2] - \tfrac{1}{2}\bar{C}_2 + \tfrac{1}{2}(\tfrac{1}{2}\bar{C}_2)^2\}. \qquad (59)$$

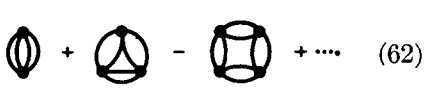
Combining with the other three contributions we get

$$\mathcal{K}^{(1)} = \tfrac{1}{2}\,\mathrm{tr}\,\{\ln[1 + \tfrac{1}{2}\bar{C}_2] - \tfrac{1}{2}\bar{C}_2 + \tfrac{1}{2}(\tfrac{1}{2}\bar{C}_2)^2\}$$
$$- \tfrac{1}{2}\,\mathrm{tr}\,\{(1/4!)(\bar{C}_2)^2\} + \mathcal{K}^{(2)}\{G_1, C_2\}$$
$$- (1/4!)v_2(1, 2, 3, 4)G_2(1, 2, 3, 4). \qquad (60)$$

Using (41) we get the entropy[18]

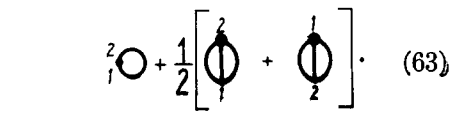
$$F^{(2)} = \tfrac{1}{2}\epsilon\,\mathrm{tr}\,\{\ln G_1 + \omega'\sigma^{\epsilon}G_1(1', 1)\}$$
$$+ \tfrac{1}{2}\,\mathrm{tr}\,\{\ln[1 + \tfrac{1}{2}\bar{C}_2] - \tfrac{1}{2}\bar{C}_2 + \tfrac{1}{2}[\tfrac{1}{2}\bar{C}_2]^2$$
$$- (1/4!)[\bar{C}_2]^2\} + \mathcal{K}^{(2)}\{G_1, C_2\}. \qquad (61)$$

Notice that diagrammatically the contribution $\mathcal{L}\{G_1, G_2\}$ of the second trace appearing in (61) is simply represented by the *alternate* series (calculated with Rule $\bar{E}$)

$\mathcal{L}\{G_1, G_2\} =$  (62)

Naturally, Eq. (61) implies an expression for W which is stationary under variations of C_2, the stationarity condition being identical with (52). Stationarity with respect to G_1 leads to the equation of motion (first equation of the Green's function hierarchy for normal systems)

$$[G_1^0]^{-1}(1, 2) - [G_1]^{-1}(1, 2) =$$

 (63)

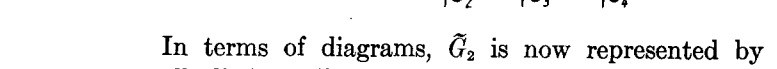
To obtain (63) we use the fact that, except for the first trace term, Eq. (61) only involves G_1 through the combination $\bar{C}_2$; use of the stationarity property with respect to C_2 leads to the compact form (63) where the appearance of an extra factor $\frac{1}{2}$ is to be noticed.

B. General Case

Now we generalize the treatment given above to the case where $G_{\frac{1}{2}}$ and $C_{\frac{3}{2}}$ are nonvanishing (superfluid Bose systems). The starting point is the result obtained in Sec. IV, and from now on we consider *only* $\frac{3}{2}$*-irreducible diagrams,* built with heavy half lines or lines ($G_{\frac{1}{2}}$ or $\tilde{G}_1$) heavy $\frac{3}{2}$-vertices ($C_{\frac{3}{2}}$) and 2-vertices ($-v_2$), and calculated with Rule (D).

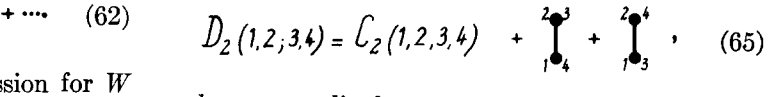
The cumulant part of the 2-body Green's function stripped of its external lines is written

$$\tilde{G}_2(\bar{1}, \bar{2}, \bar{3}, \bar{4})[\tilde{G}_1]^{-1}(\bar{1}, 1)[\tilde{G}_1]^{-1}(\bar{2}, 2)$$
$$\times [\tilde{G}_1]^{-1}(\bar{3}, 3)[\tilde{G}_1]^{-1}(\bar{4}, 4) =$$

 (64)

In terms of diagrams, $\tilde{G}_2$ is now represented by all distinct ($\frac{3}{2}$-irreducible) 2-diagrams. The contribution of the three diagrams of second order in $C_{\frac{3}{2}}$ has been singled out in (64) for they play a particular role, analogous to the one played by a 2-vertex; for that reason we call those special 2-diagrams, *pseudo 2-vertices.*

(1) Equations (44) to (48) are immediately generalized. It suffices to let the quantity D_2 (12; 34) defined by

 (65)

and correspondingly,

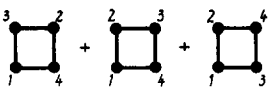
$$\bar{D}_2(1, 2; 3, 4) = D_2(\bar{1}, \bar{2}; \bar{3}, \bar{4})[\tilde{G}_1]^{\frac{1}{2}}(\bar{1}, 1)$$
$$\times [\tilde{G}_1]^{\frac{1}{2}}(\bar{2}, 2)[\tilde{G}_1]^{\frac{1}{2}}(\bar{3}, 3)[\tilde{G}_1]^{\frac{1}{2}}(\bar{4}, 4) \qquad (66)$$

play the role played by $C_2(1, 2, 3, 4)$ and $\bar{C}_2(1, 2, 3, 4)$ in Eqs. (45) and (48).

However here, there exist three special 2-diagrams which are nonsimple with respect to *two* different pairings of the external lines (instead of only one). They are the 2-diagrams built with two pseudo 2-vertices and they contribute (after removal of

[18] Equation (61) is a rewriting in matrix form of Eq. (2.67) of Ref. 13, or of Eq. (5.6) of the preliminary report (J. Math. Phys. 4, 255 (1963)). Notice that a factor 1/4 is missing in front of the $\sum$ of the last mentioned equation; in the same report, the left hand side of Eq. (5.2) (of which Eq. (52) is a rewriting in matrix form) ought to be read as symmetrized.

their external lines)

$Q_2(1,2,3,4)$ =  (67)

Each one of these three 2-diagrams belongs to *two* different T_2 functionals, and the sum of all nonsimple 2-diagrams (each one being counted once) is

$$T_2(1,2;3,4) + T_2(1,3;2,4) + T_2(1,4;2,3) - Q_2(1,2,3,4), \quad (68)$$

with

$$D_2(1,2;3,4) = S_2(1,2;3,4) + T_2(1,2;3,4). \quad (69)$$

In the symbolic notation used previously, we have

$$\bar{T}_2 = \bar{D}_2[1 + \tfrac{1}{2}\bar{D}_2]^{-1}\tfrac{1}{2}\bar{D}_2. \quad (70)$$

The analysis of the 2-diagrams which are simple with respect to all pairings of the external lines is only slightly changed. All pairs of a 0-diagram can be distributed unambiguously into cycles of multiplicity $\mu(\mu \geq 1)$ except when the 2-diagram obtained by cutting open one pair belongs to Q_2. That is the case for the tetrahedron build with four (heavy) $\frac{3}{2}$-vertices. A cycle with $\mu = 2$ is *trivial* if one and only one of the two disconnected parts obtained by cutting open the two pairs of lines of the cycle is a single 2-vertex or a pseudo 2-vertex. By definition, a *2-irreducible 0-diagram* contains only cycles with $\mu = 1$, or trivial cycles with $\mu = 2$, and $\mathcal{K}^{(2)}$ is the sum of the contributions of such diagrams. The class of 2-irreducible 2-diagrams K_2 is defined as above (in (50), with $\tilde{G}_1$ in place of G_1).

The class of 2-diagrams which are simple with respect to all pairings of the external lines (with the exception of the 2-diagrams which reduce to a single 2-vertex or pseudo 2-vertex) is obtained as follows. Each 2-vertex $(-v_2)$ in each 2-irreducible 2-diagram is replaced by all members of the class of connected 2-diagrams (stripped of external lines) which do not reduce to a single pseudo 2-vertex (C_2); the contribution of the class constructed thereby is

$$K_2\{\bar{1},\bar{2},\bar{3},\bar{4};C_2\}\tilde{G}_1(\bar{1},1)\tilde{G}_1(\bar{2},2)\tilde{G}_1(\bar{3},3)\tilde{G}_1(\bar{4},4).$$

We thus obtain

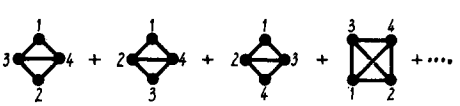
$$\begin{aligned} S_2(1,2;3,4) &= T_2(1,3;2,4) + T_2(1,4;2,3) \\ &\quad - Q_2(1,2,3,4) + K_2\{1,2,3,4;C_2\} \\ &\quad - v_2(1,2,3,4) + C_{\frac{3}{2}}(1,5,3)\tilde{G}_1(5,6)C_{\frac{3}{2}}(2,6,4) \\ &\quad + C_{\frac{3}{2}}(1,5,4)\tilde{G}_1(5,6)C_{\frac{3}{2}}(3,6,2). \end{aligned} \quad (71)$$

With Eqs. (69) and (65), this equation determines v_2 as a functional of C_2 (and $C_{\frac{3}{2}}$),

$$\begin{aligned} v_2(1,2,3,4) &= T_2(1,2;3,4) + T_2(1,3;2,4) \\ &\quad + T_2(1,4;2,3) - Q_2(1,2,3,4) \\ &\quad - C_2(1,2,3,4) + K_2\{1,2,3,4;C_2\}, \end{aligned} \quad (72)$$

T_2 being expressed in terms of C_2 (and $C_{\frac{3}{2}}$)through Eqs. (70), and (65). Here K_2 and $\mathcal{K}^{(2)}$ are represented by 2-irreducible 2-diagrams (stripped of their external lines), and 0-diagrams, respectively. These diagrams are built with heavy $\frac{3}{2}$- and 2-vertices ($\frac{3}{2}$- and 2-correlation functions) and heavy lines (propagators),

$\mathcal{K}^{(2)}\{C_{3/2},C_2\}$ =  (73)

$K_2\{1,2,3,4;C_{3/2},C_2\}$ =  (74)

K_2 and $\mathcal{K}^{(2)}$ are related through Eq. (55) (with $\tilde{G}_1$ in place of G_1), associated contributions being calculated with the *final Rule* (E) which differs from Rule (A) through

(Ei) With each heavy $\frac{3}{2}$- or 2-vertex, associate $C_{\frac{3}{2}}$ or C_2.

(Eii) With each heavy line or half line, associate $\tilde{G}_1$ or $G_{\frac{1}{2}}$.

(Eiv) Identical to (Biv).

(2) In the general case, the extra topological structure introduced by $\frac{3}{2}$-vertices (which resulted in the appearance of the term Q_2 for 2-diagrams) alters relation (56) in the following ways [Appendix B, relation (B16)]:

(i) Let $N(p_2)$ be the number of pseudo 2-vertices. On the left-hand side, $N(v_2)$ is replaced by $N(v_2) + N(p_2)$.

(ii) Let $N(Q_2)$ be the number of ways that one can separate out of a 0-diagram a 2-diagram con-

tributing to Q_2, by cutting open four lines. On the left-hand side of (56) one adds the extra term $-N(Q_2)$.

(iii) The new topological relation holds for all $\frac{3}{2}$-irreducible 0-diagrams except (a) the tetrahedron, (b) the (twisted) triangular prism, and (c) the cube (twisted or not); these diagrams are built with heavy $\frac{3}{2}$-vertices for corners. A relation closely analogous to (60) follows for $\mathcal{K}^{(\frac{3}{2})}$. In the first three terms which can be calculated together as above in (59), $\bar{D}_2$ now plays the role played by $\bar{C}_2$ in (60); furthermore, five extra terms reflect the comments (ii) and (iii) above.

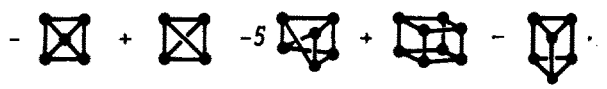
Using (41) we get for the entropy

$$F^{(2)} = \tfrac{1}{2}\epsilon \operatorname{tr}\{\ln \tilde{G}_1 + \sigma' \omega' \tilde{G}_1(1', 1)\}$$

$$- \tfrac{1}{2}(1/3!)C_{\frac{3}{2}}(1, 2, 3)\tilde{G}_1(1, \bar{1})\tilde{G}_1(2, \bar{2})\tilde{G}_1(3, \bar{3})C_{\frac{3}{2}}(\bar{1}, \bar{2}, \bar{3})$$

$$+ \tfrac{1}{2} \operatorname{tr}\{\ln [1 + \tfrac{1}{2}\bar{D}_2] - \tfrac{1}{2}\bar{D}_2 + \tfrac{1}{2}[\tfrac{1}{2}\bar{D}_2]^2\}$$

$$- \tfrac{1}{2} \operatorname{tr}(1/4!)(\bar{C}_2)^2 + \mathcal{K}^{(2)}\{\tilde{G}_1, C_{\frac{3}{2}}, C_2\} \qquad (75)$$



(i) The three terms in $\bar{D}_2$ in Eq. (75) differ from the corresponding terms in (61) in that $\bar{D}_2$ has now replaced $\bar{C}_2$. The following term in (75) is unchanged. In (61) it corresponded to $-N(\text{art q})$ the number of articulation quartets; here, it corresponds to $-[N\ (\text{art q}) - N(p_2)]$, the number of articulation quartets excluding the ones which connect one pseudo 2-vertex to the rest of the diagram.

(ii) The first diagram has the same contribution as the sum of all $\frac{3}{2}$-irreducible 0-diagrams [calculated with rule (D)] weighted by $-N(Q_2)$.

(iii) The exceptional contributions to $\mathcal{K}^{(\frac{3}{2})}$, that is, those for which the topological theorem of Appendix B is not supposed to hold, have to be added; account has to be taken of the fact that they may appear in the terms commented on in (i), (ii) with a wrong weight.

(a) The tetrahedron ($S_V = 24$) is counted nowhere else and is simply added.

(b) The twisted triangular prism ($S_V = 72$) is counted six times in (i) [in the $(\bar{D}_2)^3$ term of the ln term] and so is subtracted five times.

(c) The cube ($S_V = 48$) and the twisted cube ($S_V = 16$) are counted, respectively, six and five times in (i), six and four times and with a minus sign in (ii). So only the cube is added once.

(d) Finally the triangular prism ($S_V = 12$) which is not $\frac{3}{2}$-irreducible, [and should not contribute to $\mathcal{K}^{(\frac{3}{2})}$] appears once in the $(\bar{D})^3$ term of the ln and is subtracted.

Equation (75) for the functional $F^{(2)}\{\tilde{G}_1, C_{\frac{3}{2}}, C_2\}$ is perhaps the most important result of this paper. It shows that the entropy is a functional of $\tilde{G}_1$, $C_{\frac{3}{2}}$, C_2 (not of $G_{\frac{3}{2}}$), and explicitly exhibits that functional as a power-series expansion.

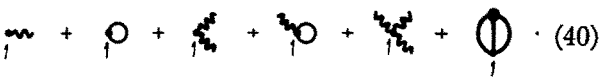
The corresponding expression for

$$W = F^{(2)}\{\tilde{G}_1, C_{\frac{3}{2}}, C_2\} - v_0 - \sum_{2\nu=1}^{4} \frac{1}{(2\nu)!} v_\nu G_\nu,$$

where the v_ν's are kept fixed, is *stationary under independent variations of* $G_{\frac{3}{2}}$, $\tilde{G}_1$, $C_{\frac{3}{2}}$, *and* C_2. Here the G_ν's are to be expressed in terms of the cumulants $\tilde{G}_\nu$ and $\tilde{G}_{\frac{3}{2}}$; $\tilde{G}_2$ are related to $C_{\frac{3}{2}}$, C_2 through Eqs. (32) and (64). The stationarity conditions provide a set of *four coupled equations* determining the four potentials v_ν as power series in the functions $G_{\frac{3}{2}}$, $\tilde{G}_1$, $C_{\frac{3}{2}}$, C_2. Equivalently this set of equations is obtained by expressing that the functional $F^{(2)}$ is stationary under variations of $G_{\frac{3}{2}}$, $\tilde{G}_1$, $C_{\frac{3}{2}}$, C_2, with the constraints of constant average energy and particle number; (the function $G_{\frac{1}{2}}$ appears only through the constraint equations).

The two equations resulting from the variations of the propagators ($G_{\frac{3}{2}}$, $\tilde{G}_1$) are the first two equations of a Green's function hierarchy for the systems considered here. The first equation is identical with (40), that is

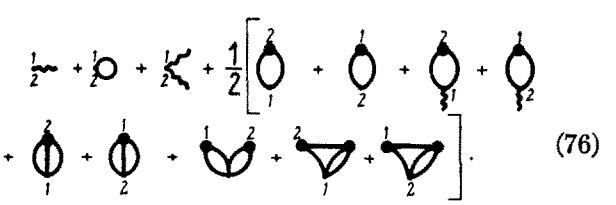
$$v_{\frac{3}{2}}(1) =$$



(40)

The second equation is obtained by remarking that $\tilde{G}_1$ occurs [except in the first line of (75)] through the combinations $\tilde{C}_{\frac{3}{2}} \equiv C_{\frac{3}{2}}(\tilde{G}_1\tilde{G}_1\tilde{G}_1)^{\frac{1}{2}}$ and $\tilde{C}_2 \equiv C_2(\tilde{G}_1\tilde{G}_1\tilde{G}_1\tilde{G}_1)^{\frac{1}{2}}$; it reads

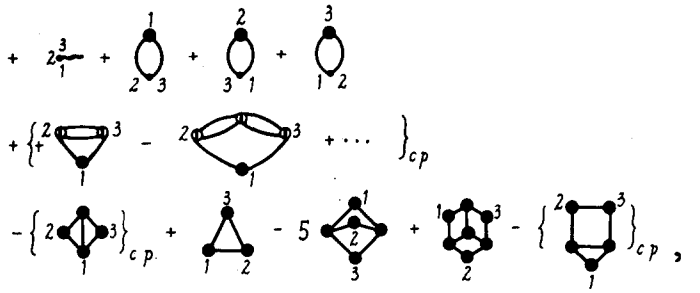
$$[G_1^0]^{-1}(1, 2) - [\tilde{G}_1]^{-1}(1, 2) =$$



(76)

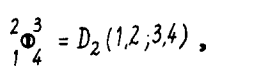
There are also two equations resulting from the variations of the correlation functions ($C_{\frac{3}{2}}$, C_2). Variation of C_2 leads to Eq. (72). Variation of $C_{\frac{3}{2}}$ leads to

$$v_{\frac{3}{2}}(1,2,3) = -C_{\frac{3}{2}}(1,2,3) + 3![\delta/\delta C_{\frac{3}{2}}(\bar{1},\bar{2},\bar{3})]\,\mathcal{K}^{(2)}\{C_{\frac{3}{2}}, C_2\}\tilde{G}_1(\bar{1},1)\tilde{G}_1(\bar{2},2)\tilde{G}_1(\bar{3},3)$$



(77)

where

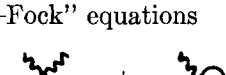


and where the subscript cp means that the $\frac{3}{2}$-diagram shown represents the three distinct diagrams obtained by circular permutations of the arguments 1, 2, 3.

Equations (40), (76), (72), and (77) together exhibit explicitly the power series for each potential v_r in terms of $G_{\frac{1}{2}}$, $\tilde{G}_1$, $C_{\frac{3}{2}}$, and C_2. Given the potentials ($v_{\frac{1}{2}}$ and $v_{\frac{3}{2}}$ vanish for the systems specifically considered here), these four (matrix) equations determine, self-consistently, $G_{\frac{1}{2}}$, $\tilde{G}_1$, $C_{\frac{3}{2}}$, and C_2. For a normal system or a superfluid Fermi system, they reduce to two equations with $G_{\frac{1}{2}} = C_{\frac{3}{2}} = 0$. "Self-consistent" approximations can be generated by properly truncating the entropy functional (75) (or the corresponding W functional). The four stationarity conditions then yield an approximate and self-consistent set of four equations.

VI. THE SIMPLEST EXAMPLE

In view of the length of the present paper and the discussion elsewhere of the equations derived here, we shall content ourselves with making contact with reality by commenting on the lowest-order approximation. This approximation (the generalized Hartree–Fock) results from linearizing the W functional with respect to the interaction, which amounts to setting $C_{\frac{3}{2}} = C_2 = 0$. The stationarity conditions with respect to $G_{\frac{1}{2}}$ and $\tilde{G}_1$ then yield a pair of generalized "Hartree–Fock" equations

$$v_1(1,\bar{1})G_{\frac{1}{2}}(\bar{1}) =$$  (78)

$$[G_1^0]^{-1}(1,2) - [\tilde{G}_1]^{-1}(1,2) =$$  (79)

Equations (78) and (79), together with the condition expressing that the system has a fixed average number of particles

$$N = (\tfrac{1}{2}\epsilon)\,\mathrm{tr}\,[\sigma^\epsilon\tau^3(\tilde{G}_1 + G_{\frac{1}{2}}G_{\frac{1}{2}})] - \partial v_0/\partial(\beta\mu), \qquad (80)$$

determine $G_{\frac{1}{2}}$, $\tilde{G}_1$, and the Lagrange parameter μ (the chemical potential). The term $-\partial v_0/\partial(\beta\mu)$ is a trivial constant resulting from the use of an ϵ-symmetrized Hamiltonian Eq. (1). The term in $G_{\frac{1}{2}}$ is the condensate contribution. For a translationally invariant system, Eq. (80) may be written, using the representation of Eq. (5), as

$$N = \frac{\Omega}{2}\int\frac{d\mathbf{k}}{(2\pi)^3}\sum_\omega([G_1(k,\omega)]_{\alpha=2,\alpha'=1} + \epsilon[G_1(k,\omega)]_{\alpha=1,\alpha'=2}) - \partial v_0/\partial(\beta\mu),$$

where Ω is the volume of the system.

For fermions, Eq. (79), where the first diagram has a vanishing contribution, gives rise to the Gorkov[19] equations with generalized spin pairings. It reduces to the BCS[20] equations in matrix form when one makes the additional assumption that the system is translationally invariant and the spin dependence of the 2 × 2 matrices in spin space has only singlet components. It has been discussed in the more general case by Balian and Werthamer.[21] For bosons, in the case of homogeneous systems, Eqs. (78) and (79) reduce to those discussed at vanishing temperature by Girardeau and Arnowitt,[22] and at finite temperature by Tolmachev.[23] Neglecting the second diagram in Eqs. (78) and (79) leads to the Bogoliubov[24] approximation, and for non-

[19] L. P. Gor'kov Zh. Eksperim i Teor. Fiz. **34**, 735 (1958) [English transl.: Soviet Phys.—JETP **7**, 505 (1958)].
[20] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **108**, 1175 (1957).
[21] R. Balian and N. Werthamer, Phys. Rev. **131**, 1553 (1963).
[22] M. Girardeau, R. Arnowitt, Phys. Rev. **113**, 755 (1959).
[23] V. Tolmachev, Dokl. Akad. Nauk SSSR **135**, 41, 825 (1960) [English transl.: Soviet Phys.—Doklady **5**, 1190, 1267 (1961)].
[24] N. Bogoliubov, Zh. Techn. Fiz. **9**, 23 (1947).

homogeneous systems, to the type of equation discussed by Gross and Pitaevskii.[25]

It need hardly be remarked that even for superfluid systems the Hartree–Fock approximation does not require the extensive analysis performed here. The advantages of a renormalized formulation: to make exact statements about (strongly) interacting systems, and to calculate in a controlled manner beyond the Hartree–Fock approximation, are being exploited.

VII. CONCLUSION AND QUESTIONS FOR FURTHER INVESTIGATION

In this paper we have derived a functional (75) of the functions $G_\nu(\nu = \frac{1}{2}, 1, \frac{3}{2}, 2)$. The requirement that this functional be stationary under variations of the G_ν's (with the constraint of fixed average energy and particle number) leads to a set of four coupled equations for the ("time"-dependent or ω-dependent) distribution functions G_ν. They are the equations that express the unrenormalized potentials v_ν in terms of the renormalized quantities G_ν. For the values of G_ν which solve this set of equations, $F^{(2)}$ is equal to the entropy of the system.

The existence of such a functional raises many questions which require further investigation. These questions fall roughly into three categories. The first concerns the possibility of making the principle a maximal one. Can one, by restricting the class of functions G_ν (presumably by imposing constraints corresponding to physical requirements), make the principle a maximal one rather than a stationary one? If so, can this maximal principle be related to the well-known one for the full density matrix? Furthermore, since the functional derived does not depend upon the equilibrium parameters, is it possible to give it a physical meaning elsewhere than at its stationary point? More particularly, in nonequilibrium situations, to what extent does such a functional play the role of a generalized Boltzmann H function?

The second set of questions concerns the possibility of discussing metastability. We should like to know whether this functional for the entropy in terms of the first four distribution functions gives rise to equations which may be satisfied by several sets of distribution functions. If there is more than one, do the other solutions correspond to metastable configurations? And if there is only one, as is likely if the first set of questions is answered affirmatively, are there still several solutions when approximations are made to the full functional? (We know for example that the Hartree–Fock equations admit more solutions than higher approximations do.) It is tempting to speculate on the parallel between the evolution in time of a metastable system and the expansion of the entropy functional. In both descriptions, higher-order terms involve more collisions. Could it be that the approximations made in treating metastable states (approximations which are not valid for very long times) are associated with truncations of the entropy functional that admit metastable states as solutions? And likewise, might not the same terms, which in sufficiently accurate expressions for the functional rule out the metastable states as stationary solutions, be responsible for the decay of these states at sufficiently long times?

In connection with this question, it should be noted that the transformation $v_\nu \leftrightarrow G_\nu$ is not necessarily well defined, and the singularities of the Jacobian (which coincide with the vanishing of the quadratic form of the second variation of $F^{(2)}$ with respect to G_ν) might provide some information about the appearance of instabilities.

This leads naturally to and cannot be dissociated from the third set of questions. To what extent is one justified in considering truncations of the functional? What properties must one require of sensible truncations? And what properties will approximations derived in this way have? We hope to discuss these questions at a later date.

ACKNOWLEDGMENT

We would like to thank the Physics Department of the University of California at Berkeley for the stimulating atmosphere provided us during the latter stages of the work on this and the previous paper. One of the authors (CD) would also like to acknowledge the hospitality extended him by the University of Washington in Seattle.

APPENDIX A. DIAGRAMS

In this appendix we derive the diagrammatic representation of the perturbation expansion, and several properties used in diagram resummations throughout the text. The general rules applying to all diagrams introduced in the text, are summarized in Sec. 6 of this appendix.

1. Algebraic Expansions and Wick Theorem

We consider the grand partition function e^W; let $U(z, 0)$ be the "evolution operator" between 0 and z,

[25] E. Gross, Ann. Phys. (N. Y.) 4, 57 (1958); L. Pitaevskii, Zh. Eksperim i Teor. Fiz. 35, 1271 (1958) [English transl.: Soviet Phys.—JETP 8, 888 (1959)].

$$U(z, 0) = e^{zH_0}e^{-zH},$$

or for a "time"-dependent Hamiltonian $H(z) = H_0 + V(z)$,

$$U(z, 0) = e^{zH_0}T\left[\exp -\int_0^z dz' H(z')\right],$$

where T is the time-ordering operator. The grand partition function is then obtained through

$$e^{W-W_0} = \operatorname{tr}[e^{-H_0}U(1, 0)]/\operatorname{tr} e^{-H_0} = \langle U(1, 0)\rangle_0, \quad \text{(A1)}$$

that is,

$$e^{W-W_0} = \sum_{p=0}^{\infty}\frac{(-)^p}{p!}\int_0^1 \cdots \int_0^1 dz_1 \cdots dz_p \times T\langle V_I(z_1) \cdots V_I(z_p)\rangle_0. \quad \text{(A2)}$$

Here $V_I(z)$ is the perturbation term in the interaction representation, the noninteracting part being H_0,

$$V_I(z) = e^{zH_0}V(z)e^{-zH_0}.$$

$V_I(z)$ contains the potentials introduced in (1) (possibly with an explicit z dependence). Typically we have

$$V_I(z) = \cdots + \frac{1}{4!}v_2(j, k, l, m) \times \Psi_I(j)\Psi_I(k)\Psi_I(l)\Psi_I(m), \quad \text{(A3)}$$

with

$$\Psi_I(j) \equiv \Psi_I(x_j, z_j) = e^{z_jH_0}\Psi(x_j)e^{-z_jH_0},$$

and for a potential with no explicit $z = z_j$ dependence,

$$v_2(j, k, l, m) = v_2(x_j, x_k, x_l, x_m) \times \delta(z_k - z_j)\delta(z_l - z_j)\delta(z_m - z_j). \quad \text{(A4)}$$

The evaluation of the average value of a T product of Ψ_I operators is straightforward and follows closely the type of analysis introduced by Wick.[26] To evaluate

$$T\langle\Psi_I(1)\Psi_I(2)\Psi_I(3) \cdots\rangle_0,$$

we introduce a pairing of the Ψ_I operators; two paired operators, denoted by

$$T\langle\cdots \overbrace{\Psi_I(j) \cdots \Psi_I(j')} \cdots\rangle_0,$$

constitute a T *contraction*. A *complete* set of T contractions is obtained by pairing all operators, for example,

$$T\langle\overbrace{\Psi_I(1)\overbrace{\Psi_I(2)\Psi_I(3)}\Psi_I(4)} \cdots\rangle_0.$$

When the number of operators Ψ_I is odd, there are no complete sets of contractions.

Theorem: The average value of a T product of Ψ_I operators is equal to the sum of all possible complete sets of T contractions. The value of each contraction is given by

$$\overbrace{\Psi_I(j)\Psi_I(j')} = T\langle\Psi_I(j)\Psi_I(j')\rangle_0. \quad \text{(A5)}$$

Each complete set of T contractions bears a sign ϵ^P, where P is the order of the permutation necessary to bring the members of each contracted pair next to each other.

(a) *Ordinary Products of Operators*

Consider the unperturbed Hamiltonian

$$H_0 = v_0 + \tfrac{1}{2}\Psi(x_1)v_1(x_1, x_2)\Psi(x_2), \quad \text{(A6)}$$

where x_1 stands for $\mathbf{r}_1$ and α_1. It is sufficient to recognize that we have

$$\Psi_I(x_1, z_1) = e^{z_1H_0}\Psi(x_1)e^{-z_1H_0} = S_{z_1}(x_1, x_2)\Psi(x_2), \quad \text{(A7)}$$

where $S_z(x_1,x_2)$ is a c-number matrix. This property may be exhibited either by direct diagonalization of (A6), or by examination of the equation of motion of $\Psi_I(x_1, z_1)$. The matrix S_z has the property

$$S_{z+z'} = S_zS_{z'}. \quad \text{(A8)}$$

From (A7) it follows that the ϵ commutator

$$[\Psi_I(1), \Psi_I(2)]_{-\epsilon} = \Psi_I(1)\Psi_I(2) - \epsilon\Psi_I(2)\Psi_I(1) \quad \text{(A9)}$$

is also c number. Given (A7) and (A9), the proof of Gaudin[27] holds.

Consider the product of operators

$$\Psi_I(1)\Psi_I(2) \cdots \Psi_I(n).$$

By ϵ-commuting $\Psi_I(1)$ to the right, we obtain the identity

$$\begin{aligned}\Psi_I(1)\Psi_I(2) \cdots \Psi_I(n) &= \dot{\Psi}_I(1)\dot{\Psi}_I(2)\Psi_I(3) \cdots \Psi_I(n) \\ &+ \dot{\Psi}_I(1)\Psi_I(2)\dot{\Psi}_I(3) \cdots \Psi_I(n) + \cdots \\ &+ \dot{\Psi}_I(1)\Psi_I(2)\Psi_I(3) \cdots \dot{\Psi}_I(n) \\ &+ \epsilon^{n+1}\Psi_I(2)\Psi_I(3) \cdots \Psi_I(n)\Psi_I(1), \quad \text{(A10)}\end{aligned}$$

where

$$\dot{\Psi}_I(1)\dot{\Psi}_I(j) = [\Psi_I(1), \Psi_I(j)]_{-\epsilon}, \quad \text{(A11)}$$

with the convention that each term containing an ϵ commutator (A11) has a sign ϵ^P where $P = j - 2$ is the order of the permutation necessary to bring $\Psi_I(j)$ next to $\Psi_I(1)$.

[26] G. Wick, Phys. Rev. 80, 268 (1950).

[27] M. Gaudin, Nucl. Phys. 15, 89 (1960).

Taking the average $\langle\ \rangle_0$ and commuting $\Psi_I(1)$ through e^{-H_0} in the last term of (A10), we obtain

$$[1 - \epsilon S_{z-1}(x_1, \bar{x}_1)]\langle\Psi_I(\bar{x}_1, z_1)\Psi_I(2)\cdots\Psi_I(n)\rangle_0$$
$$= \langle\dot{\Psi}_I(1)\Psi_I(2)\Psi_I(3)\cdots\Psi_I(n)\rangle_0$$
$$+ \langle\Psi_I(1)\dot{\Psi}_I(2)\Psi_I(3)\cdots\Psi_I(n)\rangle_0 + \cdots$$
$$+ \langle\Psi_I(1)\Psi_I(2)\Psi_I(3)\cdots\dot{\Psi}_I(n)\rangle_0.$$

That is,

$$\langle\Psi_I(1)\Psi_I(2)\cdots\Psi_I(n)\rangle_0 = \langle\overbrace{\Psi_I(1)\Psi_I(2)}\Psi_I(3)\cdots\Psi_I(n)\rangle_0$$
$$+ \langle\overbrace{\Psi_I(1)\Psi_I(2)\Psi_I(3)}\cdots\Psi_I(n)\rangle_0 + \cdots$$
$$+ \langle\overbrace{\Psi_I(1)\Psi_I(2)\Psi_I(3)\cdots\Psi_I(n)}\rangle_0, \qquad \text{(A12)}$$

with

$$\overbrace{\Psi_I(1)\Psi_I(j)} = [1 - \epsilon S_{z-1}(x_1, \bar{x}_1)]^{-1}$$
$$\times \dot{\Psi}_I(\bar{x}_1, z_1)\Psi_I(x_j, z_j), \qquad \text{(A13)}$$

and the same sign convention as used for (A10). Relation (A12) for $n = 2$ identifies the contraction with the average value of the ordinary product of two operators

$$\overbrace{\Psi_I(1)\Psi_I(j)} = \langle\Psi_I(1)\Psi_I(j)\rangle_0. \qquad \text{(A14)}$$

The Wick theorem for an ordinary product of Ψ_I operators is true for 2 operators; (A12) shows that if it is true for $n - 1$ it is true for n operators.

(b) Transformation to T Products

A permutation in (A12) of $\Psi_I(j)$ and $\Psi_I(j + 1)$ changes by a factor ϵ the terms which do not involve a contraction of $\Psi_I(j)$ and $\Psi_I(j + 1)$; in the remaining terms, $\langle\Psi_I(j)\Psi_I(j + 1)\rangle_0$ is replaced by

$$\langle\Psi_I(j + 1)\Psi_I(j)\rangle_0.$$

If in (A12) all "times" are ordered $z_1 > z_2 > \cdots > z_n$, we have

$$T\langle\Psi_I(1)\Psi_I(2)\cdots\Psi_I(n)\rangle_0$$
$$= \langle\Psi_I(1)\Psi_I(2)\cdots\Psi_I(n)\rangle_0. \qquad \text{(A15)}$$

(A14) is identical to (A5) and the theorem for T products is verified. We choose now a different "time" ordering and show that the theorem remains true. It suffices to consider the ordering obtained by permutation of two consecutive "times"

$$z_1 > z_2 \cdots > z_{j+1} > z_j \cdots > z_n.$$

By definition of a T product, we have

$$T\langle\Psi_I(1)\cdots\Psi_I(j)\Psi_I(j + 1)\cdots\Psi_I(n)\rangle_0$$
$$= \epsilon\langle\Psi_I(1)\cdots\Psi_I(j + 1)\Psi_I(j)\cdots\Psi_I(n)\rangle_0. \qquad \text{(A16)}$$

Using (A12) and the above remark, we see that the terms of (A16) which do not involve a contraction of $\Psi_I(j)$ with $\Psi_I(j + 1)$ remain identical to the corresponding terms of (A15). The terms which involve such a contraction have

$$\langle\Psi_I(j)\Psi_I(j + 1)\rangle_0, \qquad (z_j > z_{j+1})$$

replaced by

$$\epsilon\langle\Psi_I(j + 1)\Psi_I(j)\rangle_0, \qquad (z_{j+1} > z_j),$$

which in both cases is

$$T\langle\Psi_I(j)\Psi_I(j + 1)\rangle_0,$$

proving the theorem for T products.

2. Representation of W With Faithful 0-Diagrams

With Wick's theorem we can make a one-to-one correspondence between each complete set of contractions arising in each term of expansion (A2) and a diagram. Diagrams are built with lines connecting ν-vertices (vertices with 2ν lines attached to them). Each line is associated with one T contraction, each ν-vertex with a ν-potential matrix.

In a diagram with p ν-vertices, each ν-vertex bears a *label* $i (1 \leq i \leq p)$ and the ν-potential matrix associated with it has 2ν *running variables* $j_i, k_i, \cdots$ to be summed over. j_i stands for momentum or position, spin, α index, and "time" (the "time" dependence reducing to $2\nu - 1$ δ functions for the usual "time" independent ν-potential matrix). Because we represent T-product averages, the diagrams are *Feynman* diagrams. This means that two diagrams are not considered distinct if they only differ by the relative position of the labeled vertices; however, they are considered distinct if the lines joining the p labeled ν-vertices correspond to distinct sets of T contractions.

The ν-vertex has to distinguish between the 2ν lines emerging from it, just as the 2ν T contractions of an operator Ψ_I with each one of the 2ν operators of a ν-potential are distinct. To avoid crowding the diagrams with labels, we shall represent the ν-potentials by "faithful" ν-vertices as shown in Fig. 1. That is to say, we introduce a dotted polygon with 2ν corners to represent faithfully a ν-vertex. The polygon is oriented and has one corner marked with a black dot to indicate an origin; a line emerges from each corner. For the ν-vertex labeled i, the associated 2ν running variables $j_i, k_i, \cdots$ are such

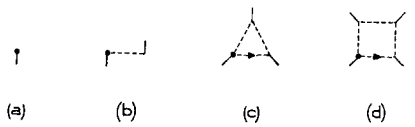


FIG. 1. Faithful representation of ν-vertices; (a) $\nu = \frac{1}{2}$; (b) $\nu = 1$; (c) $\nu = \frac{3}{2}$; (d) $\nu = 2$. The 2ν lines connecting the ν-vertices to the rest of the diagram, are also shown.

that the order of succession of the 2ν operators (from left to right) corresponds to the order of succession of the 2ν corners on the ν-vertex (from the origin on).

The term of pth order in the expansion (A5) for e^{W-W_0} is thus represented by all possible (Feynman) 0-diagrams, built with p faithful ν-vertices bearing a label i ($1 \leq i < p$). The actual algebraic contribution of each "faithful" labeled 0-diagram is obtained with the following Rule (A_f):

(A_fi) With each ν-vertex i, associate a ν-potential matrix

$$-[1/(2\nu)!]v_\nu(j_i, k_i, \cdots),$$

where $\nu = \frac{1}{2}, \frac{3}{2}, 2$; the arguments $j_i, k_i, \cdots$ are 2ν in number.

($A_f$ii) With each line joining the j_ith corner of the ν-vertex i with the $k_{i'}$th corner of the ν-vertex i', associate the contraction (unperturbed propagator)

$$G_1^0(j_i, k_{i'}) = T\langle\Psi_I(j_i)\Psi_I(k_{i'})\rangle_0.$$

($A_f$iii) Sum freely over all running variables $j_i, k_i, \cdots$, ($1 \leq i \leq p$).

($A_f$iv) With the whole diagram, associate a factor $(p!)^{-1}$.

The sign rule is examined in Sec. 3, this appendix.

We introduce the Fourier series

$$G_1^0(x, z; x', z') = \sum_{\omega, \omega'} e^{-(\omega z + \omega' z')} G_1^0(x, \omega; x', \omega'), \quad \text{(A17)}$$

where $\omega/i\pi$ and $\omega'/i\pi$ take even (Bose systems) or odd (Fermi systems) integer values. Translational invariance in "time", i.e.,

$$G_1^0(z; z') = G_1^0(z - z'),$$

guarantees that

$$G_1^0(\omega; \omega') = \delta_{\omega+\omega';0} G_1^0(\omega),$$

with

$$G_1^0(\omega) = \int_0^1 d(z - z') e^{\omega(z-z')} G_1^0(z - z').$$

Using (A17) and integrating over all "times", transforms points (i) to (iii) of the above rule into the corresponding points of Rule (A) [except for a factor $[(2\nu)!]^{-1}$ in (i)]; in the running variables, the "times" $z_i, z_k, \cdots$ are now replaced by energy variables $\omega_i, \omega_k, \cdots$ which satisfy the conservation law $\sum \omega_i = 0$ at each vertex.[28]

Faithful labeled 0-diagrams will prove useful to carry out proofs. They are, however, cumbersome to manipulate. In Sec. 5, this appendix, we shall introduce the reduced diagram representation which has been used throughout the text.

3. Representation of Green's Function with Faithful ν-Diagrams: Signs

As is well known, the ν-body distribution function (Green's function) defined in (2),

$$G_\nu(1, 2, \cdots, 2\nu) = T\langle\Psi(1)\Psi(2) \cdots \Psi(2\nu)\rangle,$$

can be rewritten as

$$= T\langle U(1, z_1)\Psi_I(1)U(z_1, z_2)\Psi_I(2) \cdots \times \Psi_I(2\nu)U(z_{2\nu}, 0)\rangle_0 / \langle U(1, 0)\rangle_0. \quad \text{(A18)}$$

The above extension of Wick's theorem allows the establishment of a one-to-one correspondence between the algebraic expansion of the numerator of (A18) and (faithful, Feynman-like) ν-diagrams, i.e., diagrams with 2ν *external* lines, originating at the external points with fixed arguments $1, 2, \cdots, 2\nu$. Only the ν-diagrams *connected* to at least one of the external lines remain after division by the normalization factor.

$G_\nu(1, 2, \cdots, 2\nu)$ is thus represented by the sum of all distinct, connected, ν-diagrams, built with p faithful vertices bearing a label $i (1 \leq i \leq p)$. Their contribution is again calculated with rule (A_f) where the summations are restricted to the

[28] When two operators belonging to the *same* vertex are contracted (that is, when a line closes upon itself at one vertex) resulting in a term, e.g.,

$$\cdots v_2(j, k, l, m)G_1^0(j, k) \cdots,$$

one must remember that in the expansion (A2) $\Psi(j)$ always occurs to the *left* of $\Psi(k)$. This may be accounted for, if, in (A4) $\delta(z_k - z_j)$ is replaced by $\delta(z_k - z_j + \eta)$, where η is an infinitesimally small, positive, quantity. Accordingly, after integration over z_k and introduction of the Fourier transforms (A17), we have

$$\int dz_k\, \delta(z_k - z_j + \eta)\, G_1^0(x_j, z_j; x_k, z_k) = \lim_{\eta \to 0} \sum_{\omega_j, \omega_k} e^{\omega_k \eta} e^{-(\omega_j + \omega_k) z_j}\, G_1^0(x_j, \omega_j; x_k, \omega_k) = \lim_{\eta \to 0} \sum_{\omega_j} e^{-\omega_j \eta}\, G_1^0(\omega_j).$$

Convergence factors $e^{\pm\omega\eta}$ will not be carried along explicitly. They have to be taken into account, however, when calculating in the ω representation the contribution of "equal time" terms. For example, for W_0 in the ω representation, one has

$$W_0 = \lim_{\eta \to 0} \tfrac{1}{2}\epsilon\, \mathrm{tr} \sum_{\omega_j, \omega_k} e^{\omega_k \eta} \ln G_1^0(x_j, \omega_j; x_k, \omega_k) = \lim_{\eta \to 0} \tfrac{1}{2}\epsilon\, \mathrm{tr} \sum_{\omega_j} e^{-\omega_j \eta} \ln G_1^0(\omega_j).$$

Note that

$$\lim_{\eta \to 0} \sum_{\omega} e^{\pm\omega\eta} = 0.$$

running variables, the variables 1, 2, $\cdots$, 2ν being kept fixed.

We now examine the sign rule. The sign is a problem only for Fermi systems, and for such systems, vertices with an odd number of operators are irrelevant. We shall only consider 2-vertices. In order to define loops in a diagram, we arbitrarily decide that at each vertex $v_2(j, k, l, m)$ the corners corresponding to the first two arguments (j, k) belong to the same loop λ, and the ones corresponding to (l, m) belong to the loop λ'. The lines of a diagram (or the contractions of a T-product average) are thereby distributed into loops; a 0-diagram contains only closed loops, a ν-diagram also contains ν open loops; each open loop begins and ends with an external point. We arbitrarily orient each closed loop; an open loop begins with its external point whose argument is further to the right in $G_\nu(1, 2, \cdots 2\nu)$.

Consider now the average value of a T product of operators. Without changing its value, we may, inside the T-product average[26]

(i) group together all pairs of operators $\Psi_I(j)\Psi_I(k)$ or $\Psi_I(l)\Psi_I(m)$ belonging to the same loop;

(ii) put these pairs in the order of succession in which they occur in the loop.

It follows from the sign rule of Wick's theorem that if we write the T contraction as $G_I^0(j, j')$ when j follows j' in the oriented loop, the overall sign ϵ^P may be split into

(a) a sign ϵ^{r+1} for each closed loop,
(b) a sign ϵ^r for each open loop,
(c) an overall sign ϵ^{P_ν}.

Here r is the number of inversions of the standard order of the arguments (j, k) or (l, m) at each vertex encountered as we proceed along the oriented loop.[29] The 2ν arguments of $G_\nu(1, 2, \cdots 2\nu)$ are associated in pairs belonging to the same open loop. P_ν is the number of permutations necessary to bring these pairs next to each other in the set $(1, 2, \cdots 2\nu)$.

Finally, we introduce a faithful, *global* representation for ν-body distribution functions which we shall use in the next section. To be specific, we consider the 2-body distribution function. A stripped 2-body correlation function is defined by Eqs. (42) and (43) for a normal system, and by Eq. (64) for a superfluid system. We want to represent the C_2 function in such a way that the ordering of its various arguments is immediately recognized

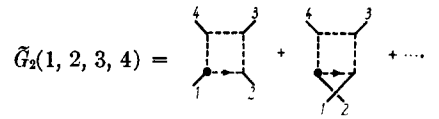


FIG. 2. Faithful representation of correlation functions ($\tilde{G}_2$, cumulant part of G_2). There are 4! diagrams in the right-hand side.

(remember that G_2 and C_2 are ϵ-symmetrical under exchange of two arguments). We therefore associate $(4!)^{-1}C_2(1, 2, 3, 4)$ with a heavy (faithful) 2-vertex and $\tilde{G}_2(1, 2, 3, 4)$ is drawn as in Fig. 2, the heavy external lines being associated with $\tilde{G}_1$ functions [in a superfluid system $\tilde{G}_2$ also contains three terms in $(C_{\frac{3}{2}})^2$].

4. Insertions: the Double Labeling Method

Throughout the text (Secs. II to V) we have considered ν-insertions in $[\nu - (\frac{1}{2})]$-irreducible diagrams, but we have assumed properties for the insertions (factorization, completeness)[14] which we shall now derive. To be specific, we consider 2-insertions in $\frac{3}{2}$-irreducible diagrams; these are the most complicated insertions introduced in the text. We suppose further for simplicity that only 2-vertices exist. Proofs worked out in what follows are thus directly relevant to Sec. V, and are easily transposed to simpler types of insertions (Secs. II to IV).

Definitions. A *free ν-diagram* is defined as a nonlabeled ν-diagram; its contribution is given by the sum of the contributions of the $p!/S_f$ distinct, labeled ν-diagrams obtained by assigning, in all possible ways, p labels to its p vertices. S_f is called the *symmetry number* of the diagram; notice that for faithful ν-diagrams, $S_f = 1$ when $\nu \neq 0$.

In a $\frac{3}{2}$-irreducible diagram, a *2-insertion* is any part of the diagram connected by four lines to the rest of it. To place a 2-insertion in a ν-diagram (γ_ν), called the skeleton, is to replace one of the p 2-vertices of (γ_ν) by any 2-diagram (stripped of its external lines); if the 2-insertion contains q 2-vertices, the new ν-diagram (Γ_ν) thus generated contains $(p + q - 1)$ 2-vertices. This operation is of particular interest when extended and characterized in the following way:

(i) The skeleton ν-diagram is always the class of $(p!/S_f)$ ν-diagrams corresponding to a given free ν-diagram.

(ii) The 2-insertion consists in a whole class of topologically specified 2-diagrams (stripped of their external lines), e.g., in the following we shall only consider the class of all connected, $\frac{3}{2}$-irreducible 2-diagrams.

[29] If we change the orientation of a closed loop attached to q vertices, r of which contribute inverted pairs, the sign is transformed from $(\epsilon)^{r+1}$ to $(\epsilon)^{q-r+1}(\epsilon)^9$ [since $G_1^0(j, k) = \epsilon\, G_1^0(k, j)$], and is therefore invariant.

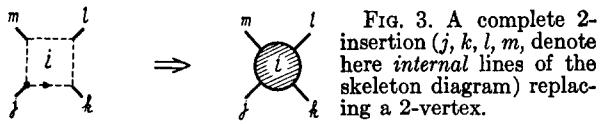


FIG. 3. A complete 2-insertion (j, k, l, m, denote here *internal* lines of the skeleton diagram) replacing a 2-vertex.

Consider the case where a single 2-vertex is replaced by any 2-insertion. If we do not distinguish two diagrams differing only through exchange of labels between the skeleton and the insertion (assuming for the moment that this distinction is unambiguous, as it generally is when $\nu \neq 0$) we have for each distinct (Γ_ν) diagram a weight

$$[1/(p+q-1)!]\ [(p+q-1)!/(p-1)!\,q!],$$

or, if we use an extra label attached globally to the insertion,

$$(1/p!)\ (1/q!).$$

From this factorization property it follows that inserting the whole class of 2-insertions characterized in (ii) above amounts to replacing a 2-vertex, labeled i, contributing $(4!)^{-1}v_2(j_i, k_i, l_i, m_i)$ by a factorized sum of 2-insertions, labeled i, contributing $C_2(j_i, k_i, l_i, m_i)$ (Fig. 3); we call this sum of 2-insertions a *complete* 2-insertion.

It is convenient to define an *ordered 2-insertion*. We replace the complete 2-insertion contributing $C_2(j_i, k_i, l_i, m_i)$ by the average of the 4! complete, ordered 2-insertions obtained by distributing, in all possible ways, four *ordering labels* 1, 2, 3, 4 among the four lines connecting the 2-insertion to the rest of the diagram (Fig. 4). Each 2-insertion ordered in this manner is then represented by a faithful heavy 2-vertex, associated with $(4!)^{-1}C_2(1, 2, 3, 4)$; that is, for the two terms shown in Fig. 4, $(4!)^{-1}C_2(j_i, k_i, l_i, m_i)$ and $(4!)^{-1}C_2(k_i, j_i, l_i, m_i)$, respectively. Putting in the complete ordered 2-insertion consists in replacing a faithful (ordered) 2-vertex i by the faithful heavy 2-vertex i with the same ordering. This replacement is schematized in Fig. 5. The 2-insertions considered are therefore finally char-

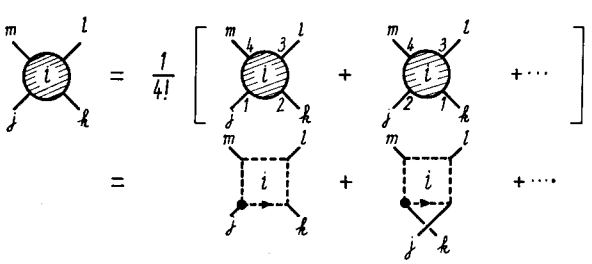


FIG. 4. Complete, ordered, 2-insertion and faithful heavy 2-vertex associated.

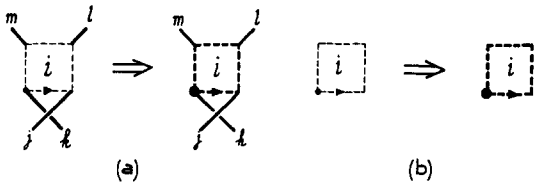


FIG. 5. Transformation under replacement of a 2-vertex by a complete, ordered, 2-insertion in a diagram, (a) symbolized by the transformation, (b) i.e.,
$$-(4!)^{-1}v_2(1_i, 2_i, 3_i, 4_i) \rightarrow (4!)^{-1}C_2(1_i, 2_i, 3_i, 4_i).$$

acterized by

(iii) The 2-insertions are *ordered*.

(iv) Each 2-vertex of the skeleton (γ_ν) is replaced by the complete, ordered, 2-insertion.

Factorization Property

We found it convenient above to attach an extra label globally to a 2-insertion. Likewise we now introduce an extra labeling (σ-labeling, σ for skeleton) for each vertex of the skeleton where an insertion is to be made. In other words, each one of the $(p!/S_t)$ distinct, labeled, ν-diagrams (the skeleton), corresponding to a given assignment of the p labels, is now replaced by $p_\sigma!$ diagrams bearing the same assignment of the usual labels and all possible distinct assignments of σ_i labels ($1 \leq i \leq p_\sigma$). Diagrams with double labeling are now weighted by $(p!)^{-1}(p_\sigma!)^{-1}$, where p_σ is the number of vertices where the insertions are made; here we take $p_\sigma = p$. ν-diagrams with double labeling have the following properties:

(a) When introducing a 2-insertion, the σ-labeling of the 2-vertex concerned is transferred globally to the 2-insertion. It we replace the 2-vertex σ_i by a 2-insertion, again globally labeled σ_i, there is no ambiguity between the skeleton and what is inserted.

(b) A permutation of the usual labels between two insertions σ_i, σ_j now leads necessarily to distinct diagrams, the σ-labeling remaining distinct. (This is not the case for 0-diagrams with the usual labeling only).

Thus each ν-diagram, obtained by inserting an ordered 2-insertion, is weighted by

$$(1/p!)\ (1/4!)^p\ (1/Q!),$$

with

$$Q = \sum_{i=1}^{p} q_i,$$

and q_i the number of 2-vertices of the σ_i 2-insertion. If we do not distinguish two ν-diagrams differing

only through permutation of the usual labels of two 2-insertions, using property (b), we obtain the new weight

$$\frac{1}{p!}\left(\frac{1}{4!}\right)^p \frac{1}{Q!}\frac{Q!}{\prod_{i=1}^{p} q_i!} = \frac{1}{p!}\prod_{i=1}^{p}\left(\frac{1}{4!}\frac{1}{q_i!}\right).$$

With this factorization property, the introduction of a complete, ordered, 2-insertion as defined above, amounts to replacing each 2-vertex by the corresponding heavy 2-vertex (Fig. 5) and in the contribution of the skeleton, each $-v_2(j, k, l, m)$ by $C_2(j, k, l, m)$.

Example: if the original skeleton is the complete class of 2-irreducible 0-diagrams (or 2-diagrams, stripped of their external lines) contributing $\mathcal{K}^{(2)}\{-v_2\}$ [or $K_2\{1, 2, 3, 4; -v_2\}$], then, after performing at each 2-vertex a complete, ordered, 2-insertion, we obtain $\mathcal{K}^{(2)}\{C_2\}$ [or $K_2\{1, 2, 3, 4; C_2\}$].

Completeness

In Sec. V, we have used the above result and then identified:

(i) the 2-diagrams contributing to $K_2\{1, 2, 3, 4; C_2\}$ as the class $(\bar{\Gamma}_2)$ of all distinct ($\frac{3}{2}$-irreducible) 2-diagrams, simple with respect to each pairing of the external lines [as defined in Sec. V.A(1) and B(1)].

(ii) The 0-diagrams contributing to $\mathcal{K}^{(2)}\{C_2\}$ as the class $(\bar{\Gamma}_0)$ of all distinct ($\frac{3}{2}$-irreducible) 0-diagrams where one 2-irreducible skeleton is distinguished.

This identification is completed if we show that $(\bar{\Gamma}_2, \bar{\Gamma}_0)$ are respectively in one-to-one correspondence with the classes (Γ_2, Γ_0) generated in the above paragraph.

The $(\bar{\Gamma}_0)$ 0-diagrams have one 2-irreducible skeleton distinguished. The $(\bar{\Gamma}_2)$ 2-diagrams have a unique 2-irreducible skeleton; this is a reflection of the fact that the symmetry number of faithful ν-diagrams ($\nu \neq 0$) is $S_f = 1$; alternatively a very simplified form of the analysis in cycles of pairs done in Sec. V and Appendix B would directly show it (the unique 2-irreducible skeleton is the one directly connected to the external lines). The $(\bar{\Gamma}_2, \bar{\Gamma}_0)$ diagrams thus have a well-defined skeleton.

We now distribute in all possible distinct ways [$p!(4!)^p$ in number], p σ-labels and p sets of four ordering labels, among the p 2-insertions of each distinct labeled diagram of $(\bar{\Gamma}_2, \bar{\Gamma}_0)$ (with the extra weight factor $[p!(4!)^p]^{-1}$ for each diagram). The diagrams thus obtained are in one-to-one correspondence with the diagrams of (Γ_2, Γ_0), which completes the identification and proves the results used in Sec. V.

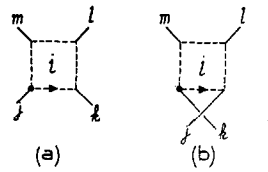


FIG. 6. The part not shown of diagrams (a) and (b) is identical; these diagrams differ only by permutation of two lines at vertex i.

Diagrams Resulting from General ν-Insertions

The faithful labeled diagrams resulting from general ν insertions have the following characteristics:

(i) The complete $\frac{1}{2}$-insertion is represented by a heavy wavy $\frac{1}{2}$-line, bearing a σ-label. The ν-vertices ($\nu = 1, \frac{3}{2}, 2$) have the usual labels. Likewise, the complete 1-insertion is represented by a heavy line bearing a σ-label, the ν-vertices ($\nu = \frac{3}{2}, 2$) having usual labels. Faithful diagrams are completely characterized by their usual labels and the σ-labels are now redundant (i.e., exchange of σ-labels always leads to distinct diagrams) so we may return to singly labeled diagrams by suppressing the σ-labels and the corresponding $(p_\sigma!)^{-1}$ factor.

(ii) The complete $\frac{3}{2}$-insertions are represented by heavy $\frac{3}{2}$-vertices bearing a σ-label, the 2-vertices having the usual label. Here the σ-label in the (generalized) $\frac{3}{2}$-vertices replaces the usual label and is not redundant; likewise, the σ-label attached to heavy 2-vertices resulting from a complete 2-insertion replaces the usual label.

5. A Reduced Diagram Representation

Faithful diagrams are useful to perform proofs. They are cumbersome to carry along. We introduce now the reduced diagrams used throughout the text.

Consider two distinct faithful, labeled diagrams differing by a permutation of two lines connecting a ν-vertex to the rest of the diagram (Fig. 6). The contribution of these two diagrams are obtained by summing over the integrands

$$(\cdots v_\nu(j_i, k_i, \cdots)G_1^0(j_i, j)G_1^0(k_i, k)\cdots),$$

and

$$\epsilon(\cdots v_\nu(k_i, j_i, \cdots)G_1^0(k_i, k)G_1^0(j_i, j)\cdots).$$

The ϵ arises because the number of closed loops (or the number of inversions when the two lines belong to the same loop) has changed by one unit. These two contributions are identical since we have

$$v_\nu(j_i, k_i, \cdots) = \epsilon v_\nu(k_i, j_i, \cdots).$$

Given a labeled diagram, it is possible to carry out $(2\nu)!$ permutations at each vertex. Each one of these permutations leads to labeled diagrams

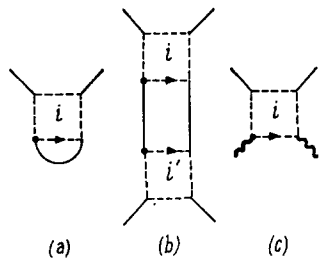


FIG. 7. (a) Line with both ends at the same vertex i. (b) Multiplet of m = two lines connecting two 2-vertices, i, i'. (c) Multiplet of n = two $\frac{1}{2}$-lines at a 2-vertex i.

which have the same contribution and are *distinct* except when

(i) one line closes itself upon one ν-vertex [Fig. 7(a)],

(ii) a multiplet of m lines ($m = 2, 3, 4$)[8] directly connects two vertices [Fig. 7(b)].

The same reasoning may be used for faithful labeled diagrams resulting from ν-insertions, i.e., containing a heavy $\frac{1}{2}$-line, or line, and heavy $\frac{3}{2}$- or 2-vertices. The conclusions are the same, except that there is one more case where the resulting labeled diagrams are not distinct, i.e., when

(iii) a multiplet of n $\frac{1}{2}$-lines ($n = 2, 3, 4$) are attached to one vertex [Fig. 7(c)].

The class of distinct, labeled, faithful diagrams obtained by this operation contains

$$R = (S_L)^{-1} \prod_{i=1}^{p} (2\nu_i)!$$

terms giving the same contribution; S_L plays the role of a symmetry factor for the lines, and is obtained by associating

(i) a factor $\frac{1}{2}$ with each line closing upon itself at one vertex;

(ii) a factor $(m!)^{-1}(m = 2, 3, 4)$ with each multiplet of m lines directly connecting two vertices;

(iii) a factor $(n!)^{-1}(n=2, 3, 4)$ with each multiplet of n half lines attached to the same vertex.

We represent this class of labeled faithful diagrams by a single labeled *reduced* diagram which is assigned the extra weight R and in which the ordering at each vertex is no longer distinguished. That is, in a reduced diagram, as used throughout the text, a ν-vertex is represented by just a black dot where 2ν lines intersect. (When diagrams are stripped of their external lines, a number of the 2ν lines of a ν vertex may be removed. They are replaced by as many *fixed arguments* written next to the vertex.) The sign of a reduced diagram is obtained by considering any of the R faithful diagrams. In particular, there always is one of these R faithful diagrams in which there are no inversions of the standard order of the paired operators at each vertex. According to the sign rule of Sec. 3, this appendix, the sign reduces to $(\epsilon)^{\lambda+P_\nu}$ where λ is the number of closed loops of the diagram.

6. Final Rule for All Reduced Diagrams

The contributions associated with reduced ν-diagrams are therefore calculated by the following *general rule* (embedding all rules stated throughout the text). These rules lead to the construction of a function of 2ν fixed arguments $\Gamma(1, 2, \cdots 2\nu)$.

(i) With each $\frac{1}{2}$-, $\frac{3}{2}$-, or 2-vertex, associate $-v_{\frac{1}{2}}(j_i)$, $-v_{\frac{3}{2}}(j_i, k_i, l_i)$, or $-v_2(j_i, k_i, l_i, m_i)$. With each heavy $\frac{3}{2}$- or 2-vertex, associate $C_{\frac{3}{2}}(j_i, k_i, l_i)$ or $C_2(j_i, k_i, l_i, m_i)$.

(ii) With each line, associate $G_1^0(j_i, k_{i'})$. With each heavy line, associate $\tilde{G}_1(j_i, k_{i'})$. With each heavy $\frac{1}{2}$-line, associate $G_{\frac{1}{2}}(j)$.

(iii) Free summations run over the variables $j_i, k_i \cdots$, $(1 \leq i \leq p)$. The external variables (2ν in number for a ν-diagram) are kept fixed.

(iv) Distribute the lines of a ν-diagram into closed loops and ν open loops. The sign of the diagram is $(\epsilon)^{\lambda+P_\nu}$; λ is the number of closed loops; P_ν the order of the permutation transforming the order $(1, 2, \cdots 2\nu)$ into the order where the arguments paired together by an open loop are next to each other.

Orient arbitrarily each closed loop; an open loop is oriented from the external point whose argument is furthest to the right in $\Gamma(1, 2, \cdots 2\nu)$.

Consider two arguments j', j appearing successively in an oriented loop; when the two arguments belong to a line associated with G_1^0, $\tilde{G}_1$, write $G_1^0(j, j')$, $\tilde{G}(j, j')$; when they belong to a 2-vertex associated with $-v_2$, C_2, write $-v_2(j, j', \cdot, \cdot)$ [or $-v_2(\cdot, \cdot, j, j')$], $C_2(j, j', \cdot, \cdot)$ [or $C_2(\cdot, \cdot, j, j')$].

(v) Associate a factor $(p!S_L)^{-1}$ with each distinct *labeled diagram* (one label for each one of the p vertices).

(vi) Or equivalently, associate a factor $(S_V S_L)^{-1}$ with each distinct *unlabeled* or *free* diagram.

The two representations considered in (v) and (vi) are equivalent. S_V is a symmetry number for the vertices.[9]

Example: Eq. (75) last line: $S_L = 1$; $S_V = 8$, 24, 72, 48, 16; Eq. (76) $S_V = 1$: first line, $S_L = 1$, and then $S_L = 2$; second line, $S_L = 6, 6, 4, 2, 2$.

Remark on functional derivatives. The functional-derivative definition of the Green's function expressed by Eq. (8) is translated into faithful labeled diagrams by saying that $G_\nu(1, 2, \cdots 2\nu)$ is obtained by suppressing one ν-vertex $-(2\nu!)^{-1}v_\nu(1, 2, \cdots 2\nu)$

in all possible ways in each 0-diagram. Since the labeling of the suppressed ν-vertex is irrelevant, we obtain all distinct, labeled, faithful ν-diagrams with the proper weight (i.e., $[(p-1)!]^{-1}$ if the original 0-diagram has p vertices).

APPENDIX B. TOPOLOGICAL THEOREMS

We prove here relation (56) and its equivalent for anomalous systems. For this purpose we analyze the topological structure of 0-diagrams with respect to 1- and 2-insertions.

1. Analysis of 1-Insertions

We recall the analysis which leads to relation (27), a relation valid for any $\frac{1}{2}$-irreducible 0-diagram in anomalous systems, or any 0-diagram in normal systems.

(a) *Cycles of Lines*

We have seen [Sec. III.(1)] that the set of all lines of a 0-diagram can be unambiguously divided into subsets, the cycles c_m, where m is the number of lines of the cycle.

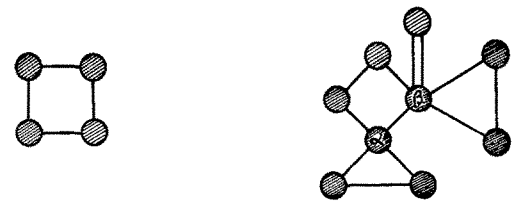
(b) *Structure of a 0-Diagram*

Definition: In a 0-diagram, a *1-skeleton* is a topological element connected to the rest of the diagram by one pair of lines and satisfying the following property. The remainder of the diagram may be viewed as a 1-insertion; if we suppress this 1-insertion and replace it by a single line the 1 skeleton is transformed into a 1-irreducible 0-diagram.

In a 0-diagram with a single cycle c_m, $m > 1$, m 1-skeletons are connected by the m lines of the cycle [see Fig. 8(a)]. Cycles with $m = 1$, that is, cycles of a single line, are not singled out in this figure.

Definition: A *q-fold 1-skeleton* is a topological element connected to the rest of the diagram by q pairs of lines. The remainder of the diagram separates into q different 1-insertions, one for each pair of lines. If we suppress one of these 1-insertions and replace it by a single line, we obtain a $(q-1)$-fold 1-skeleton. A 1-fold 1-skeleton is identical with the above-defined 1-skeleton.[30] By convention a 1-irreducible 0-diagram is a 0-fold 1-skeleton.

[30] The skeletons defined are a particular case of the skeletons introduced in Appendix A. Here, however, the skeletons are *always 1-irreducible*; throughout the text we have indeed used the terminology ν-irreducible skeletons. Here we wish to indicate also the relationship of the ν-irreducible skeletons with the diagrams which they generate through ν-insertions. For example if the rest of the diagram is generated by inserting q ν-insertions in a ν-irreducible skeleton, we then speak of a q-fold (ν-irreducible) skeleton, or for short, a q-fold ν-skeleton.



(a) (b)

FIG. 8. (a) Cycle c_4; four lines connect four 1-skeletons. (b) Tree of cycles; the q-fold 1-skeleton $\alpha(q = 2)$ belongs to two cycles c_3, c_4; the q-fold 1-skeleton $\beta(q = 3)$ belongs to three cycles c_4, c_3, c_2.

In a 0-diagram with several cycles c_m, $m > 1$, a q-fold 1-skeleton belongs to q distinct cycles c_m [see Fig. 8(b)].

It follows from the definition of a cycle that two cycles cannot have in common more than one q-fold 1-skeleton. Equivalently, several cycles cannot form a ring of cycles; they necessarily form a *tree* of cycles.[31]

(c) *Topological Relation*

For any tree of cycles, the topological relation (27) (a variant of the Euler–Poincaré relations)

$$N(\mathrm{c}) - N(\mathrm{l}) + N(\mathcal{K}^{(1)}) = +1 \qquad \text{(B1)}$$

is immediately verified by induction [if true for a tree containing $N(\mathrm{c})$ cycles, $N(\mathrm{l})$ lines, $N(\mathcal{K}^{(1)})$ (q-fold] 1-skeletons, it remains true when one more cycle is added]. Notice that (B1) can also be written as

$$N(C) - N(\mathcal{K}^{(1)}q) + N(\mathcal{K}^{(1)}) = +1, \qquad \text{(B2)}$$

or

$$\sum_c [1 - N_c(\mathcal{K}^{(1)})] + N(\mathcal{K}^{(1)}) = +1, \qquad \text{(B3)}$$

where C is used for any nontrivial cycles c_m ($m > 1$). $N(\mathcal{K}^{(1)}q)$ is the number of q-fold 1-skeletons (shared by q cycles) weighted by their multiplicity q, and $N_c(\mathcal{K}^{(1)})$ is the number of (q-fold) 1-skeletons contained in a cycle c.

2. Analysis of 2-Insertions: Normal Systems

We prove (56) for 1-irreducible 0-diagrams in normal systems.

[31] Let c and c' be two cycles. Each path joining c and C' contains lines belonging to a set of cycles $(c^{(1)}, c^{(2)} \cdots)$ cutting open any two lines of c (or c') separates the diagram into two disconnected parts, hence there cannot exist two nonoverlapping sets $(c^{(1)}, c^{(2)} \cdots)$. This property is the characteristic of a tree structure.

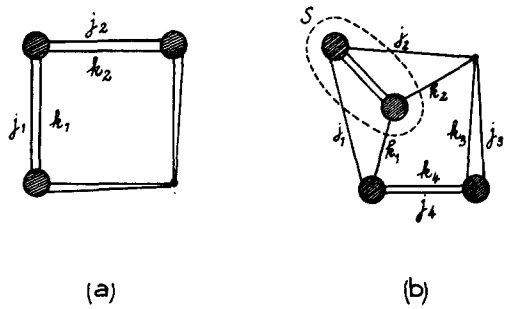


FIG. 9. (a) Cycle γ_4; four pairs of lines connect three 2-skeletons and one 2-vertex; the diagram also contains several γ_2's. (b) Cycle γ_4: four pairs of lines $(j_1\ k_1; j_2\ k_2; j_3\ k_3; j_4\ k_4)$ connect one 2-vertex, two 2-skeletons, and one 2-insertion S simple with respect to $(j_1\ k_1; j_2\ k_2)$, built with two 2-skeletons. The diagram also contains one γ_3 (and several γ_2's).

(a) *Cycles of Pairs of Lines*

We have seen (Sec. V.A, Normal Case) that the set of all possible *pairs* of lines of a (1-irreducible) 0-diagram can be unambiguously divided into subsets, the cycles γ_μ, where μ is the number of pairs of the cycle ($\mu \geq 1$).

Definition: A *2-skeleton* is a topological element connected to the rest of the diagram by a quadruplet of lines and satisfying the following property. The remainder of the diagram may be viewed as a 2-insertion; if we suppress this 2-insertion and replace it by a 2-vertex, the 2-skeleton is transformed into a 2-irreducible 0-diagram.

We consider a cycle γ_μ. The μ pairs of this cycle may connect only 2-skeletons or 2-vertices. In this case, in addition to γ_μ, the diagram can only contain cycles $\gamma_{\mu'}$ with $\mu' = 1, 2$; indeed, any two lines j_1 and j_2 chosen from two different pairs of the μ pairs of γ_μ constitute a pair which belongs to a cycle $\gamma_2[(j_1, j_2), (k_1, k_2)]$, and cannot belong to a cycle of higher multiplicity [see Fig. 9(a)]. From the definition of a 2-skeleton, any pair of lines other than the μ pairs of γ_μ can only belong to a cycle γ_1 or γ_2.

The most general situation occurs when the μ pairs of the cycle connect[32] 2-insertions which are not necessarily 2-irreducible, but need only be *simple* with respect to the two pairs which they connect [Fig. 9(b)], i.e., two pairs (j_1, k_1), (j_2, k_2), occuring in succession in γ_μ are connected by a 2-insertion, simple with respect to this pairing.

Consider now two cycles γ_{μ_1}, γ_{μ_2}. The following situations may occur:

(i) They have no lines or a single line in common.

(ii) They have two lines j, k in common.

(a) The lines j, k (or any two lines common to γ_{μ_1}, γ_{μ_2}) do *not* constitute a pair of γ_{μ_1}, γ_{μ_2}.

Let j_1, k_1 and j_2, k_2 be the lines paired with j, k in γ_{μ_1} and γ_{μ_2}. The pairs (j, k), (j_1, k_1), (j_2, k_2) belong to a cycle γ_μ which shares four lines with γ_{μ_1} and four lines with γ_{μ_2}.

Let[33] $\mu_1 > 2$ and $\mu_2 > 2$; then (j, j_1) and (k, k_1) are necessarily successive pairs in γ_{μ_1}, [Fig. 10(a)]. The $\mu - 2$ pairs other than (j, j_1), (k, k_1) of γ_{μ_1} are "internal" lines of the 2-insertion $S^{(1)}$ [nonsimple with respect to $(j, j_1)(k, k_1)$, and hence simple with respect to the other pairings]. Likewise the $\mu_2 - 2$ pairs other than (j, j_2), (k, k_2) of γ_{μ_2} are internal lines of $S^{(2)}$. This is enough to prove that γ_{μ_1}, and γ_{μ_2} cannot have more than two lines in common.

(b) The lines (j, k) constitute a pair of γ_{μ_1}, γ_{μ_2}.

Let j_2, k_2 be the lines paired with j, k in γ_{μ_2}. Clearly (j_2, k_2) is also a pair of γ_{μ_1}, and the two cycles share four lines which constitute two pairs of γ_{μ_1} and two pairs of γ_{μ_2}.

Let $\mu_1 > 2$ and $\mu_2 > 2$: (j, k), (j_2, k_2) are necessarily successive pairs of γ_{μ_1}, and (j, j_2), (k, k_2) of γ_{μ_2} [Fig. 10(b)]. The $\mu_1 - 2$ pairs other than (j, k), (j_2, k_2) in γ_{μ_1} are "internal" lines of the 2 diagram $S^{(1)}$ [nonsimple with respect to (j, k), (j_2, k_2)], and the $\mu_2 - 2$ pairs other than (j, j_2), (k, k_2) of γ_{μ_2} are internal lines of $S'^{(2)}$ [nonsimple with respect to (j, j_2), (k, k_2)]. A 2-insertion can be

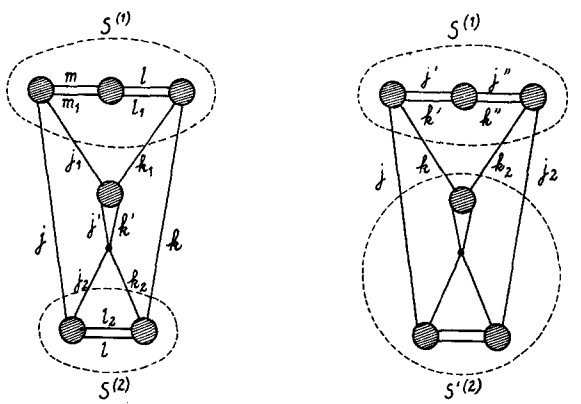


FIG. 10. (a) Cycles; γ_{μ_1}: (j, j_1), (k, k_1), (l, l_1), (m, m_1); γ_{μ_2}: (j, j_2), (k, k_2), (l, l_2); γ_μ: (j, k) (j_1, k_1) (j', k') (j_2, k_2). (b) γ_{μ_1}: (j, k), (j', k'), (j'', k''), $(j_2, k_2) \equiv \gamma_{\mu_1}$ above. $\gamma_{\mu_2} \equiv \gamma_\mu$ above.

[32] A 2-skeleton may be viewed also as a 2-irreducible 2-insertion, i.e., if we separate a 2-skeleton from the rest of the diagram by cutting open its quadruplet of lines, it becomes a 2-irreducible 2-diagram. Likewise, a 1-skeleton is also a 1-irreducible 1-insertion, but a q-fold 1-skeleton is neither a 1- nor a q-insertion.

[33] The particular cases, where γ_{μ_1} or γ_{μ_2} (or both) is a γ_2, can be discussed in the same way and do not lead to different results.

nonsimple in a unique way; hence $S^{(1)}$ and $S'^{(2)}$ cannot coincide. This shows that γ_{μ_1} and γ_{μ_2} cannot have in common more than the four lines connecting $S^{(1)}$ and $S'^{(2)}$.

To sum up we have shown that two cycles[33] can only have in common:

(i) zero or one line;

(ii) two unpaired lines. In this case there exists a cycle $\gamma_\mu(\mu > 2)$ sharing four lines with γ_{μ_1} and four lines with γ_{μ_2};

(iii) four lines which are always distributed into two pairs inside γ_{μ_1} and γ_{μ_2}. These two pairs are successive in γ_{μ_i} if $\mu_i > 2$, $(i = 1, 2, i \neq j)$.

Definitions: two cycles of pairs having four lines in common are said to be *linked*. A set of linked cycles is said to be *connected* if all its cycles are linked two-by-two in such a way as to link indirectly (or directly) any two cycles of the set.

In the partition of cycles, a *connected* subset of nontrivial cycles (with $\mu \geq 2$) will be called a *parquet*. A 0-diagram may contain no parquet (2-irreducible 0-diagram), one parquet (*parquet diagram* where all nontrivial cycles of pairs are connected), or several parquets.

The parquets play here the role played in Sec. 1 of this appendix by the cycles of lines.

(b) *Structure of a 0-Diagram*

(1) *Single parquet.* It is convenient to schematize the diagrams used thus far and introduce "*structure diagrams*". In a structure diagram we represent a cycle of pairs γ_μ by a set of μ lines (each line standing for a pair) connecting μ black dots (standing for 2-vertices) or shaded bubbles (standing for 2-skeletons) into a closed figure which we call a circle of pairs [Fig. 11(a)]. Two linked cycles γ_μ, $\gamma_{\mu'}$, are represented by two circles of pairs linked by a link (white dot), each circle having, respectively, $\mu - 1$ and $\mu' - 1$ black dots or shaded bubbles [Fig. 11(b)]. The introduction of a link in this simple fashion is possible because the pair of lines common to two cycles γ_μ, $\gamma_{\mu'}$ $(\mu, \mu' > 2)$ are distributed into two *successive* pairs inside each cycle.

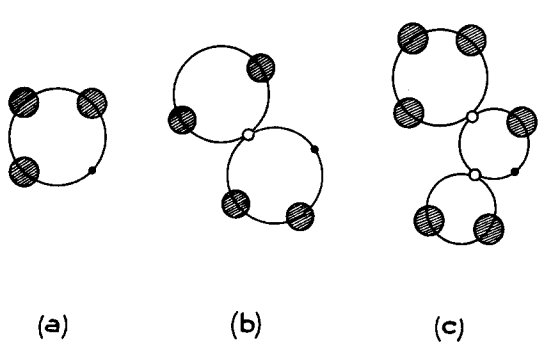


FIG. 11. "Structure diagrams" for (a) Fig. (9a): one circle of pair; (b) Fig. (9b): two circles of pairs connected by one link; (c) Fig. (10): three circles of pairs connected by two links. The cycles γ_{μ_1}, γ_μ, γ_{μ_2} of Fig. 10(a) are respectively represented by the circles with three, one, and two shaded bubbles.

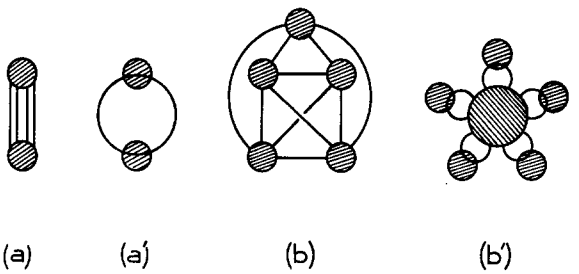


FIG. 12. 0-diagrams containing only γ_2's and their structure diagram representation. (a) connected γ_2's (single parquet), (b) disconnected γ_2's (five parquets).

In this way the topological relation between cycles is cast in evidence. We have seen that when a γ_2 shares four lines with a γ_μ $(\mu > 2)$ the two pairs of γ_μ formed by these four lines are not necessarily in succession. The relation γ_2, γ_μ could not and will not be represented in the above fashion. The explicit representation of γ_2 is, however, not necessary since γ_2 is completely embedded in γ_μ, an explicitly represented cycle. The "structure diagrams" still represent the topological relation between γ_μ's $(\mu > 2)$ and contain faithfully all the elements of the original diagram. Likewise, the trivial γ_2's internal to a 2-skeleton need not be displayed, being embedded in an explicitely represented 2-skeleton.

Only in a parquet which reduces to nontrivial cycles γ_2 is it necessary to display these cycles explicitly, since they are not embedded in cycles γ_μ already displayed, or in 2-skeletons. Such parquets arise from 0-diagrams which reduce to two 2-skeletons or two 2-vertices [Fig. 12(a)] connected by a quadruplet of lines (constituting three connected cycles γ_2). We depict in Fig. 12(a') one of the cycles γ_2 by a circle of pairs (the two other cycles γ_2 are embedded this time in the one cycle γ_2 represented). Clearly the counting of cycles γ_2 in a structure diagram would be rather delicate, but we shall never have to do this.

A parquet is then represented by connected circles of pairs, the connection being insured by links. From the properties [(i) to (iii) above] of the cycles of pairs, it follows that two circles of pairs have *at most* one link in common. Consequently, circles of pairs cannot form a ring; they necessarily constitute a *tree*.

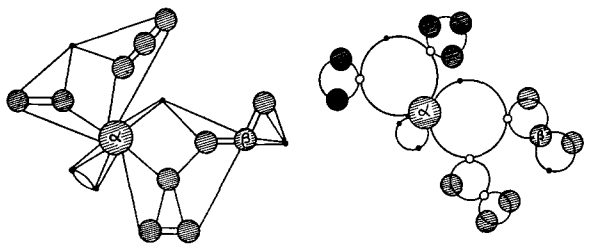


FIG. 13. (a) Tree of parquets: the 3-fold 2-skeleton α belongs to three parquets P_3, P_1, P_4 (labeled according to the number of cycles with $\mu \geq 3$ which they contain); the 2-fold 2-skeleton β belongs to two parquets P_4, P_1'. (b) "Structure diagram" for (a). The four parquets P_3, P_1, P_4, P_1' contain respectively 3, 1, 4, and 1 circles of pairs; P_3, P_1, P_4 share the shaded bubble α; P_4 and P_1' share the shaded bubble β.

A *parquet* is represented by a *tree of circles of pairs.*

(*2*) *Several parquets. Definition:* A *p-fold 2-skeleton* is a topological element connected to the rest of the diagram by p quadruplets of lines. The remainder of the diagram separates into p different 2-insertions, one for each quadruplet; if one of these p 2-insertions is suppressed and replaced by a 2-vertex, a $(p-1)$-fold 2-skeleton is obtained. When $p = 1$ we have the above-defined 2-skeleton. By convention a 2-irreducible 0-diagram is a 0-fold 2-skeleton.

A 0-diagram contains in general several parquets; a 2-vertex can only belong to a single parquet; a p-fold 2-skeleton[34] belongs to p different parquets [Fig. 13(a)] just as in Sec. 1 of this appendix, a q-fold 1-skeleton belonged to q different cycles of lines. In terms of the structure diagrams, two parquets cannot have in common a black or a white dot, they can only share a shaded bubble [Fig. 13(b)].

From the definitions it follows that two parquets, cannot have in common more than one p-fold 2-skeleton. Equivalently, parquets cannot form a ring; they necessarily form a tree.

A 0-diagram has the structure of a *tree of parquets.*

(*c*) *Topological Relations*

(*1*) For a tree of parquets, the following (Euler–Poincaré) relation is again immediately proved by induction:

$$N(P) - N(\mathcal{K}^{(2)} \cdot p) + N(\mathcal{K}^{(2)}) = +1. \qquad \text{(B4)}$$

$N(P)$ is the number of parquets, $N(\mathcal{K}^{(2)})$ is the number of shaded bubbles (i.e., p-fold 2-skeletons $N(\mathcal{K}^{(2)} \cdot p)$ is the number of shaded bubbles weighted by their multiplicity p. The relation (B4) can also be written

$$\sum_P [1 - N_P(\mathcal{K}^{(2)})] + N(\mathcal{K}^{(2)}) = +1, \qquad \text{(B5)}$$

where $N_P(\mathcal{K}^{(2)})$ is the number of shaded bubbles belonging to the parquet P.

(*2*) Consider now a *single parquet.* Being a tree of circles of pairs we again have

$$N\ (\text{Circles}) - N\ (\text{vertices} \cdot \text{multiplicity}) + N\ (\text{vertices}) = +1. \qquad \text{(B6)}$$

Among the vertices, the black dots and shaded bubbles have multiplicity one; the white dots have multiplicity two. Relation (B6) can be written

$$\sum_{\Gamma \in P} [1 - N_\Gamma(\pi)] + [N_P(v_2) - N_P(v_2^i)] + N_P(\mathcal{K}^{(2)}) + N_P(\text{links}) = 1. \qquad \text{(B7)}$$

In the single parquet P considered, the displayed cycles γ_μ ($\mu > 2$ except when P reduces to γ_2's) are denoted by Γ_μ. $N_\Gamma(\pi)$ is the number of vertices in each circle of pairs, that is, the number of pairs of each cycle Γ_μ; $[N_P(v_2) - N_P(v_2^i)]$ is the number of black dots. In a parquet diagram this is the number of 2-vertices not counting the ones which only belong to a trivial γ_2, i.e., the 2-vertices "internal" to each 2-skeleton. $N_P(\mathcal{K}^{(2)})$ is the number of shaded bubbles (2-skeletons).

Consider a set of four lines in a 0-diagram. It is called an articulation quartet of lines if the 0-diagram separates into two disconnected parts, when the four lines are cut. We introduce now the number of ways that a parquet diagram can be separated into two disconnected parts by cutting four lines belonging to a *nontrivial* cycle. This is also the number of articulation quartets of lines of a parquet diagram minus the number of 2-vertices internal to 2-skeletons, that is, $N_P(\text{art q}) - N_P(v_2^i)$. In a parquet we have

$$[N_P(\text{art q}) - N_P(v_2^i)] = -N_P(\text{links}) + \sum_{\Gamma \in P} N_\Gamma((\pi)^2). \qquad \text{(B8)}$$

$N_\Gamma((\pi)^2)$ is the number of pairs of pairs of lines in each cycle Γ_μ ($\frac{1}{2}[\mu(\mu-1)]$ in number); relation (B8) follows from the fact that wherever we have a link, we have four lines belonging to *two* different cycles.

[34] Notice that a p-fold 2-skeleton may reduce to a mere topological relation (i.e., contain no actual element of the diagram). For example, consider the 0-diagram generated out of the lowest-order 2-irreducible 0-diagram [first diagram in Eq. (3.53)] by replacing each of its five 2-vertices by a 2-skeleton [Fig. 12(b)]. Its structure diagram [Fig. 12(b')] involves five 2-skeletons connected through one 5-fold 2-skeleton; this 5-fold 2-skeleton reduces to a mere topological relation. This was not the case for a q-fold 1-skeleton since it necessarily contains 2-vertices.

Relation (B7) becomes

$$1 - N_P(\mathcal{K}^{(2)}) = \sum_{\Gamma \in P} [1 - N_\Gamma(\pi) + N_\Gamma((\pi)^2)]$$
$$- [N_P(\text{art q}) - N_P(v_2^i)] + [N_P(v_2) - N_P(v_2^i)]. \quad \text{(B9)}$$

This relation is verified for each parquet, and is additive. As a consequence, Eq. (B5) can be transformed into

$$\sum_P \{ \sum_{\Gamma \in P} [1 - N_\Gamma(\pi) + N_\Gamma((\pi)^2)]$$
$$- [N_P(\text{art q}) - N_P(v_2^i)]$$
$$+ [N_P(v_2) - N_P(v_2^i)]\} + N(\mathcal{K}^{(2)}) = +1. \quad \text{(B10)}$$

Thus, for a general 0-diagram,

$$\sum_\Gamma [1 - N_\Gamma(\pi) + N_\Gamma((\pi)^2)] - N(\text{art q})$$
$$+ N(\mathcal{K}^{(2)}) + N(v_2) = +1. \quad \text{(B11)}$$

The restriction in the summation, to the explicitly displayed cycles Γ, is in fact irrelevant, since the bracket vanishes for $\mu = 1$ and $\mu = 2$. Thus (B11) becomes identical with (56).

The proof given here is preferable to the one given in reference [13] in that it is less *ad hoc*, following logically from the analysis of the structure of a 0-diagram.

3. Analysis of 2-Insertions: Superfluid Systems

We extend (56) to $\frac{3}{2}$-irreducible 0-diagrams.

(a) *Cycles of Pairs of Lines*

Again, the pairs of lines of a ($\frac{3}{2}$-irreducible) 0-diagram can be unambiguously assigned to cycles of pairs [Secs. VA(1) and B(1)]. The tetrahedron-like 0-diagram is the only exceptional case. Indeed by cutting open two opposite lines in it one obtains a 2-diagram contributing to Q_2 [Eq. (67)]. This 2-diagram is nonsimple in two distinct ways and

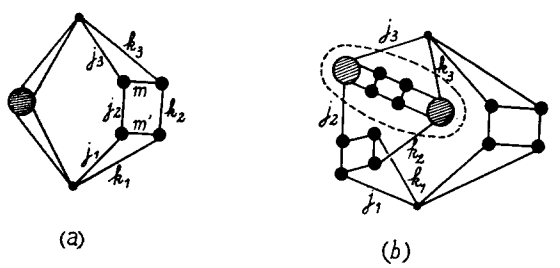


FIG. 14. (a) Cycle γ_5: (j_1, k_1), (j_2, k_2), (j_3, k_3), $\cdots$; five pairs of lines connect one 2-skeleton, two 2-vertices, and two pseudo 2-vertices. (b) Cycle γ_6: (j_1, k_1), (j_2, k_2), (j_3, k_3), $\cdots$, the six pairs connect successively a diamond simple 2-diagram, a simple 2-diagram (inside the dotted line), a 2-vertex, . . . etc. Cycle γ_5: (j_2, j_3), (k_2, k_3), $\cdots$; γ_6 and γ_5 have in common (j_2, j_3, k_2, k_3). There are also four cycles $\bar{\gamma}_3$ and two of them contain the four lines (j_1, j_2, k_1, k_2).

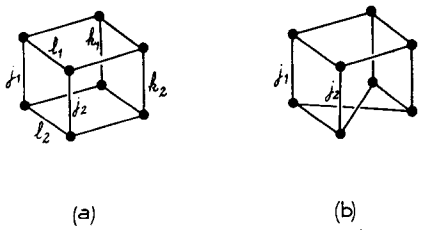


FIG. 15. (a) Cube: three distinct $\bar{\gamma}_4$ containing, respectively, (j_1, j_2), (l_1, l_2), (j_2, k_2). (b) Twisted cube: single $\bar{\gamma}_4$ containing (j_1, j_2).

therefore the distribution of its pairs into cycles is not unique. In the following, we consider 0-diagrams, excluding the *tetrahedron*.

The analysis of the cycle structure of a 0-diagram follows the same patterns as in the previous section of this appendix, only here the situation is slightly more complicated due to the presence of pseudo 2-vertices.

Consider again a cycle γ_μ. The μ pairs of the cycle may connect only 2-skeletons, 2-vertices or pseudo 2-vertices. In this case, consider two lines j_1, j_i chosen in two different pairs of γ_μ. These two lines may belong to

(a) a γ_2 as in normal diagrams or as (j_1, j_2), (k_1, k_2) in Fig. 14(a);

(b) a γ_3 as (j_1, j_3), (m, m'), (k_1, k_3) in Fig. 14(a). In this case the pair (m, m') does not belong to γ_μ. We denote by $\bar{\gamma}_3$ this particular type of cycle γ_3 which connects three 2-insertions among which there are at least two pseudo 2-vertices;

(c) a γ_4 when the 0-diagram considered is a cube diagram and j_1, j_2 are two parallel lines of this cube. We denote by $\bar{\gamma}_4$ the particular cycles γ_4 which connect four pseudo 2-vertices in the two possible cubelike diagrams (Fig. 15).

The most general situation occurs when the μ pairs connect μ 2-insertions which are not necessarily 2-irreducible, but need only be *simple* with respect to the pairs of γ_μ through which they are connected to the rest of the diagram. Consider one of these simple 2-insertions S and let (j_1, k_1), (j_2, k_2) be the pairs of γ_μ connecting it to the rest of the diagram. In general, S can be nonsimple with respect to only one of the two other pairings $(j_1, j_2; k_1, k_2)$, $(j_1, k_2; j_2, k_1)$; this property was satisfied in all normal diagrams. Here however, if S is *the one* simple 2-insertion which belongs to Q_2, S is nonsimple with respect to *both* pairings $(j_1, j_2; k_1, k_2)$ and $(j_1, k_2; j_2, k_1)$ [Fig. 14(b)]. This peculiar simple 2-insertion in which each one of the pairs (j_1, k_1), (j_2, k_2) ends up at opposite corners of a square

(a) (b)

FIG. 16. (a) Structure diagram for Fig. 7(a). (b) Structure diagram for Fig. 7(b); bars across a circle stand for pseudo 2-vertices; a diamond stands for a diamondlike 2-insertion.

(built with four heavy $\frac{3}{2}$-vertices), we call a diamond simple 2-insertion.

Again we consider two cycles γ_{μ_1}, γ_{μ_2} and proceed as in Sec. 2(a) of this appendix to analyze how many lines these two cycles may have in common. The analysis follows the same pattern (only here one has to pay a special attention to the peculiar cycles $\bar{\gamma}_3$, $\bar{\gamma}_4$), the possible situations are now[35]

(i) and (ii) identical to the corresponding (i) and (ii) of Sec. 2 of this appendix;

(iii) γ_{μ_1} and γ_{μ_2} may have four lines in common always distributed into two pairs; these two pairs are successive in $\gamma_{\mu_i}(i = 1, 2)$ if $\mu_j > 2$ $(j \neq i)$, and if $\gamma_{\mu_j} \neq \bar{\gamma}_3, \bar{\gamma}_4$.

The definitions of linked cycles, connected sets of cycles, and parquets is unchanged.

(b) *Structure of a 0-Diagram*

In the "structure diagrams" of Sec. 2, this appendix, a cycle of pairs γ_μ was represented by a circle of pairs, that is a set of μ lines connecting μ black dots (2 vertices), shaded bubbles (2 skeletons). Circles of pairs were connected by white dots (links). Here, however, the cycles $\bar{\gamma}_3$, $\bar{\gamma}_4$ cannot be explicitly represented with the topological links previously used since when $\bar{\gamma}_3$, $\bar{\gamma}_4$ share two pairs with a γ_μ, these two pairs do not necessarily occur *in succession*.

In Sec. 2 the cycles γ_2 presented the same problem [four lines common to a γ_2, and a γ_μ $(\mu > 2)$ did not necessarily constitute two successive pairs in γ_μ]. This problem was solved by noticing that in general the γ_2's were embedded in cycles γ_μ $(\mu > 2)$ so that, it was not necessary to display the γ_2's explicitly.

In the same fashion we shall embed the $\bar{\gamma}_3$'s in cycles Γ_μ $(\mu > 2, \Gamma_\mu \neq \bar{\gamma}_3, \bar{\gamma}_4)$. The $\bar{\gamma}_3$'s appear in a diagram when a 2-diagram belonging to Q_2 is connected by two pairs of a cycle to the rest of the diagram; that is when two successive pairs of a cycle connect a diamond simple 2-diagram, or when three successive pairs connect two pseudo 2-vertices. In the "structure diagrams", the $\bar{\gamma}_3$'s will be embedded in a faithful way if we now represent each Γ_μ by a circle of pairs where in addition to the black dots and shaded bubbles we have diamonds (representing diamond simple 2-diagrams) and bars (each bar representing a pseudo 2-vertex connected by two pairs of Γ_μ) (Fig. 16). The $\bar{\gamma}_4$'s occur only in the two *cubelike* 0-diagrams (Fig. 15) which, like the tetrahedron, are treated separately.

(a) (b)

FIG. 17. (a) Parquet reducing to two $\bar{\gamma}_3$' (and γ_2's) and its structure diagram representation in (b).

Again the only cases where cycles $\bar{\gamma}_3$, γ_2 need be displayed are cases in which parquets reduce to $\bar{\gamma}_3$'s and γ_2's, since only then is it not possible to embed the $\bar{\gamma}_3$'s and γ_2's in Γ_μ's. Parquets reducing to $\bar{\gamma}_3$'s and γ_2's correspond to parquet diagrams where a quadruplet of lines connect

(i) either two 2-skeletons or two 2-vertices. These cases have been examined in Sec. 2 of this appendix [Fig. 12(a)];

(ii) one 2-vertex and one pseudo 2-vertex, a case which is disposed of similarly;

(iii) one 2-skeleton (or one 2-vertex) and one 2-insertion belonging to Q_2; this case is represented as in Figs. 17(a) and 17(b) (one γ_2 explicitly represented, two $\bar{\gamma}_3$'s embedded);

(iv) one pseudo 2-vertex and one 2-insertion belonging to Q_2 [the *twisted triangular prism*, Fig. 18(a)]; this diagram contains six $\bar{\gamma}_3$'s and it is simpler to consider it apart with the other exceptional cases already noted (tetrahedron and cubelike diagrams). This is the last exceptional diagram which we shall encounter.

(a) (b)

FIG. 18. (a) Twisted triangular prism; notice that the triangular prism (b) is $\frac{3}{2}$-reducible and is to be discarded.

[35] Again, when γ_{μ_1} or γ_{μ_2} (or both) is a γ_2, a $\bar{\gamma}_3$, or a $\bar{\gamma}_4$, the result remains the same.

Again, a parquet is a tree of circles of pairs, and a general 0-diagram is a tree of parquets.

(c) *Topological Relations*

Relations (B4), (B5), (B6) are unchanged. (B7) is now written for a single parquet,

$$\sum_{\Gamma\in P}[1 - N_\Gamma(\pi)] + [N_P(v_2) - N_P(v_2^i)] + [N_P(p_2) - N_P(p_2^i)] + N_P(\mathcal{K}^{(2)}) + N_P(Q_2^s) + N_P(\text{links}) = 1. \quad \text{(B12)}$$

The Γ's are the explicitly displayed cycles Γ_μ ($\mu > 2$, $\bar\gamma_3$ excluded) of the parquet P considered. Here we have introduced $N_P(p_2) - N_P(p_2^i)$, the number of bars; that is, the number of pseudo 2-vertices in a parquet diagram, excluding the ones "internal" to a 2-skeleton or to a diamond 2 insertion and those which play the role of lines in displayed cycles Γ_μ. For example, in Figs. 14(b) and 16(b), $N_P(p_2) - N_P(p_2^i) = 4$. $N(Q_2^s)$ is the number of diamonds (diamond simple 2-insertions), that is, how many times (in the *displayed* cycles Γ_μ) the simple 2-insertion which belongs to Q_2 connects two pairs of a Γ_μ.

Relation (B8) now becomes

$$N_P(\text{art q}) - N_P(v_2^i) - N_P(p_2^i) = -N_P(\text{links}) + \sum_{\Gamma\in P} N_\Gamma[(\pi)^2], \quad \text{(B13)}$$

and leads, instead of (B11), to

$$\sum_\Gamma \{1 - N_\Gamma(\pi) + N_\Gamma[(\pi)^2]\} + N(Q_2^s) - N(\text{art q}) + N(\mathcal{K}^{(2)}) + N(v_2) + N(p_2) = 1. \quad \text{(B14)}$$

The cycles Γ are the ones explicitly displayed in the structure diagram. Let $N(Q_2)$ be the number of ways in which, by cutting four lines, one can separate in a 0-diagram a 2-insertion contributing to Q_2. The following relation holds:

$$N(Q_2) = \sum_{\bar\gamma_3} \{1 - N(\pi) + N[(\pi)^2]\} - N(Q_2^s), \quad \text{(B15)}$$

where the summation is extended over all $\bar\gamma_3$ cycles of a 0-diagram. Indeed there only are two distinct ways of having $\bar\gamma_3$ cycles in a 0-diagram. Expressed in terms of the structure diagrams, a $\bar\gamma_3$ is either embedded in two consecutive bars [and the pairs connecting them to the rest of the diagram; e.g., $(j_1, j_3; m', m; k_1, k_3)$ in Figs. 14(a) and 16(a)]. In this case,

$$\{1 - N(\pi) + N[(\pi)^2]\}_{\bar\gamma_3} = 1, \qquad N(Q_2^s) = 0,$$

or $\bar\gamma_3$ is embedded in a diamond (and the pairs connecting it to the rest of the diagram). In this case there always are *two* $\bar\gamma_3$'s [respectively containing $(j_1, j_2; k_2, k_1)$ and $(j_1, k_2; j_2, k_1)$ in Figs. 14(b) and 16(b)]. In this case we have

$$\{1 - N(\pi) + N[(\pi)^2]\}_{\bar\gamma_3} = 2, \qquad N(Q_2^s) = 1.$$

There are no other possible ways on which a 2-insertion contributing to Q_2 can appear in a 0-diagram; and so (B15) is proven. We finally can rewrite (B14) as

$$\sum_\gamma \{1 - N_\gamma(\pi) + N_\gamma[(\pi)^2]\} - [N(\text{art q}) - N(p_2)] + N(\mathcal{K}^{(2)}) + N(v_2) - N(Q_2) = +1, \quad \text{(B16)}$$

where the summation now extends over all cycles. This relation is valid for all $\frac{3}{2}$-irreducible 0-diagrams, save for the exceptional diagrams encountered above; these must be considered separately. The final equation (75) follows directly form (B16).

Notice that the twisted cube [Fig. 15(b)] occurs in the algebraic expression associated with the first bracket of (B16) with its proper weight ($S = 16$). Hence only the tetrahedron, the twisted triangular prism, and the cube diagrams need be treated separately.

On the van der Waals Theory of the Vapor–Liquid Equilibrium. III. Discussion of the Critical Region

P. C. HEMMER,* M. KAC, AND G. E. UHLENBECK
The Rockefeller Institute, New York, New York
(Received 20 September 1963)

The discussion of the properties of the Kac one-dimensional fluid model presented in Parts I and II of this series of papers breaks down near the critical point. In Sec. II of the present paper we develop a new successive-approximation method for the eigenvalues and eigenfunctions of the Kac integral equation which is valid in the critical region and which connects smoothly with the developments in the one- and two-phase regions given in Part I. The perturbation parameter is $(\gamma\delta)^{\frac{1}{2}}$ where $\gamma\delta$ is the ratio of the ranges of the repulsive and attractive forces. The main physical consequence is that in the critical region the long-range behavior of the two-point distribution function is represented by an infinite series of decreasing exponentials with ranges all of order $1/\gamma(\gamma\delta)^{\frac{1}{2}}$, and with amplitudes of order $(\gamma\delta)^{\frac{1}{2}}$. This leads to deviations from the Ornstein–Zernike theory and to a specific heat anomaly which are discussed in Sec. V. We conclude with some comments on the possible relevance of our results for the three-dimensional problem.

I. INTRODUCTION

As we noted already in the first two parts of this series of papers,[1] the discussion of the approach to the van der Waals limit which was given there breaks down near the critical point. This is evident from the way the quantity B^2, which is proportional to $\partial p/\partial l$, enters into the expansions for the eigenvalues and eigenfunctions of the Kac equation and into the expansions for the short- and long-range behavior of the two-point distribution function. For $B = 0$ all these expansions blow up. It is therefore necessary to develop a new asymptotic treatment of the Kac equation in the critical region and this is presented in this paper.

It turns out that the critical region should be defined by the ranges in temperature and specific volume given by

$$(T - T_c)/T_c \approx (\gamma\delta)^{\frac{1}{2}}, \qquad (l - l_c)/l_c \approx (\gamma\delta)^{\frac{1}{2}}, \qquad (1)$$

where T_c and l_c are the van der Waals critical values and where $\gamma\delta$ is again the ratio of the ranges of the repulsive and attractive forces. In Sec. II we will show that in this region one can again "tame" the Kac equation by an appropriate change of variables, and that this leads to a consistent successive-approximation method for the eigenfunctions and eigenvalues. In contrast to the one- and two-phase regions, the expansion parameter is no longer $\gamma\delta$ but $(\gamma\delta)^{\frac{1}{2}}$, and also the zeroth-order eigenfunctions are no longer the harmonic oscillator eigenfunctions (≡ Weber functions), but correspond to the eigenfunctions of an oscillator for which the potential energy is a fourth-degree polynomial in the deflection. As a result, the eigenfunctions and eigenvalues can no longer be determined explicitly. However, it is possible to discuss qualitatively how the eigenfunctions change when one goes away from the critical region, and it is of special interest to note the difference in behavior when one goes towards the one-phase region or towards the two-phase region. In the first case, the eigenfunctions become again the Weber functions with only slight distortions, but in the second case (two-phase region) they approach the eigenfunctions for a potential which has two equal or almost equal minima. It is well known that for such a two-minimum potential, the lowest eigenvalue is almost doubly degenerate, and this corresponds to the beginning of the degeneracy which is characteristic for the two-phase region.

In Sec. III the consequences are discussed with respect to the form of the isotherms in the critical region. In I we have already pointed out that, to any order in $\gamma\delta$, the successive-approximation method always gives a phase transition and a van der Waals-like equation of state. We now find that in the critical region (i.e., for finite $\gamma\delta$), there is a *qualitative* difference of the isotherm net from the prediction of the simple van der Waals theory, (i.e., for $\gamma\delta = 0$). For finite $\gamma\delta$, the transition from the one-phase to the two-phase regime occurs continuously in the critical region.[2] Strictly speaking,

* Present address: Institute for Theoretical Physics, N.T.H. Trondheim, Norway.

[1] J. Math. Phys. 4, 216, 229 (1963), hereafter referred to as I and II, respectively.

[2] The concept of a critical *region* in contrast to a critical *point* has often been discussed especially in order to explain various anomalous critical phenomena which apparently were in conflict with the van der Waals equation. For a review

for $T < T_c$ the isotherms do *not* have a horizontal portion, and as a result the densities of the two coexistent phases cannot be defined unambiguously. However, although it is difficult to make precise statements because of the uncertainty one always encounters when one tries to join two different asymptotic developments of the same function, it is possible to make a reasonable extrapolation of the locus of the coexistent points from the two-phase region into the critical region. We then find that, for finite $\gamma\delta$, the critical density remains at the van der Waals value, but the critical temperature is lowered by ΔT_c with $\Delta T_c/T_c \cong 0.702(\gamma\delta)^{\frac{1}{3}}$. At the new critical point, the locus of the coexistent points is still parabolic but it is much flatter than the van der Waals theory predicts.[3]

In Sec. IV and V the consequences are discussed with respect to the two-point distribution function. The main result is that in the critical region the long-range behavior of this distribution function is represented by an infinite series of decreasing exponentials. The ranges of these exponentials are all of order $1/\gamma(\gamma\delta)^{\frac{1}{3}}$, and also the amplitudes are all of the same order of magnitude, namely $(\gamma\delta)^{\frac{1}{3}}$. The Ornstein–Zernike theory is therefore *not* valid in the critical region. However, we can show that if one goes away from the critical region towards the one-phase region, our results go over into the expansions of the long-range behavior of the correlation function derived in II, of which the Ornstein–Zernike exponential was the first term. The situation seems to be that, if one approaches the critical point, more and more exponential terms become excited, so to say. *Outside* the critical region these terms are of decreasing order of magnitude, but *in* the critical region they all contribute.

It seems to us that this picture of how the deviations from the Ornstein–Zernike theory develop if one approaches the critical point, is probably also valid in three dimensions, and that it is independent of our assumption of an exponential attractive potential. In fact, we show in Sec. VI that the whole discussion of Sec. IV can be generalized to the case where the attractive potential consists of a sum of m exponentials, and that this does not change the qualitative picture.

In Sec. VII we conclude with a few comments on some recent experimental results on the critical opalescence and on the so-called specific-heat anomaly in order to see whether there are some experimental indications for the existence of a critical region.

II. THE EIGENFUNCTIONS AND EIGENVALUES IN THE CRITICAL REGION

We will develop all quantities around the critical point of the zeroth-order equation of state, that is the van der Waals equation

$$s = 1/(l - \delta) - \nu_0/l^2,$$

in the notation used in I and II. The critical quantities are given by:

$$s_c = 1/8\delta; \qquad l_c = 3\delta; \qquad \nu_{0c} = \tfrac{27}{8}\delta. \tag{2}$$

Instead of I, Eq. (33), it turns out that around the critical point one can "tame" the basic Kac integral equation [I, Eq. (10)] by the substitution

$$z = (\gamma\delta)^{\frac{1}{3}}[x - \eta_c(2/\gamma)^{\frac{1}{3}}], \tag{3}$$

where $\eta_c \equiv \eta(s_c) = (2\nu_{0c})^{\frac{1}{3}}/l_c$. This is motivated by the fact that in the one-phase region the "tamed" eigenfunctions depend on the variable [see I, Eq. (38)]

$$z = (B/l)^{\frac{1}{3}}[x - \eta(2/\gamma)^{\frac{1}{3}}]. \tag{4}$$

If one now *defines* the critical region in temperature and specific volume by the equations corresponding

of the older literature, see for instance J. P. Kuenen, *Die Zustandsgleichung* (Vieweg, Braunschweig, 1907), Chap. 5, especially for the effects of gravity and of impurities on the critical phenomena. The only attempt we know of to deduce a critical region from the ideas of van der Waals is by G. Bakker [Z. Physik. Chem. **49**, 609 (1904)]. He pointed out that there may be a range in temperature where the thickness of the capillary layer is of the same order as $v_1^{\frac{1}{3}} - v_2^{\frac{1}{3}}$. In this region one would not see a meniscus, although the two phases are still present. This is also suggested by Mayer. For a review of his arguments, which are partially formal and partially physical, and are based on the fugacity and virial expansions, see the book by J. E. Mayer and M. G. Mayer [*Statistical Mechanics*, (John Wiley & Sons, Inc., New York, 1940), Chap. 14]. Mayer's conclusion is that there may be a temperature $T_m < T_c$ where the meniscus would disappear, although there are still horizontal parts in the isotherms for temperatures between T_m and T_c. In this temperature range, the isotherms enter the two-phase region with a horizontal tangent. Because of the resulting shape of the coexistence curve near the critical point, this region is often referred to as Mayer's derby hat.

From the experimental point of view the existence of such a critical region is still controversial. For a review of recent experimental results see the book by T. O. Hirschfelder, Ch. F. Curtiss, and R. B. Bird [*Molecular Theory of Gases and Liquids* (John Wiley & Sons, Inc., New York, 1954), Chap. 5, Sec. 2]. See also the work by H. W. Habgood and W. G. Schneider on Xenon [Can. J. Chem. **32**, 98, 164 (1954)].

[3] The question of the shape of the coexistence curve near the critical point has been much discussed. Guggenheim [J. Chem. Phys. **13**, 253 (1945)], pointed out that the data are better represented by the law $(v_1 - v_2) \sim (T_c - T)^{\frac{1}{3}}$ than by the van der Waals result $(v_1 - v_2) \sim (T_c - T)^{\frac{1}{2}}$ which follows from the parabolic form of the coexistence curve near the critical point. Compare also M. A. Weinberger and W. G. Schneider, Can. J. Chem. **30**, 422 (1952), and the more recent discussion by B. Widom and O. K. Rice, J. Chem. Phys. **23**, 1250 (1955). The Guggenheim result would imply that at the critical point also $(\partial^3 p/\partial v^3)_T$ would be zero, and B. H. Zimm [J. Chem. Phys. **19**, 1019 (1951)] has given a plausibility argument for the vanishing of *all* derivatives at the critical point. However, also in this respect, there are no results which are generally accepted.

to (1),

$$\nu_0 = \nu_{0c}[1 + \nu_1(\gamma\delta)^{\frac{1}{3}}], \tag{5}$$

$$l = l_c[1 + l_1(\gamma\delta)^{\frac{1}{3}}], \tag{6}$$

then one finds that B/l is of order $(\gamma\delta)^{\frac{1}{3}}$, so that (4) goes over into (3). However, one must of course still show that the substitution (3) works! We proceed in the same way as in I. For the variable y in the Kac equation, one makes the analogous substitution

$$z' = (\gamma\delta)^{\frac{1}{3}}[y - \eta_c(2/\gamma)^{\frac{1}{3}}],$$

and we put

$$z' = ze^{-\gamma\tau} + (\gamma\delta)^{\frac{1}{3}}\zeta(1 - e^{-2\gamma\tau})^{\frac{1}{3}},$$

$$\psi(x) = \psi[(\gamma\delta)^{-1/6}z + \eta_c(2/\gamma)^{1/2}] \equiv (\gamma\delta)^{1/12}H(z),$$

where the factor $(\gamma\delta)^{1/12}$ is introduced in order to keep $H(z)$ normalized to one. Finally we put for s the expansion[4]

$$s = s_c[1 - 3\nu_1(\gamma\delta)^{2/3} + s_1(\gamma\delta) + s_2(\gamma\delta)^{4/3} + \cdots], \tag{7}$$

where the numbers s_1, s_2 $\cdots$ will depend on the given *numbers* ν_1 and l_1 through the equation of state, which is still given by [see I, Eq. (20)]

$$l = -\lambda_0'(s)/\lambda_0(s), \tag{8}$$

in terms of the maximum eigenvalue $\lambda_0(s)$ of the Kac equation.

Making all these substitutions, introducing the expansions (5) and (7), and developing the kernel in powers of $(\gamma\delta)^{1/3}$ up to order $(\gamma\delta)^{5/3}$, one finds, after carrying out the integrations over ζ and τ and using (2),

$$\frac{d^2H}{dz^2}[3(\gamma\delta)^{4/3} - (6)^{1/2}z(\gamma\delta)^{5/3}] - (6)^{1/2}(\gamma\delta)^{5/3}\frac{dH}{dz}$$
$$+ H(z)\Big[3\Lambda + \Big(\frac{3}{2}\Big)^{1/2}z\Big\{\Big(\frac{s_1}{4} - 1\Big)(\gamma\delta)^{4/3}$$
$$+ \Big(\frac{s_2}{4} - \frac{39}{16}\nu_1^2\Big)(\gamma\delta)^{5/3}\Big\} + z^2\Big\{\frac{3\nu_1}{4}(\gamma\delta)^{4/3}$$
$$+ \Big(1 - \frac{s_1}{4}\Big)(\gamma\delta)^{5/3}\Big\} - \frac{3}{8}\Big(\frac{3}{2}\Big)^{1/2}\nu_1(\gamma\delta)^{5/3}z^3$$
$$- \frac{1}{16}(\gamma\delta)^{4/3}z^4 + \frac{1}{20}\Big(\frac{3}{2}\Big)^{1/2}(\gamma\delta)^{5/3}z^5\Big] = 0 \tag{9}$$

where the new eigenvalue parameter Λ is related to the original λ by the equation

$$\frac{\lambda}{\omega(s_c)} = 1 + \frac{9\nu_1}{4}(\gamma\delta)^{2/3} + \frac{3}{2}\Big(1 - \frac{s_1}{4}\Big)\gamma\delta$$
$$+ \frac{3}{8}\Big(-s_2 + \frac{27}{4}\nu_1^2\Big)(\gamma\delta)^{4/3} + \cdots - 3\Lambda. \tag{10}$$

From (9) one sees that one obtains a consistent approximation procedure by putting

$$H(z) = H^{(0)}(z) + (\gamma\delta)^{\frac{1}{3}}H^{(1)}(z)$$
$$+ (\gamma\delta)^{\frac{2}{3}}H^{(2)}(z) + \cdots,$$
$$\Lambda = \Theta(\gamma\delta)^{4/3} + \Omega(\gamma\delta)^{5/3} + \Xi(\gamma\delta)^2 + \cdots. \tag{11}$$

In lowest order [i.e., $(\gamma\delta)^{4/3}$] one gets

$$\Big[\frac{d^2}{dz^2} - \frac{z^4}{48} + \frac{\nu_1}{4}z^2$$
$$+ (6)^{-\frac{1}{3}}\Big(\frac{s_1}{4} - 1\Big)z + \Theta\Big]H^{(0)}(z) = 0, \tag{12}$$

and in next order [i.e., $(\gamma\delta)^{5/3}$],

$$\Big[\frac{d^2}{dz^2} - \frac{z^4}{48} + \frac{\nu_1}{4}z^2 + (6)^{-\frac{1}{3}}\Big(\frac{s_1}{4} - 1\Big)z + \Theta\Big]H^{(1)}(z)$$
$$= \Big[\Big(\frac{2}{3}\Big)^{\frac{1}{3}}\frac{d}{dz}z\frac{d}{dz} - \frac{z^5}{20(6)^{\frac{1}{3}}} + \frac{\nu_1}{8}\Big(\frac{3}{2}\Big)^{\frac{1}{3}}z^3$$
$$+ \frac{s_1 - 4}{12}z^2 - \Big(\frac{2}{3}\Big)^{\frac{1}{3}}\Big(\frac{s_2}{8} - \frac{39}{32}\nu_1^2\Big)z - \Omega\Big]H^{(0)}(z). \tag{13}$$

These equations determine, in principle, the eigenfunctions $H_n^{(0)}(z)$, $H_n^{(1)}(z)$ and the corresponding eigenvalues Θ_n, Ω_n as functions of the constants ν_1, s_1, and s_2.

Turning now to the equation of state, we start from the general expression

$$-\lambda_0'(s) = \iint_{-\infty}^{+\infty} dx\,dy\psi_0(x)\psi_0(y)\Big[\frac{W(x)}{W(y)}\Big]^{\frac{1}{3}}$$
$$\exp\Big\{\frac{(\nu_0\gamma)^{\frac{1}{3}}}{2}(x + y)\Big\}\int_s^\infty d\tau\,\tau e^{-s\tau}P_\tau(x \mid y, \tau).$$

By expanding ν_0 and s around the critical values, and by using the same taming substitutions for x and y ,one obtains:

$$-\lambda_0'(s) = \omega(s_c)\Big[3\delta - \delta(6)^{\frac{1}{3}}(\gamma\delta)^{\frac{1}{3}}\int_{-\infty}^{+\infty} dz\,zH_0^{(0)}(z)^2$$
$$+ (\gamma\delta)^{\frac{1}{3}}\Big\{-2\delta(6)^{\frac{1}{3}}\int_{-\infty}^{+\infty} dz\,zH_0^{(0)}(z)H_0^{(1)}(z)$$
$$+ 2\delta\int_{-\infty}^{+\infty} dz\,z^2H_0^{(0)}(z)^2 + \frac{33}{4}\delta\nu_1\Big\} + \cdots\Big]. \tag{14}$$

From (10) and the equation of state (8), one then finds that one must have, by comparing the result for l with the defining equation (6),

[4] It follows from the van der Waals equation and Eqs. (5) and (6), that in (7) the coefficient of the term of order $(\gamma\delta)^{\frac{1}{3}}$ must be $(-3\nu_1)$. It is also not difficult to show that only in this way one obtains a consistent approximation procedure for the Kac equation.

$$\int_{-\infty}^{+\infty} dz\, zH_0^{(0)}(z)^2 = -l_1(\tfrac{3}{2})^{\frac{1}{3}} \tag{15}$$

$$-2(\tfrac{2}{3})^{\frac{1}{3}} \int_{-\infty}^{+\infty} dz\, zH_0^{(0)}(z)H_0^{(1)}(z) + \frac{2}{3}\int_{-\infty}^{+\infty} dz\, z^2 H_0^{(0)}(z)^2 + \tfrac{1}{2}\nu_1 = 0, \tag{16}$$

and these equations determine, in principle, the dependence of s_1 and s_2 on ν_1 and l_1.

III. ISOTHERM NET IN THE CRITICAL REGION

Since the equation of state is expressed in such an implicit fashion in terms of the lowest eigenvalue and eigenfunction of Eq. (12) which can not be determined in closed form, it is convenient for the discussion to make use of the following physical analogy. Writing Eq. (12) in the form

$$d^2H^{(0)}/dz^2 + [\Theta - V(z) + Fz]H^{(0)} = 0, \tag{17}$$

with

$$V(z) = \tfrac{1}{48}z^4 - (\tfrac{1}{4}\nu_1)z^2, \quad F = (s_1 - 4)/4(6)^{\frac{1}{3}}, \tag{18}$$

one sees that Eq. (12) can be looked upon as a Schrödinger equation for a particle moving in the potential $V(z)$ and under the influence of a quasi-"electric" field of strength F. Furthermore it is easy to show that from Eq. (15) it follows that

$$l_1 = (\tfrac{2}{3})^{\frac{1}{3}}\, \partial\Theta_0/\partial F. \tag{19}$$

Hence l_1 is proportional to the "electric moment" produced by F, and the discussion of the isotherm becomes equivalent to the discussion of the Stark effect for the lowest state of the motion of the particle in the potential $V(z)$.

Consider first the case when F is small. Since $V(z) = V(-z)$, the lowest eigenfunction $H_0^{(0)}(z)$ is also symmetric in z when $F = 0$. Hence one has a quadratic Stark effect, so that $l_1 = 0$ for $F = 0$, which follows also immediately from (15). By the usual perturbation theory one finds

$$\Theta_0(F) = \Theta_0 - F^2 \sum_{n=1}^{\infty} \frac{1}{\Theta_n - \Theta_0} \times \left[\int_{-\infty}^{+\infty} dz\, zH_0^{(0)}H_n^{(0)}\right]^2 + O(F^4), \tag{20}$$

where the Θ_n and $H_n^{(0)}$ are the eigenvalues and eigenfunctions of the unperturbed equation. Therefore for $F = 0$, $\partial^2\Theta_0/\partial F^2 < 0$, and not only $\partial\Theta_0/\partial F = 0$, but also $\partial^3\Theta_0/\partial F^3 = 0$. This implies that at the critical density, that is for $l_1 = 0$, all isotherms have an inflection point with a negative slope. In fact one obtains from (19) and (20)

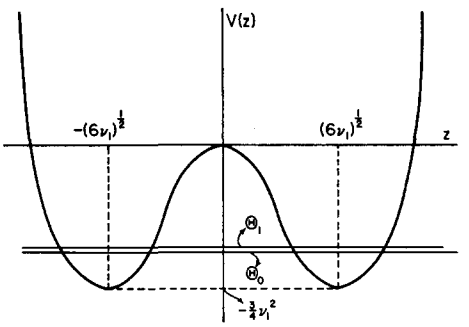


FIG. 1. The potential $V(z)$ of Eq. (17) for a temperature much below the critical temperature.

$$\left(\frac{\partial s}{\partial l}\right)_{l_1=0} = -\frac{(\gamma\delta)^{\frac{1}{3}}}{4\delta^2} \Big/ \sum_{n=1}^{\infty} \frac{1}{\Theta_n - \Theta_0} \times \left[\int_{-\infty}^{+\infty} dz\, zH_0^{(0)}H_n^{(0)}\right]^2, \tag{21}$$

which is always negative. The isotherms have therefore *no* horizontal part in the critical region. However the behavior of the slope (21) as function of the temperature is quite different for positive or negative ν_1, that is, for $T < T_c$ or $T > T_c$. Clearly, for negative ν_1 ($T > T_c$), the potential $V(z)$ has only a minimum at $z = 0$, and for $\nu_1 \ll -1$, $V(z)$ becomes a parabola and the eigenfunctions become the Weber functions. For $\nu_1 \ll -1$, it is simple to develop a perturbation method for the Θ_n and $H_n^{(0)}(z)$ and hence for the slope (21). One finds for $\nu_1 \ll -1$

$$\left(\frac{\partial s}{\partial l}\right)_{l_1=0} = \frac{(\gamma\delta)^{\frac{1}{3}}}{4\delta^2}\nu_1\left[1 + \frac{1}{2(-\nu_1)^{\frac{1}{2}}} + O(-\nu_1)^{-3}\right]. \tag{22}$$

It is also easy to see that this is in agreement with the equation of state in the one-phase region,

$$s = \frac{1}{l - \delta} - \frac{\nu_0}{l^2} + \frac{\gamma}{2}\left[1 - \frac{l}{B}\left\{1 - \frac{\nu_0(l^2 - \delta^2)}{l^3}\right\}\right], \tag{23}$$

with

$$B^2/l^2 = 1 - 2\nu_0(l - \delta)^2/l^3,$$

which was given in I, Sec. IV. Using (5) and (6) one obtains from (23) the expansion (7) for s with

$$s_1 = 6\nu_1 l_1 - \tfrac{3}{2}l_1^3 + 4 - 6l_1(3l_1^2 - 4\nu_1)^{-\frac{1}{2}}, \tag{24}$$

from which (22) follows.

The discussion is quite different for the case $\nu_1 > 0$, that is $T < T_c$. The potential $V(z)$ now has *two* minima (see Fig. 1), and for $\nu_1 \gg 1$, this leads to the well-known near degeneracy of the lowest energy level corresponding to the symmetric

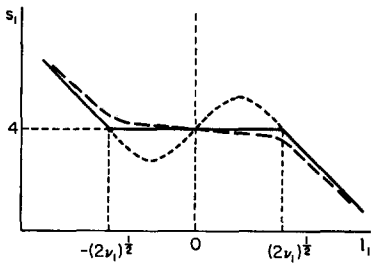


FIG. 2. Shape of the isotherm below the critical point; ——— Eq. (30); - - - - van der Waals; - - - - Eq. (30a).

and antisymmetric combination of the two lowest harmonic-oscillator eigenfunctions around the two minima. Using the WKB method[5] one finds

$$\Theta_1 - \Theta_0 \cong [(2\nu_1)^{\frac{1}{2}}/\pi] \exp[-(2\nu_1)^{\frac{1}{2}}],$$

$$\int_{-\infty}^{+\infty} dz\, zH_0^{(0)}H_1^{(0)} \cong (6\nu_1)^{\frac{1}{2}}.$$

Only the first term in the sum in (21) contributes, so that for $\nu_1 \gg 1$,

$$\left(\frac{\partial s}{\partial l}\right)_{l_1=0} \cong -\frac{(\gamma\delta)^{\frac{1}{2}}}{12\pi\delta^2(2\nu_1)^{\frac{1}{2}}} \exp[-(2\nu_1)^{\frac{1}{2}}]. \quad (25)$$

The slope of the isotherm drops therefore very fast to zero if ν_1 increases.

Consider next the case when F and therefore l_1 are not small. Again, to discuss $s_1(l_1, \nu_1)$, one must distinguish between the cases $\nu_1 \ll -1$ and $\nu_1 \gg 1$. For temperatures *above* $T_c(\nu_1 \ll -1)$, the "potential" function $V(z) - Fz$ has still only one minimum at $z = z_0$, determined by

$$\tfrac{1}{12}z_0^3 - (\tfrac{1}{2}\nu_1)z_0 - F = 0. \quad (26)$$

The "energy" Θ_0 is then given approximately by

$$\Theta_0 = V(z_0) - Fz_0 + 2^{-\frac{3}{2}}(z_0^2 - 2\nu_1)^{\frac{1}{2}}, \quad (27)$$

where the last term represents the "zero-point energy" $\frac{1}{2}\hbar\omega$ of the oscillation around z_0 [with $\hbar = 1$, $m = \frac{1}{2}$, potential energy $\frac{1}{4}\omega^2(z - z_0)^2 = \frac{1}{2}(d^2V/dz^2)_0(z - z_0)^2$, so that $\omega = 2^{\frac{1}{2}}(d^2V/dz^2)_0^{\frac{1}{2}} = 2^{-\frac{1}{2}}(z_0^2 - 2\nu_1)^{\frac{1}{2}}$]. From (19) and (27) one gets

$$(\tfrac{3}{2})^{\frac{1}{2}}l_1 = -z_0 + 2^{-\frac{3}{2}}z_0(z_0^2 - 2\nu_1)^{-\frac{1}{2}}\,\partial z_0/\partial F$$
$$= -z_0 + 2^{\frac{1}{2}}z_0/(z_0^2 - 2\nu_1)^{\frac{1}{2}},$$

using (26). In *zeroth* approximation, the minimum $z_0 = -(\frac{3}{2})^{\frac{1}{2}}l_1$, and substituting this in (26), and expressing F in s_1 by (18), one obtains the first three terms of the equation of state (24) which also follow from the van der Waals equation. One can say that these van der Waals terms correspond to the "classical" approximation of the Stark effect. In *first* approximation, z_0 becomes

$$z_0 = -(\tfrac{3}{2})^{\frac{1}{2}}l_1[1 - 2(6)^{\frac{1}{2}}l_1/(3l_1^2 - 4\nu_1)^{\frac{1}{2}}]$$

and this leads to the last term in the equation of state (24), which represents therefore, so to say, the first "quantum" correction to the van der Waals equation. Note that the corresponding result for the energy levels Θ_n is given by

$$\Theta_n = -\tfrac{9}{64}l_1^4 + \tfrac{3}{8}l_1\nu_1 - \nu_1/(3l_1^2 - 4\nu_1)^{\frac{1}{2}} + \tfrac{1}{2}n(3l_1^2 - 4\nu_1)^{\frac{1}{2}}, \quad (28)$$

and that in this approximation the eigenfunctions are the Weber functions $D_n(y)$, where

$$y = 2^{-\frac{1}{2}}(3l_1^2 - 4\nu_1)^{\frac{1}{4}}\{z + (3/2)^{\frac{1}{2}}l_1\}. \quad (29)$$

It is also not difficult to find the higher approximations. The results are collected in Appendix I, since they are needed in Sec. IV.

For temperatures *below* $T_c(\nu_1 \gg 1)$, the potential function $V(z) - Fz$ can have *two* minima which will be of unequal depth and the sign of F [or of $(s_1 - 4)$] determines which minimum is deeper. If in zeroth approximation we assume again that the energy Θ_0 is given by the value of the potential function at the deeper minimum, then one finds (in exactly the same way as above) for the equation of state

$$l_1 = -2\nu_1^{1/2}\,\mathrm{sign}\,(s_1 - 4) - (s_1 - 4)/12\nu_1 + O(\nu_1^{-5/2}), \quad (30)$$

and this follows also from the van der Waals equation [first three terms in (24)] plus the Maxwell rule (see Fig. 2). Again one can say that the van der Waals equation is the "classical" approximation of the corresponding Stark-effect problem. Especially the horizontal portion of the isotherm at $s_1 = 4$ is due to the fact that the deeper minimum switches from the right to the left when F changes sign. Adding to the minimum energy, the zero-point energy does *not* change the qualitative picture; the horizontal portion remains. The finite slope of the isotherm, which we discussed before [see Eq. (25)] is due to a second "quantum mechanical" effect, namely the penetration of the barrier between the two minima if they become of almost equal depth. By using the WKB method one can show that this penetration effect changes Eq. (30) to

$$l_1 = -2\nu_1^{\frac{1}{2}}(s_1 - 4)/[(s_1 - 4)^2 + \epsilon^2]^{\frac{1}{4}} - (s_1 - 4)/12\nu_1, \quad (30a)$$

where

$$\epsilon = (2^{\frac{3}{2}}/\pi)\exp[-(2\nu_1)^{\frac{1}{2}}]$$

and this agrees with Eq. (25).

[5] Compare D. M. Dennison and G. E. Uhlenbeck, Phys. Rev. 41, 313 (1932).

So far we have discussed the isotherm net mainly for $\nu_i \ll -1$ and $\nu_1 \gg 1$. Although it is qualitatively clear how the isotherms change going through the critical region, it is difficult to make precise statements if ν_1 is of order one, and it is clearly *not* possible to answer unambigously the question of what the shape of the coexistence curve is in the critical region. However, it seems reasonable to say that, because of the symmetry of the function $V(z)$, the critical density is not changed if $\gamma\delta$ is finite, but that the critical temperature is lowered because for $\nu_1 > 0$, $V(z)$ will have two minima, which is somehow an indication for the separation of the two phases. One can then define the new critical temperature as that value of ν_1 for which the lowest energy level is just at zero, which is the value of $V(z)$ at the maximum. To determine this value one can use the variant of the WKB method which was developed by Kramers and Ittmann.[6] With this method one finds that the relation between ν_1 and Θ_0, for $l_1 = 0$ and Θ_0 close to zero, is given by

$$\nu_1 = \left(\frac{3\pi}{16}\right)^{\frac{1}{3}} + \frac{|\lambda|}{6}\left(\frac{2}{3\pi}\right)^{\frac{1}{3}}\left[\ln(36\pi) + C + \frac{\pi}{2}\right], \qquad (31)$$

where $\lambda = 2\Theta_0\nu_1^{-\frac{1}{3}}$ and C is the Euler constant. Hence $\nu_1 = (\frac{3}{16}\pi)^{\frac{1}{3}} = 0.702$ if $\Theta_0 = 0$, so that the critical temperature changes by ΔT_c with $\Delta T_c/T_c = 0.702(\gamma\delta)^{\frac{1}{3}}$. For still lower temperatures, that is, for $\nu_1 > 0.702$, the lowest eigenfunction will have two maxima. It is tempting to identify the positions of these maxima with the value of $l_1(\frac{3}{2})^{\frac{1}{3}}$ for the incipient "gas" and "liquid" phase. Doing this one obtains

$$l_1^2 = 4\nu_1^{-\frac{1}{3}}|\lambda|. \qquad (32)$$

Eliminating $|\lambda|$ between (31) and (32) then gives for the extrapolated coexistence curve[7] the equation

$$\nu_1 = \left(\frac{3\pi}{16}\right)^{\frac{1}{3}} + \frac{1}{48}\left[\ln(36\pi) + C + \frac{\pi}{2}\right]l_1^2$$
$$= 0.702 + 0.143l_1^2. \qquad (33)$$

This is still a parabola, but it is appreciably flatter than the van der Waals result $\nu_1 = \frac{1}{4}l_1^2$ (see Fig. 3).

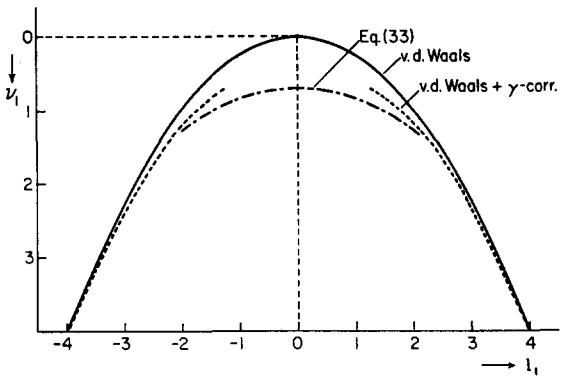


FIG. 3. Shape of the coexistence curve near the critica point; ——— van der Waals; – – – – van der Waals + γ corrections [Eq. (23)]; ·—·— Eq. (33).

IV. THE TWO-POINT DISTRIBUTION FUNCTION IN THE CRITICAL REGION

To determine the long-range behavior of the two-point distribution function in the critical region we will follow the same method as used in II, Sec. V for the one-phase region. However, since in the latter region $\bar{n}_2(x; l)$ decays to $1/l^2$ according to the Ornstein–Zernike exponential $\exp(-B\gamma x/l)$ [II, Eq. (54)], and since in the critical region B/l is of order $(\gamma\delta)^{\frac{1}{3}}$, one should expect that in the critical region the range in x is of order $1/\gamma(\gamma\delta)^{\frac{1}{3}}$. We therefore replace, in the basic formula [II, Eq. (9a)] for the Laplace transform of $\bar{n}_2(x; l)$, the variable σ by $\sigma\gamma(\gamma\delta)^{\frac{1}{3}}$ and we start from

$$l\int_0^\infty dx\, \bar{n}_2(x; l)\exp[-\sigma\gamma(\gamma\delta)^{\frac{1}{3}}x]$$
$$= -1 + \sum_{n=0}^{\infty}\frac{\lambda_0(s)}{\lambda_0(s) - \lambda_n(s + \sigma\gamma(\gamma\delta)^{\frac{1}{3}})}\,b_n^2, \qquad (34)$$

with

$$b_n = \int_{-\infty}^{+\infty} dx\, \psi_0(x; s)\psi_n(x; s + \sigma\gamma(\gamma\delta)^{\frac{1}{3}}).$$

We first compute b_n^2 up to order $(\gamma\delta)^{\frac{1}{3}}$. Introduce the variable z of Eq. (3) and the expansion (11) for the eigenfunctions. Notice that $s + \sigma\gamma(\gamma\delta)^{\frac{1}{3}}$ is obtained from the expansion (7) for s by adding 8σ to s_2, so that the equation (12) for $H^{(0)}$ is *not* affected by the change of s to $s + \sigma\gamma(\gamma\delta)^{\frac{1}{3}}$. Using the orthogonality and normalization relations, one obtains

$$b_n^2 = \delta_{n0} + (\gamma\delta)^{\frac{1}{3}}\Big\{\int_{-\infty}^{+\infty} dz H_0^{(0)}(z, s)$$
$$\times [H_n^{(1)}(z, s + \sigma\gamma(\gamma\delta)^{\frac{1}{3}}) - H_n^{(1)}(z, s)]\Big\}^2$$
$$+ 2(\gamma\delta)^{\frac{1}{3}}\delta_{n0}\int_{-\infty}^{+\infty} dz\{H_0^{(0)}(z, s)$$
$$\times [H_0^{(2)}(z, s + \sigma\gamma(\gamma\delta)^{\frac{1}{3}}) - H_0^{(2)}(z, s)]$$
$$+ H_0^{(1)}(z, s)[H_0^{(1)}(z, s + \sigma\gamma(\gamma\delta)^{\frac{1}{3}}) - H_0^{(1)}(z, s)]\}. \qquad (35)$$

[6] H. A. Kramers and G. P. Ittmann, Z. Physik 58, 217 (1929). See especially Sec. 6.

[7] Note that the equation for the coexistence curve as derived from the van der Waals equation plus the corrections of $O(\gamma)$ (see I, Sec. 4) diverges at the critical point, so that a direct extrapolation is not possible.

The second term can be calculated as follows. From Eq. (13) for $H^{(1)}$ and the fact that $H_n^{(0)}$ and Θ_n are not affected by the change $s \to s + \sigma\gamma(\gamma\delta)^{\frac{1}{2}}$, one shows that

$$\left[\frac{d^2}{dz^2} - \frac{z^4}{48} + \frac{\nu_1 z^2}{4} + \frac{s_1 - 4}{4(6)^{\frac{1}{2}}} z + \Theta_n\right] \times [H_n^{(1)}(z, s + \sigma\gamma(\gamma\delta)^{\frac{1}{2}}) - H_n^{(1)}(z, s)] = [-(\tfrac{2}{3})^{\frac{1}{2}}\sigma z - \Omega_n(\sigma) + \Omega_n(0)]H_n^{(0)}(z, s), \quad (36)$$

where $\Omega_n(\sigma) \equiv \Omega_n(s + \sigma\gamma(\gamma\delta)^{\frac{1}{2}})$. Multiplying (36) with $H_0^{(0)}(z, s)$ and integrating gives for $n \neq 0$

$$\int_{-\infty}^{+\infty} dz H_0^{(0)}(z, s)\{H_n^{(1)}(z, s + \sigma\gamma(\gamma\delta)^{\frac{1}{2}}) - H_n^{(1)}(z, s)\} = -(\tfrac{2}{3})^{\frac{1}{2}} \frac{\sigma}{\Theta_n - \Theta_0} \int_{-\infty}^{+\infty} dz\, z H_0^{(0)}(z, s) H_n^{(0)}(z, s). \quad (37)$$

To calculate next the terms in (34) containing the eigenvalues, it turns out that we need to extend the expansion (10) up to order $(\gamma\delta)^2$. This leads to

$$\lambda_0(s) - \lambda_n(s + \sigma\gamma(\gamma\delta)^{\frac{1}{2}}) = \omega(s_c) \times \left\{3(\Theta_n - \Theta_0 + \sigma)(\gamma\delta)^{4/3} + 3[\Omega_n(\sigma) - \Omega_0(0)](\gamma\delta)^{5/3} + 3\left[\Xi_n(\sigma) - \Xi_0(0) + \frac{11}{4}\nu_1\sigma\right](\gamma\delta)^2 + \cdots\right\},$$

which gives

$$3(\gamma\delta)^{4/3}\sigma \frac{\lambda_0(s)}{\lambda_0(s) - \lambda_n(s + \sigma\gamma(\gamma\delta)^{\frac{1}{2}})} = \frac{\sigma}{\sigma + \Theta_n - \Theta_0} - \frac{(\gamma\delta)^{\frac{1}{2}}\sigma}{(\sigma + \Theta_n - \Theta_0)^2} \times [\Omega_n(\sigma) - \Omega_0(0)] + \frac{(\gamma\delta)^{\frac{1}{2}}\sigma^2}{(\sigma + \Theta_n - \Theta_0)^2} \times \left\{+\frac{11\nu_1}{4} + \frac{9\nu_1}{4}\left(1 + \frac{\Theta_n - \Theta_0}{\sigma}\right) + \frac{[\Omega_n(\sigma) - \Omega(0)]^2}{\sigma[\sigma + \Theta_n - \Theta_0]} - \frac{\Xi_n(\sigma) - \Xi(0)}{\sigma}\right\}. \quad (38)$$

Notice also that from (36) for $n = 0$ one gets, by multiplying with $H_0^{(0)}(z, s)$ and integrating,

$$\Omega_0(\sigma) - \Omega_0(0) = -(\tfrac{2}{3})^{\frac{1}{2}}\sigma \int_{-\infty}^{+\infty} dz\, z H_0^{(0)}(z, s)^2 = \sigma l_1, \quad (39)$$

where we have used the equation of state (15).

Substituting all these partial results in (34) one finds after some simple rearrangements

$$l_c^2\sigma\gamma(\gamma\delta)^{\frac{1}{2}} \int_0^\infty dx \exp[-\sigma\gamma(\gamma\delta)^{\frac{1}{2}}x]\left[\bar{n}_2(x; l) - \frac{1}{l^2}\right] = (\gamma\delta)^{\frac{1}{2}} \frac{2}{3} \sum_{n=1}^{\infty} \frac{\sigma}{\sigma + \Theta_n - \Theta_0} \times \left[\int_{-\infty}^{+\infty} dz\, z H_0^{(0)} H_n^{(0)}\right]^2 + (\gamma\delta)^{\frac{1}{2}}R. \quad (40)$$

Here R is a collection of terms for which we refer to Appendix II, where it will be proved that in fact $R = 0$. Of course Eq. (40) is valid only in the critical region, so that l is always given by Eq. (6), and the eigenvalues Θ_n and eigenfunctions $H_n^{(0)}$ must be determined from Eq. (12). With $R = 0$ one can invert the Laplace transform, and one finds

$$\bar{n}_2(x, l) = \frac{1}{l^2} + \frac{2}{3}\frac{(\gamma\delta)^{\frac{1}{2}}}{l_c^2} \sum_{n=1}^{\infty} \left[\int_{-\infty}^{+\infty} dz\, z H_0^{(0)} H_n^{(0)}\right]^2 \times \exp[-(\Theta_n - \Theta_0)\gamma x(\gamma\delta)^{\frac{1}{2}}]. \quad (41)$$

This is the final result. It shows that in the critical region the decay of $\bar{n}_2(x, l)$ to the uncorrelated value $1/l^2$ is represented by an infinite series of decreasing exponentials which are all of order $(\gamma\delta)^{\frac{1}{2}}$ and have ranges of order $1/\gamma(\gamma\delta)^{\frac{1}{2}}$.

Before starting a detailed discussion of Eq. (41) we conclude with some remarks.

(a) The fluctuation theorem [see II, Eq. (34a)] can be checked as follows. From (41) one gets:

$$\int_0^\infty dx\left[\bar{n}_2(x, l) - \frac{1}{l^2}\right] = \frac{2}{3}\frac{\delta}{l_c^2}(\gamma\delta)^{-\frac{1}{2}} \times \sum_{n=1}^{\infty} \frac{1}{\Theta_n - \Theta_0}\left[\int_{-\infty}^{+\infty} dz\, z H_0^{(0)} H_n^{(0)}\right]^2,$$

and from (21) one sees that to order $(\gamma\delta)^{-\frac{1}{2}}$, the right-hand side is equal to $(-\frac{1}{2}l^3)/(\partial s/\partial l)$ and this is also the leading term of the fluctuation integral.

(b) In the range $x \sim 1/\gamma$, Eq. (41) gives, to lowest order,

$$\bar{n}_2(x, l) = \frac{1}{l^2} + \frac{2}{3l_c^2}(\gamma\delta)^{\frac{1}{2}} \times \sum_{n=0}^{\infty}\left\{\int_{-\infty}^{+\infty} dz[z + (\tfrac{3}{2})^{\frac{1}{2}}l_1]H_0^{(0)}H_n^{(0)}\right\}^2 = \frac{1}{l^2} + \frac{2}{3l_c^2}(\gamma\delta)^{\frac{1}{2}} \int_{-\infty}^{+\infty} dz[z + (\tfrac{3}{2})^{\frac{1}{2}}l_1]^2 H_0^{(0)}(z)^2, \quad (42)$$

where the equation of state (15) has been used to extend the summation from zero to infinity which then allows the application of Parseval's theorem. One can derive (42) also directly from the basic formula [II, Eq. (9a)] replacing σ by $\sigma\gamma$ and expanding to order $(\gamma\delta)^{\frac{1}{2}}$, but we will leave this to the reader. Eq. (42) clearly shows that in the critical region $1/\gamma$ is *not* the ultimate range, since in this

range $\bar{n}_2(x, l)$ still differs from $1/l^2$ by a constant of order $(\gamma\delta)^{\frac{1}{2}}$.

(c) To check the virial theorem [see II, Eq. (40a)]

$$s = \frac{1}{l} + \delta\bar{n}_2(\delta^+, l) - \nu_0\gamma^2 \int_\delta^\infty dx\, xe^{-\gamma x}\bar{n}_2(x; l)$$

in the critical region, one still needs the correction of order $(\gamma\delta)^{\frac{1}{2}}$ in the short-range behavior of $\bar{n}_2(x; l)$. In Appendix III we prove that

$$\bar{n}_2(\delta^+, l) = \frac{1}{l(l-\delta)} + \frac{9}{4l_c^2}(\gamma\delta)^{\frac{1}{2}} \times \int_{-\infty}^{+\infty} dz[z + (\tfrac{3}{2})^{\frac{1}{3}}l_1]^2 H_0^{(0)}(z)^2. \quad (43)$$

Using (42) and the expansions (5), (6), and (7) for ν_0, l, and s, it is then easy to check the virial theorem up to order $(\gamma\delta)^{\frac{1}{2}}$.

V. DISCUSSION OF EQ. (41)

We will show that Eq. (41) goes over into the results obtained in II for the long-range behavior of $\bar{n}_2(x; l)$ in the one- and two-phase region.

Consider first the two-phase region. We found (see II, Sec. VI) that there, in the range $x \sim 1/\gamma$ and up to order γ,

$$\bar{n}_2(x, l) = \frac{\xi_l l_l}{l}\left[\frac{1}{l_l^2} + \frac{\nu_0(l_l-\delta)^4\gamma}{l_l^5 B_l}\exp\left(-\frac{B_l}{l_l}\gamma x\right)\right] + \frac{\xi_g l_g}{l}\left[\frac{1}{l_g^2} + \frac{\nu_0(l_g-\delta)^4\gamma}{l_g^5 B_g}\exp\left(-\frac{B_g}{l_g}\gamma x\right)\right], \quad (44)$$

where ξ_l, ξ_g are the mole fractions of the liquid and vapor phase which have the specific volumes l_l and l_g, so that

$$\xi_l + \xi_g = 1, \qquad \xi_l l_l + \xi_g l_g = l.$$

Using the expansion (5) for ν_0 and the expansion (6) both for l_l and l_g, one finds

$$B_i/l_i = (\gamma\delta)^{\frac{1}{3}}(\tfrac{3}{4}l_{1i}^2 - \nu_1)^{\frac{1}{3}}, \text{ with } i = l \text{ or } g, \quad (45)$$

$$\frac{1}{l}\left(\frac{\xi_l}{l_l} + \frac{\xi_g}{l_g}\right) = \frac{1}{l^2} + \frac{\xi_l\xi_g(l_{1g}-l_{1l})^2}{l_c^2}(\gamma\delta)^{\frac{1}{3}} + O(\gamma\delta).$$

Furthermore, near the critical point, $l_{1l} = -2\nu_1^{\frac{1}{2}}$ and $l_{1g} = +2\nu_1^{\frac{1}{2}}$, since the van der Waals coexistence curve is given by $l_1^2 = 4\nu_1$. Hence $B/l = (\gamma\delta)^{\frac{1}{3}}(2\nu_1)^{\frac{1}{3}}$ both for the liquid and the vapor, so that the two exponentials in (44) become identical. Finally one has

$$l_1 = \xi_l l_{1l} + \xi_g l_{1g} = 2\nu_1^{\frac{1}{2}}(\xi_g - \xi_l),$$

which gives

$$\left.\begin{matrix}\xi_l\\ \xi_g\end{matrix}\right\} = \frac{1}{2}\left[1 \mp \frac{l_1}{2\nu_1^{\frac{1}{2}}}\right].$$

Substituting all this in (44) one obtains

$$\bar{n}_2(x, l) = \frac{1}{l^2} + \frac{2}{3}\frac{(\gamma\delta)^{\frac{1}{3}}}{l_c^2}\left\{6\nu_1 - \tfrac{3}{2}l_1^2 + \frac{1}{(2\nu_1)^{\frac{1}{3}}} \times \exp\left[-(2\nu_1)^{\frac{1}{2}}\gamma(\gamma\delta)^{\frac{1}{3}}x\right]\right\}. \quad (46)$$

This follows also from Eq. (41) when $\nu_1 \gg 1$. Let us verify it only in the symmetric case ($l_1 = 0$). We saw in Sec. III that then the energy levels Θ_n are a series of narrow doublets corresponding to the symmetric and antisymmetric combinations of the oscillator eigenfunctions around the two minima. In the series (41) only the odd values of n contribute, and in first approximation only $n = 1$ and $n = 3$ need to be considered. One finds

$$\Theta_1 - \Theta_0 \cong 0, \qquad \Theta_3 - \Theta_0 \cong (2\nu_1)^{\frac{1}{2}};$$

$$\int dz\, zH_0^{(0)}H_1^{(0)} \cong (6\nu_1)^{\frac{1}{3}}, \qquad \int dz\, zH_0^{(0)}H_3^{(0)} \cong (2\nu_1)^{-\frac{1}{2}};$$

so that Eq. (41) becomes equal to (46) with $l_1 = 0$.

Consider next the transition to the one-phase region, which is of greater interest since more detailed results are available [see II, Sec. V, Eqs. (54) and (55)]. For $\nu_1 \ll -1$, or better, for $(3l_1^2 - 4\nu_1) \gg 1$, we found in Sec. III that in first approximation [see Eq. (28)]

$$\Theta_n - \Theta_0 = \tfrac{1}{2}n(3l_1^2 - 4\nu_1)^{\frac{1}{2}},$$

and that the eigenfunctions are just the Weber functions $D_n(y)$, where y is given by Eq. (29). It is then easy to verify that the first term of the sum in Eq. (41) becomes

$$\frac{4}{3l_c^2}\frac{(\gamma\delta)^{\frac{1}{3}}}{(3l_1^2 - 4\nu_1)^{\frac{1}{3}}}\exp\left[-\tfrac{1}{2}(3l_1^2 - 4\nu_1)^{\frac{1}{2}}\gamma x(\gamma\delta)^{\frac{1}{3}}\right],$$

and that this is also precisely what the Ornstein–Zernike term [see II, Eq. (54)]

$$[\gamma\nu_0(l-\delta)^4/l^5 B]\exp[-(B/l)\gamma x]$$

becomes in the critical region up to order $(\gamma\delta)^{\frac{1}{3}}$. One only has to use (45) for B/l, and replace ν_0 and l by their critical values.

In the next approximation, using the results derived in Appendix I, one finds

$$\Theta_n - \Theta_0 = \tfrac{1}{2}n(3l_1^2 - 4\nu_1)^{\frac{1}{2}} + \frac{1}{2(3l_1^2 - 4\nu_1)}\left[n^2 + n - \frac{2l_1^2}{3l_1^2 - 4\nu_1}(5n^2 - n)\right],$$

and up to order ϵ^2,

$$\int_{-\infty}^{+\infty} dz\, zH_0^{(0)}H_n^{(0)} = (\tfrac{3}{4}l_1^2 - 4\nu_1)^{-\frac{1}{2}} \times \left\{\delta_{n1}\left[1 + \frac{\epsilon^2}{3}\frac{12\nu_1 - 5l_1^2}{3l_1^2 - 4\nu_1}\right] - 2(\tfrac{2}{3})^{\frac{1}{3}}\frac{\epsilon l_1}{(3l_1^2 - 4\nu_1)}\delta_{n2}\right\},$$

omitting terms which do not contribute after squaring. In this order, therefore, the first and second term of the sum in Eq. (41) contribute. One obtains

$$\bar{n}_2(x;\, l) - \frac{1}{l^2} = \frac{(\gamma\delta)^{\frac{1}{3}}}{l_c^2}\left[\frac{4}{3}\frac{1}{(3l_1^2 - 4\nu_1)^{1/2}} + \frac{4}{3}\frac{l_1^2 + 4\nu_1}{(3l_1^2 - 4\nu_1)^{5/2}}\gamma x(\gamma\delta)^{1/3} - \frac{8}{9}\frac{5l_1^2 - 12\nu_1}{(3l_1^2 - 4\nu_1)^3}\right] \times \exp\left[-\tfrac{1}{2}(3l_1^2 - 4\nu_1)^{1/2}\gamma x(\gamma\delta)^{1/3}\right] + \frac{(\gamma\delta)^{2/3}}{l_c^2}\frac{32}{9}\frac{l_1^2}{(3l_1^2 - 4\nu_1)^3} \times \exp\left[-(3l_1^2 - 4\nu_1)^{1/2}\gamma x(\gamma\delta)^{1/3}\right] + \cdots, \qquad (47)$$

and one can verify that the last three terms in (47) correspond precisely to the three corrections of order γ^2 to the Ornstein–Zernike result in the one-phase region [see II, Eq. (55)] if these corrections are computed in the critical region up to order $(\gamma\delta)^{\frac{1}{3}}$. This confirms the general picture mentioned in the introduction, and it shows again that the Ornstein–Zernike exponential is the leading term only *outside* the critical region.

Since in the experimental investigation of the critical opalescence one observes the Fourier transform of the two-point distribution function, it is of interest to discuss also for our model the function

$$\tilde{g}(k) = \int_{-\infty}^{+\infty} dx\, e^{ikx} g(x) = l\int_{-\infty}^{+\infty} dx\, e^{ikx}\left[\bar{n}_2(x,\, l) - \frac{1}{l^2}\right].$$

For the Ornstein–Zernike exponential (in any number of dimensions), $\tilde{g}(k) \sim 1/(k^2 + b^2)$, where b^2 is proportional to the compressibility, and it is therefore customary to plot $1/\tilde{g}(k) \equiv f(k)$ vs k^2, since this should be a straight line intersecting the $f(k)$ axis at a point which goes to zero at the critical point if the Ornstein–Zernike theory is correct. Since Eq. (41) is of the form

$$g(x) = \sum_{n=0}^{\infty} a_n e^{-b_n x},$$

where the a_n and b_n are positive, we obtain for our model, in the critical region,

$$\frac{1}{f(k)} \equiv \tilde{g}(k) = \sum_0^{\infty} \frac{2a_n b_n}{b_n^2 + k^2}.$$

For k^2 large compared to the b_n^2, $f(k)$ is almost linear in k^2. In fact,

$$f(k) = \frac{k^2}{2\sum a_n b_n} + \frac{\sum a_n b_n^3}{2(\sum a_n b_n)^2} + O\left(\frac{1}{k^2}\right). \qquad (48)$$

To study the curvature for smaller k^2, note that one can write

$$\frac{f''}{2f^3} = \left[\sum_n \frac{a_n b_n}{(b_n^2 + k^2)^2}\right]^2 - \left(\sum_n \frac{a_n b_n}{b_n^2 + k^2}\right)\left[\sum_n \frac{a_n b_n}{(b_n^2 + k^2)^3}\right],$$

where the primes denote differentiations after k^2. Hence, by Schwartz's inequality, one concludes that $f'' \leq 0$, so that the curve is convex towards the k^2 axis.

For the discussion of the temperature dependence of the Ornstein–Zernike plot, we will restrict ourselves to the case that $l_1 = 0$ (critical density) and that $\nu_1 \ll -1$. One then can use Eq. (47); since the last term vanishes, one can write $g(x)$ as one exponential by putting the second term back in the exponent, and one obtains

$$g(x) = a_1 e^{-b_1 x},$$

with

$$a_1 = \frac{2}{3l_c}\frac{(\gamma\delta)^{\frac{1}{3}}}{(-\nu_1)^{\frac{1}{2}}}\left[1 - \frac{1}{4(-\nu_1)^{\frac{3}{2}}} + \cdots\right],$$

$$b_1 = (\gamma\delta)^{4/3}\frac{(-\nu_1)^{\frac{1}{2}}}{\delta}\left[1 + \frac{1}{4(-\nu_1)^{\frac{3}{2}}} + \cdots\right].$$

Since the slope of the Ornstein–Zernike plot is $\frac{1}{2}a_1 b_1$, one sees that the temperature dependence drops out up to order $(-\nu_1)^{\frac{1}{2}}$, so that the straight parts of the plot are very nearly parallel. From (48) one further sees that the intersection with the $f(k)$ axis is $b_1/2a_1$, or

$$f(0) = \tfrac{9}{4}(\gamma\delta)^{\frac{1}{3}}(-\nu_1)[1 + 1/2(-\nu_1)^{\frac{3}{2}} + \cdots].$$

Since $-(\gamma\delta)^{\frac{1}{3}}\nu_1 = (T - T_c)/T_c$ to this order in $(\gamma\delta)$ one sees that, in first approximation, $f(0) \sim (T - T_c)$ which is the Ornstein–Zernike result. However, closer to the critical point, $f(0)$ bends upwards and one can show [using the exact Eq. (41)] that $f(0)$ is finite for $T = T_c$. The deviations from the Ornstein–Zernike theory which follow from our model, look therefore about as shown in Fig. 4.

Finally we will discuss briefly the so-called *specific heat anomaly*.[8] It is well known that according to the van der Waals theory the specific heat at constant volume is constant ($= \frac{3}{2}k$) in the one-phase region and increases suddenly if one crosses at a fixed density into the two-phase region. Especially at the critical density one finds

[8] We are indebted to Dr. J. V. Sengers for pointing out to us the interest of discussing this anomaly for our model.

$$\Delta c_v = \begin{cases} 0 & \text{for } T > T_c, \\ \frac{9}{2}k[1 - \frac{28}{25}(T_c - T)/T_c + \cdots] & \text{for } T < T_c. \end{cases} \tag{49}$$

Since the experimental results are quite different (this is the anomaly), it is of interest to see what the implications are of Eq. (41) for the specific heat. We start from the equation

$$\epsilon = \frac{kT}{2} - \frac{\alpha_0}{l_c} - \frac{2}{3}\frac{\alpha_0}{l_c}(\gamma\delta)^{\frac{1}{2}} \times \int_{-\infty}^{+\infty} dz[z + l_1(\tfrac{3}{2})^{\frac{1}{2}}]^2 H_0^{(0)}(z)^2, \tag{50}$$

which follows from the general result [II, Eq. (41)] by introducing the expression (42) for $n_2(x; l)$. Since in the critical region

$$\frac{\partial}{\partial T} = -\frac{\alpha_0}{kT_c^2}\frac{1}{\nu_{0c}(\gamma\delta)^{\frac{1}{2}}}\frac{\partial}{\partial \nu_1},$$

one obtains

$$\Delta c_v \equiv \left(\frac{\partial \epsilon}{\partial T}\right)_l - \frac{k}{2} = \frac{3}{4}k\frac{\partial}{\partial \nu_1}\int_{-\infty}^{+\infty} dz\, z^2 H_0^{(0)2}(z, \nu_1, l_1) = -3k\left(\frac{\partial^2 \Theta_0}{\partial \nu_1^2}\right)_{l_1} + \frac{3k}{8} l_1 \left(\frac{\partial^2 s_1}{\partial \nu_1^2}\right)_{l_1}, \tag{51}$$

by differentiating Eq. (12) for $H_0^{(0)}(z)$ after ν_1 at constant l_1, and by using the equation of state (15). This shows that our model gives corrections to the specific heat which are of order k in the critical region, so that in this sense our model gives an anomalous behavior of c_v. To discuss Eq. (51) further, consider first the case $l_1 = 0$. It is easy to show that then

$$\Delta c_v = -3k\left(\frac{\partial^2 \Theta_0}{\partial \nu_1^2}\right)_{l_1} = \frac{3k}{8}\sum_{n=1}^{\infty}\frac{1}{\Theta_n - \Theta_0}\left[\int_{-\infty}^{+\infty} dz\, z^2 H_0^{(0)} H_n^{(0)}\right]^2,$$

and by using the results of Sec. III one finds for $\nu_1 \ll -1$

$$\Delta c_v = \tfrac{3}{8}k[1/(-\nu_1)^{\frac{1}{2}}] + \cdots,$$

and for $\nu_1 \gg 1$[9]

$$\Delta c_v = \tfrac{9}{2}k - \tfrac{3}{4}k[1/(2\nu_1)^{\frac{1}{2}}] + \cdots.$$

At the critical density our model therefore smooths out the discontinuity (49) in c_v which follows from the van der Waals equation. Next, consider the dependence of Δc_v on l_1 at fixed ν_1. From (51) follows

$$[\partial(\Delta c_v)/\partial l_1]_{\nu_1} = \tfrac{3}{8}k(\partial^2 s_1/\partial \nu_1^2)_{l_1}, \tag{52}$$

using

$$8(\partial \Theta_0/\partial l_1)_{\nu_1} = l_1(\partial s_1/\partial l_1)_{\nu_1},$$

which is a consequence of (19). Eq. (52) is just a transcription of the thermodynamic formula

$$(\partial c_v/\partial v)_T = T(\partial^2 p/\partial T^2)_v,$$

and it shows that, at the critical density, Δc_v has always an extremum, because $(s_1 - 4)$ is an odd function of l_1 so that $\partial^2 s_1/\partial \nu_1^2$ is zero for $l_1 = 0$. Furthermore, for small l_1,

$$\Delta c_v = \Delta c_v(l_1 = 0) + l_1^2 \tfrac{3}{16}k(\partial^3 s_1/\partial \nu_1^2\, \partial l_1)_{l_1=0} + \cdots,$$

which gives for $\nu_1 \ll -1$

$$\Delta c_v = \Delta c_v(l_1 = 0) - \tfrac{27}{64}k[l_1^2/(-\nu_1)^{5/2}] + \cdots,$$

and for $\nu_1 \gg 1$

$$\Delta c_v = \Delta c_v(l_1 = 0) - (27k/8\pi)(2\nu_1)^{\frac{1}{2}} \times \exp[-(2\nu_1)^{\frac{1}{2}}]l_1^2 + \cdots,$$

so that the extremum is probably always a maximum. The maximum becomes sharper when one approaches the critical point from the one-phase region, and then becomes rapidly very flat in the two-phase region. Presumably, the width of the maximum is smallest for ν_1 of order one.

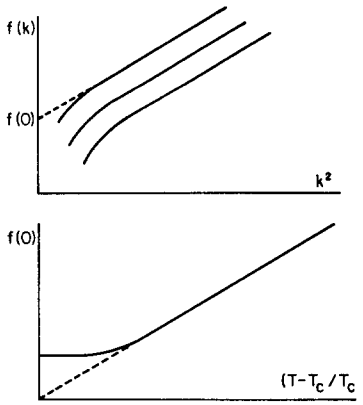


FIG. 4. Ornstein-Zernike plot according to Eq. (47).

VI. GENERALIZATION OF EQ. (41)

In order to show that qualitatively the behavior of the two-point distribution function in the critical region does not depend sensitively on the form of the attractive potential (provided it has a long range), we will generalize the discussion of Sec. IV to the case where the attractive potential consists of a sum of m exponentials. Since the formulas are very long, we will only present an outline of the

[9] To derive this result it is simpler not to use (51), but to go back to the general result [See II, Eq. (41)] and then use Eq. (46) for $\tilde{n}_2(x; l)$.

calculations. As in II, Sec. VII, we write the attractive potential in the form

$$\varphi_{attr}/kT = -\gamma \sum_{i=1}^{m} \nu_i \exp(-\gamma\sigma_i x),$$

where $\nu_i = \alpha_i/kT$, and we start to tame the Kac integral equation in the same way as in II, except that we replace η by $\eta_c \equiv \eta(s_c)$. One expands the van der Waals $\nu_0 = \sum_i (\nu_i/\sigma_i)$ [see II, Eq. (79)] and s as in (5) and (7), and one then introduces new variables y_m by an orthogonal transformation which diagonalizes, just as in II, Eq. (75), the matrix

$$M_{ij} = \sigma_i^2\delta_{ij} - \sigma_i\sigma_j(c_ic_j)^{\frac{1}{2}},$$

except that now,

$$c_i = \frac{2\nu_{ic}(l_c - \delta)^2}{\sigma_i l_c^3} = \frac{8\nu_{ic}}{27\sigma_i\delta} = \frac{\nu_{ic}}{\sigma_i\nu_{0c}} = \frac{\nu_i}{\sigma_i\nu_0},$$

where $\nu_{ic} = \alpha_i/kT_c$. It follows that $\sum c_i = 1$, in agreement with II, Eq. (80), since $B = 0$, and from the discussion given in II, this implies that the smallest eigenvalue $A_m = 0$, so that

$$\sum_{i,j} M_{ij}z_iz_j = \sum_{k=1}^{m-1} A_k y_k^2.$$

The variable y_m plays a special role; *only* in this variable one must make the second, or critical-region "taming",

$$y_m = z/(\gamma\delta)^{\frac{1}{2}},$$

and one then gets a consistent successive-approximation scheme with the parameter $(\gamma\delta)^{1/3}$. Up to order $(\gamma\delta)^{4/3}$ one obtains[10]

$$\gamma\delta \sum_{k=1}^{m-1}\left(\frac{\partial^2 h}{\partial y_k^2} - \frac{A_k}{4}y_k^2 h\right) + (\gamma\delta)^{4/3}\left[\frac{\partial^2 h}{\partial z^2} - \frac{(\mu z)^4}{48}h + \frac{\nu_1(\mu z)^2}{4}h + \frac{\mu z}{6^{\frac{1}{3}}}\left(\frac{s_1}{4} - \sum_{i=1}^{m}\sigma_i\right)h\right] + (\gamma\delta)^{4/3}\frac{\mu z h}{2(6)^{\frac{1}{3}}}\sum_{k=1}^{m-1} A_k y_k^2 + \Lambda h = 0, \quad (53)$$

which replaces Eq. (9) up to this order. In (53), h is a function of $y_1, y_2, \cdots, y_{m-1}, z$, and the new eigenvalue Λ is defined by

$$3\Lambda = 1 + \frac{9\nu_1}{4}(\gamma\delta)^{2/3} + \tfrac{3}{2}(\gamma\delta)\left(\sum_{i=1}^{m}\sigma_i - \frac{s_1}{4}\right) + (\gamma\delta)^{4/3}\left(-\frac{3}{8}s_2 + \frac{81}{32}\nu_1^2\right) - \frac{\lambda}{\omega(s_c)}, \quad (54)$$

which replaces Eq. (10). Finally the quantity μ is defined by

$$\sum_{i=1}^{m}\sigma_i^2 a_{im}^2 = \left(\sum_{i=1}^{m}\sigma_i c_i^{\frac{1}{2}} a_{im}\right)^2 \equiv \mu^2,$$

where a_{ik} is the orthogonal matrix, which diagonalizes M_{ij}. By introducing instead of the y_k the new variables w_k by

$$y_k = w_k\{1 + [\mu z/2(6)^{\frac{1}{3}}](\gamma\delta)^{\frac{1}{3}}\},$$

Eq. (53) can be written (always up to order $(\gamma\delta)^{4/3}$) in the form

$$\gamma\delta\left[1 - \frac{\mu z}{6^{\frac{1}{3}}}(\gamma\delta)^{\frac{1}{3}}\right]\sum_{k=1}^{m-1}\left[\frac{\partial^2 h}{\partial w_k^2} - \frac{A_k}{4}w_k^2 h\right] + (\gamma\delta)^{4/3}\left[\frac{\partial^2 h}{\partial z^2} - \frac{(\mu z)^4}{48}h + \frac{\nu_1(\mu z)^2}{4}h + \frac{\mu z}{6^{\frac{1}{3}}}\left(\frac{s_1}{4} - \sum_i \sigma_i\right)h\right] + \Lambda h = 0, \quad (55)$$

and hence it becomes separable. It is easy to verify that (55) is fulfilled by putting

$$\Lambda = \gamma\delta\sum_{k=1}^{m-1} A_k^{\frac{1}{2}}(n_k + \tfrac{1}{2}) + \Theta(\gamma\delta)^{4/3},$$
$$h(w_1 \cdots w_{m-1}, z) = \prod_{k=1}^{m-1} N_{n_k} D_{n_k}(A_k^{\frac{1}{2}}w_k)H^{(0)}(z), \quad (56)$$

if $H^{(0)}(z)$ and Θ satisfy the equation

$$\left\{\frac{d^2}{dz^2} - \frac{(\mu z)^4}{48} + \frac{\nu_1(\mu z)^2}{4} + \frac{\mu z}{6^{\frac{1}{3}}}\left[\frac{s_1}{4} - \sum_{i=1}^{m}\sigma_i + \sum_{k=1}^{m-1} A_k^{\frac{1}{2}}(n_k + \tfrac{1}{2})\right] + \Theta\right\}H^{(0)} = 0, \quad (57)$$

which is of the same form as (12). In fact by introducing the new variables

$$\hat{z} = z\mu^{2/3}, \qquad \hat{\nu}_1 = \nu_1\mu^{-2/3}, \qquad \hat{\Theta} = \Theta\mu^{-4/3},$$
$$\hat{s}_1 = 4 + \frac{1}{\mu}\left(s_1 - 4\sum_{i=1}^{m}\sigma_i + 2\sum_{k=1}^{m-1}A_k^{\frac{1}{2}}\right), \quad (58)$$

and by putting $H^{(0)}(z) \equiv \mu^{\frac{1}{2}}\hat{H}(\hat{z})$, Eq. (57) becomes

$$\left[\frac{d^2}{d\hat{z}^2} - \frac{\hat{z}^4}{48} + \frac{\hat{\nu}_1\hat{z}^2}{4} + \frac{\hat{z}}{6^{\frac{1}{3}}}\left(\frac{\hat{s}_1}{4} - 1 + \frac{1}{\mu}\sum_{k=1}^{m-1}A_k^{\frac{1}{2}}n_k\right) + \hat{\Theta}_{\mathbf{n}}\right]\hat{H}_{\mathbf{n}}(\hat{z}) = 0, \quad (59)$$

which is quite similar to (12). The only difference is that, because of the extra term containing the n_k, the eigenvalues $\hat{\Theta}_{\mathbf{n}}$ and eigenfunctions $\hat{H}_{\mathbf{n}}$ will depend on the whole set of "quantum" numbers $\mathbf{n} \equiv n_1, n_2 \cdots n_m$. Note still, that with the definitions (58) and with

[10] In this equation, ν_1 should not be confused with $\nu_1 = \alpha_1/kT$. From now on, only the critical quantities ν_{ic} will appear and ν_1 will always mean the number defined by Eq. (5).

$$\hat{\gamma} = \mu\gamma, \tag{60}$$

Eq. (5) for the critical temperature region can be written as

$$\nu_0 = \nu_{0c}[1 + \hat{\nu}_1(\hat{\gamma}\delta)^{\frac{1}{3}}],$$

while Eq. (7) for the pressure region becomes

$$s = s_c\Big[1 - 3\hat{\nu}_1(\hat{\gamma}\delta)^{\frac{1}{3}} + \Big(\hat{s}_1 - 4 + \frac{4}{\mu}\sum_{i=1}^{m}\sigma_i - \frac{2}{\mu}\sum_{k=1}^{m-1}A_k^{\frac{1}{2}}\Big)(\hat{\gamma}\delta) + \hat{s}_2(\hat{\gamma}\delta)^{4/3} + \cdots\Big],$$

with $\hat{s}_2 = s_2\mu^{-4/3}$. Finally, Eqs. (54) and (56) for the eigenvalue λ_n can be combined to

$$\frac{\lambda_n(s)}{\omega(s_c)} = 1 + \frac{9}{4}\hat{\nu}_1(\hat{\gamma}\delta)^{\frac{1}{3}} + \tfrac{3}{2}\hat{\gamma}\delta \times\Big[1 - \frac{\hat{s}_1}{4} - \frac{1}{\mu}\sum_{k=1}^{m-1}A_k^{\frac{1}{2}}(2n_k + \tfrac{1}{2})\Big] + (\hat{\gamma}\delta)^{4/3}\Big(-3\hat{\Theta}_n - \frac{3}{8}\hat{s}_2 + \frac{81}{32}\hat{\nu}_1^2\Big) + \cdots$$

Clearly it follows from (59) that the *maximum* eigenvalue and corresponding eigenfunction, for which the whole set $\mathbf{n} = 0$, obey in the capped variables exactly the same equation as before. Therefore the equation of state is still determined by

$$\int_{-\infty}^{+\infty} d\hat{z}\,\hat{z}\hat{H}_{0\ldots0}^2(\hat{z}) = -(\tfrac{3}{2})^{\frac{1}{3}}\hat{l}_1, \tag{61}$$

where $\hat{l}_1$ is defined by:

$$l = l_c[1 + \hat{l}_1(\hat{\gamma}\delta)^{\frac{1}{3}}],$$

similar to Eq. (6).

Turning now to the behavior of the two-point distribution function in the range of order $1/\hat{\gamma}(\hat{\gamma}\delta)^{\frac{1}{3}}$ up to order $(\hat{\gamma}\delta)^{\frac{1}{3}}$, it is not difficult to prove that also in this case the calculation goes in the same way as before. The starting point is the obvious generalization of Eq. (34). One then shows that *only* the terms with $n_1 = n_2 = \cdots = n_{m-1} = 0$ contribute to order $(\hat{\gamma}\delta)^{\frac{1}{3}}$, and since in this case the equation for $\hat{H}_{0,0\ldots0}$, $n_m(\hat{z})$ is the same as Eq. (12) for $H_n(z)$, one obtains, in complete analogy with Eq. (41), that, in the general case,

$$\bar{n}_2(x; l) = \frac{1}{l^2} + \frac{2}{3}\frac{(\hat{\gamma}\delta)^{\frac{1}{3}}}{l_c^2}\sum_{n=1}^{\infty}\Big[\int_{-\infty}^{+\infty} d\hat{z}\,\hat{z}\hat{H}_0\hat{H}_n\Big]^2 \times \exp\,[-(\hat{\Theta}_n - \hat{\Theta}_0)x\hat{\gamma}(\hat{\gamma}\delta)^{\frac{1}{3}}], \tag{62}$$

where $\hat{H}_n$ is the *same* function of $\hat{z}$ as $H_n^{(0)}$ is of z in (41), and $\hat{\Theta}_n$ is equal to the Θ_n occurring in (41).[11] The only essential change is therefore the replacement of γ by $\hat{\gamma} = \mu\gamma$. Now we saw in II [see the equation following (85)] that, at the critical point,

$$a_{im} = (c_i^{\frac{1}{2}}/\sigma_i)/[\sum_i (c_i/\sigma_i^2)]^{\frac{1}{2}}.$$

Hence,

$$\mu^2 = \sum_{i=1}^{m}\sigma_i^2 a_{im}^2 = \frac{\sum_i c_i}{\sum_i (c_i/\sigma_i^2)} = \frac{\sum_i (\alpha_i/\sigma_i)}{\sum_i (\alpha_i/\sigma_i^3)}$$

since $c_i = (\alpha_i/\sigma_i)(1/\nu_0 kT)$. Therefore we can write

$$\frac{1}{\hat{\gamma}^2} = \frac{1}{\mu^2\gamma^2} = \frac{\frac{1}{2}\int_{-\infty}^{+\infty} dx\,x^2\varphi_{\text{attr}}(|x|)}{\int_{-\infty}^{+\infty} dx\,\varphi_{\text{attr}}(|x|)}, \tag{63}$$

so that $1/\hat{\gamma}$ is just the range R of the Ornstein–Zernike theory [see II, Eq. (72)]. The argument given in Sec. V, showing that the first term of the series (41) becomes the Ornstein–Zernike exponential in the one-phase region, can therefore be repeated in exactly the same way in the general case. We have not investigated in detail what happens in the higher approximations, but it seems very likely that qualitatively at least everything will remain the same.

VII. CONCLUDING REMARKS

It is well known that the Fourier transform of the Ornstein–Zernike exponential [the function $\tilde{g}(k)$ of Sec. V] is independent of the number of dimensions, and one may hope therefore that the deviations from the Ornstein–Zernike theory which follow from our model with regard to $\tilde{g}(k)$ will at least qualitatively be the same in three dimensions. This is confirmed by the results of P. C. Hemmer, who has succeeded in generalizing to three dimensions the discussion of the long-range behavior of the correlation function reported in II.[12] It is therefore of interest to inquire whether there are any experimental indications of the deviations from the Ornstein–Zernike theory discussed in Sec. V [see especially Fig. 4].[13] Unfortunately, at present the experiments are still not decisive. The recent work

[11] These Θ_n should not be confused with the Θ occurring in Eq. (57), which is related to $\hat{\Theta}$ by Eq. (58).

[12] This work will be reported in Part IV of this series.

[13] The question of the validity of the Ornstein–Zernike theory close to the critical point has been discussed especially by M. S. Green [J. Chem. Phys. **33**, 1403 (1960)]. Although we do not agree with his theoretical arguments, we are very much indebted to Dr. Green for telling us about his work, and about the experimental results indicating deviations from the O–Z theory. Compare also the review by O. K. Rice in *Thermodynamics and Physics of Matter* (Princeton University Press, Princeton, New Jersey, 1955), Sec. E, where one finds a very complete bibliography.

by Thomas and Schmidt[14] on the critical opalescence in Argon does not show any deviation from the Ornstein–Zernike theory. On the other hand, deviations have been reported for various binary mixtures and these appear to point in the same direction as those shown in Figure 4.[15] It seems to us of great interest to see whether more refined experiments will show deviations from the Ornstein–Zernike theory, since such deviations *near* the critical point are perhaps the clearest indication of the existence of a critical *region*.

With regard to the specific-heat anomaly, the experiments[16] clearly show that c_v is *not* a constant in the one-phase region, and that it has a pronounced maximum as function of the density at constant temperature near the critical point. The maximum occurs close to the critical density, and this aspect of the anomaly agrees therefore qualitatively with the results for our model. However, the temperature dependence is quite different. Instead of smoothing out of the van der Waals discontinuity predicted by our model, recent experiments for Argon[17] show that at the critical density, c_v has as function of the temperature an asymmetric logarithmic singularity at the critical temperature. It seems therefore that in three dimensions, one has both a critical region *and* a critical point! Whether this is an universal phenomenon, more or less independent of the intermolecular forces, remains to be seen.

APPENDIX I. HIGHER APPROXIMATIONS OF THE EIGENFUNCTIONS NEAR THE ONE-PHASE REGION

For a systematic application of the perturbation theory to the basic equation (12) when $\nu_1 \ll -1$, it is best to replace z by the variable y of Eq. (29). Eq. (12) then becomes

$$\left[\frac{d^2}{dy^2} - (3l_1^2 - 4\nu_1)^{-3/2}\frac{y^4}{6} + (3l_1^2 - 4\nu_1)^{-5/4} \times \frac{l_1 y^3}{3^{\frac{1}{2}}} - \frac{y^2}{4} + \alpha y + \vartheta\right]K(y) = 0, \tag{A1}$$

where

$$K(y) = (\tfrac{3}{4}l_1^2 - \nu_1)^{-\frac{1}{4}}H^{(0)}(z),$$

$$\alpha = \frac{2}{3^{\frac{1}{2}}}(3l_1^2 - 4\nu_1)^{-\frac{1}{2}}\left(\frac{s_1}{4} - 1 + \tfrac{3}{8}l_1^3 - \tfrac{3}{2}\nu_1 l_1\right),$$
$$\vartheta = (3l_1^2 - 4\nu_1)^{-\frac{1}{2}} \times \left(2\Theta + l_1 - \frac{3}{32}l_1^4 + \frac{3}{4}l_1^2\nu_1 - \frac{s_1}{4}l_1\right). \tag{A2}$$

For the small perturbation parameter, choose

$$\epsilon = (3l_1^2 - 4\nu_1)^{-\frac{1}{4}}$$

Note that $l_1/(3l_1^2 - 4\nu_1)^{\frac{1}{2}}$ is always of order one. Write (A1) in the form

$$\left[\frac{d^2}{dy^2} - \frac{\epsilon^2 y^4}{6} - \left(1 - \frac{4\nu_1}{3l_1^2}\right)^{-\frac{1}{2}}\frac{\epsilon y^3}{3} - \frac{y^2}{4} + \alpha(\epsilon)y + \vartheta(\epsilon)\right]K(y) = 0,$$

and expand

$$K(y) = K^{(0)}(y) + \epsilon K^{(1)}(y) + \epsilon^2 K^{(2)}(y) + \cdots,$$
$$\alpha(\epsilon) = \alpha_0 + \epsilon\alpha_1 + \epsilon^2\alpha_2 + \cdots,$$
$$\vartheta(\epsilon) = \vartheta^{(0)} + \epsilon\vartheta^{(1)} + \epsilon^2\vartheta^{(2)} + \cdots.$$

In zeroth order, one gets

$$[d^2/dy^2 - \tfrac{1}{4}y^2 + \alpha_0 y + \vartheta^{(0)}]K^{(0)} = 0. \tag{A3}$$

Since the equation of state (15) implies that for the ground state

$$\int_{-\infty}^{+\infty} dy\, yK_0(y)^2 = 0, \tag{A4}$$

it follows that α_0 must be zero, and hence

$$K_n^{(0)}(y) = N_n D_n(y), \qquad \vartheta_n^{(0)} = n + \tfrac{1}{2}. \tag{A5}$$

In first order one then gets

$$\left[\frac{d^2}{dy^2} - \frac{y^2}{4} + n + \frac{1}{2}\right]K_n^{(1)}(y) = \left[-\left(1 - \frac{4\nu_1}{3l_1^2}\right)^{-\frac{1}{2}}\frac{y^3}{3} - \alpha_1 y - \vartheta_n^{(1)}\right]K_n^{(0)}. \tag{A6}$$

Since $K_n^{(0)}(y)^2$ is symmetric in y, multiplication of (A6) with $K_n^{(0)}$ and integration over y gives

$$\vartheta_n^{(1)} = 0. \tag{A7}$$

For $n = 0$, one gets from (A6)

$$K_0^{(1)}(y) = \frac{1}{9}\left(1 - \frac{4\nu_1}{3l_1^2}\right)^{-\frac{1}{2}}D_3(y) + \left[\left(1 - \frac{4\nu_1}{3l_1^2}\right)^{-\frac{1}{2}} + \alpha_1\right]D_1(y).$$

[14] J. E. Thomas and P. W. Schmidt, J. Chem. Phys. (to be published).

[15] See especially D. McIntyre, A. Wims, and M. S. Green, J. Chem. Phys. 37, 3019 (1962).

[16] The most complete data are for Argon, but other substances show the same behavior. See J. M. H. Levelt, Dissertation, University of Amsterdam, 1958.

[17] M. I. Bagatskii, A. V. Voronel, and V. G. Gusak, Zh. Eksperim. i Teor. Fiz. 43, 728 (1962). We are indebted to Dr. M. Fisher for pointing out this reference to us.

Since it follows from (A4) that

$$\int_{-\infty}^{+\infty} dy\, yK_0^{(0)}K_0^{(1)} = 0,$$

one concludes that

$$\alpha_1 = -[1 - \tfrac{4}{3}(\nu_1/l_1^2)]^{-\frac{1}{2}}. \tag{A8}$$

From (A2) together with (A5), (A7) and (A8) one obtains the results (24) and (28) for the equation of state and the energy Θ_n. For the first-order correction to the eigenfunction, one then gets from (A6)

$$K_n^{(1)}(y) = N_n\left(1 - \frac{4\nu_1}{3l_1^2}\right)^{-\frac{1}{2}}\left[\frac{1}{9}D_{n+3} + nD_{n+1} - n(n-1)D_{n-1} - \frac{n(n-1)(n-2)}{9}D_{n-3}\right].$$

In second order one obtains

$$\left[\frac{d^2}{dy^2} - \frac{y^2}{4} + n + \frac{1}{2}\right]K_n^{(2)}(y) = -\left(1 - \frac{4\nu_1}{3l_1^2}\right)^{-\frac{1}{2}} \times \left(y - \frac{y^3}{3}\right)K_n^{(1)}(y) + \left(\frac{y^2}{6} - \alpha_2 y - \vartheta_n^{(2)}\right)K_n^{(0)}(y), \tag{A9}$$

from which one finds

$$\vartheta_n^{(2)} = n^2 + n + \tfrac{1}{2} - [3l_1^2/(3l_1^2 - 4\nu_1)] \times \left(\frac{10}{3}n^2 - \frac{2}{3}n + \frac{2}{9}\right), \tag{A10}$$

and from (A4) one concludes that

$$\int_{-\infty}^{+\infty} dy\, y(K_0^{(1)2} + 2K_0^{(0)}K_0^{(2)}) = 0,$$

which leads to $\alpha_2 = 0$. One then can determine $K_n^{(0)}(y)$ in the usual way; and so on!

APPENDIX II. COMPLETION OF THE PROOF OF EQ. (41)

The quantity R left undefined in Eq. (40) is given by

$$R = \frac{2}{3}\sum_{n=1}^{\infty}\frac{\sigma(\sigma - \Theta_n + \Theta_0)}{(\Theta_n - \Theta_0)^2}\left[\int_{-\infty}^{+\infty} dz\, zH_0^{(0)}H_n^{(0)}\right]^2 + 2\int_{-\infty}^{+\infty} dz\, H_0^{(0)}\left\{H_0^{(2)}(z, s + \sigma\gamma(\gamma\delta)^{\frac{1}{2}}) - H_0^{(2)}\right\} + 2\int_{-\infty}^{+\infty} dz\, H_0^{(1)}\left\{H_0^{(1)}(z, s + \sigma\gamma(\gamma\delta)^{\frac{1}{2}}) - H_0^{(1)}\right\} - \frac{\Xi_0(\sigma) - \Xi_0(0)}{\sigma} - \frac{\nu_1}{2}, \tag{A11}$$

where the arguments of the eigenfunctions which are not indicated are always meant to be z, s. This can be simplified by using the fact that

$$\sum_{n=0}^{\infty} b_n^2 = 1.$$

Substituting (35) and using (37) shows that the second and third terms in (A11) are equal to

$$-\frac{2}{3}\sum_{n=1}^{\infty}\frac{\sigma^2}{(\Theta_n - \Theta_0^2)}\left[\int_{-\infty}^{+\infty} dz\, zH_0^{(0)}H_n^{(0)}\right]^2,$$

so that one gets

$$R = -\tfrac{2}{3}\sigma\sum_{n=1}^{\infty}\frac{1}{\Theta_n - \Theta_0}\left[\int_{-\infty}^{+\infty} dz\, zH_0^{(0)}H_n^{(0)}\right]^2 - \frac{\Xi_0(\sigma) - \Xi_0(0)}{\sigma} - \frac{\nu_1}{2}. \tag{A12}$$

One now has to calculate $\Xi_0(\sigma) - \Xi_0(0)$, for which the equation for $H_n^{(2)}(z)$ is needed, which requires the extension of the expansion in Eq. (9) to order $(\gamma\delta)^2$. We will not give the explicit expression, but only note that the general form of the equations for the successive order eigenfunctions $H_0^{(2)}(z)$ is

$$L_1H_0^{(0)} = 0, \tag{A13a}$$

$$L_1H_0^{(1)} = L_2H_0^{(0)}, \tag{A13b}$$

$$L_1H_0^{(2)} = L_2H_0^{(1)} + L_3H_0^{(0)}, \tag{A13c}$$

where the operators L_1 and L_2 can be read of Eqs. (12) and (13). The replacement of s by $s + \sigma\gamma(\gamma\delta)^{\frac{1}{2}}$ does not affect L_1 but changes L_2 and L_3, and we indicate this by writing $L_2(\sigma)$, $L_3(\sigma)$ for the new operators. The eigenvalue Ξ_0 occurs in L_3, and to evaluate $\Xi_0(\sigma) - \Xi_0(0)$ one needs $L_3(\sigma) - L_3$. From the extension of the expansion (9) to order $(\gamma\delta)^2$ one finds

$$[L_3(\sigma) - L_3]H_0^{(0)}(z) = \frac{2\sigma}{3}z^2H_0^{(0)}(z) - [\Xi_0(\sigma) - \Xi_0(0)]H_0^{(0)}(z),$$

so that

$$\Xi_0(\sigma) - \Xi_0(0) = \frac{2\sigma}{3}\int_{-\infty}^{+\infty} dz\, z^2H_0^{(0)2} - \int_{-\infty}^{+\infty} dz\, H_0^{(0)}[L_3(\sigma) - L_3]H_0^{(0)}.$$

The last integral can be transformed with the help of (A13c), and then one gets

$$\Xi_0(\sigma) - \Xi_0(0) = \frac{2\sigma}{3}\int_{-\infty}^{+\infty} dz\, z^2H_0^{(0)2} + \int_{-\infty}^{+\infty} dz\, H_0^{(0)}[L_2(\sigma)H_0^{(1)}(z, s + \sigma\gamma(\gamma\delta)^{\frac{1}{2}}) - L_2H_0^{(1)}]$$

$$= -\frac{\nu_1\sigma}{2} + \sigma(\tfrac{2}{3})^{\frac{1}{3}} \int_{-\infty}^{+\infty} dz\, zH_0^{(0)}H_0^{(1)} - \sigma(\tfrac{2}{3})^{\frac{1}{3}}$$

$$\times \int_{-\infty}^{+\infty} dz\, zH_0^{(0)}[H_0^{(1)}(z, s + \sigma\gamma(\gamma\delta)^{\frac{1}{3}}) - H_0^{(1)}]$$

$$+ \int_{-\infty}^{+\infty} dz\, H_0^{(0)}L_2[H_0^{(1)}(z, s + \sigma\gamma(\gamma\delta)^{\frac{1}{3}}) - H_0^{(1)}], \tag{A14}$$

where we have used the equation of state (16), the relation

$$L_2(\sigma) = L_2 - \sigma(\tfrac{2}{3})^{\frac{1}{3}}z + \Omega_0(0) - \Omega_0(\sigma), \tag{A15}$$

and the fact that

$$\int_{-\infty}^{+\infty} dz\, H_0^{(1)}(z, s')H_0^{(0)}(z, s) = 0$$

for $s' = s$ and for $s' = s + \sigma\gamma(\gamma\delta)^{\frac{1}{3}}$.

Substituting (A14) in (A12) one obtains

$$R = -\frac{2\sigma}{3}\sum_{n=1}^{\infty}\frac{1}{\Theta_n - \Theta_0}\left[\int_{-\infty}^{+\infty} dz\, zH_0^{(0)}H_n^{(0)}\right]^2$$

$$- \left(\frac{2}{3}\right)^{\frac{1}{3}} \int_{-\infty}^{+\infty} dz\, zH_0^{(0)}H_0^{(1)}$$

$$+ \left(\frac{2}{3}\right)^{\frac{1}{3}} \int_{-\infty}^{+\infty} dz\, zH_0^{(0)}[H_0^{(1)}(z, s$$

$$+ \sigma\gamma(\gamma\delta)^{\frac{1}{3}}) - H_0^{(1)}]$$

$$- \frac{1}{\sigma}\int_{-\infty}^{+\infty} dz\, H_0^{(0)}L_2[H_0^{(1)}(z, s$$

$$+ \sigma\gamma(\gamma\delta)^{\frac{1}{3}}) - H_0^{(1)}]. \tag{A16}$$

Now express $H_0^{(1)}(z, s + \sigma\gamma(\gamma\delta)^{\frac{1}{3}})$ in terms of the complete set of functions $H_n^{(0)}$. From (A13b) written with s replaced by $s + \sigma\gamma(\gamma\delta)^{\frac{1}{3}}$, one obtains

$$H_0^{(1)}(z, s + \sigma\gamma(\gamma\delta)^{\frac{1}{3}})$$

$$= -\sum_{n=1}^{\infty}\frac{H_n^{(0)}}{\Theta_n - \Theta_0}\int_{-\infty}^{+\infty} dz\, H_n^{(0)}L_2(\sigma)H_0^{(0)}.$$

Eliminate in this way all the functions $H^{(1)}$ from the expression (A16) for R. Using again (A15) and the fact that the operator L_2 is Hermitian, one finds the desired result

$$R = 0.$$

APPENDIX III. PROOF OF EQ. (43)

We start from the equation

$$\bar{n}_2(\delta^+; l) = \frac{e^{-s\delta}}{l\lambda_0(s)}\iint_{-\infty}^{+\infty} dx\, dy\, \psi_0(x)\psi_0(y)\left[\frac{W(x)}{W(y)}\right]^{\frac{1}{3}}$$

$$\times P_\gamma(x \mid y, \delta)\exp\{\tfrac{1}{2}(\nu_0\gamma)^{\frac{1}{3}}(x + y)\}, \tag{A17}$$

which is an exact consequence of the general formula [see II, Eq. (28a)] for the Laplace transform of $\bar{n}_2(x; l)$. To show this, multiply Eq. (28a) of II by $\sigma\exp(\sigma\delta)$ and go the limit $\sigma \to \infty$. The left-hand side becomes $l\bar{n}_2(\delta^+; l)$. Using in the right-hand side the series [see II, Eq. (29)] for the resolvent $R_{s+\sigma}$, and the expression [See I, Eq. (9)] for the kernel $K_{s+\sigma}(x, y)$, it is easy to see that the first term of the series leads to (A17) since

$$\lim_{\sigma\to\infty} \sigma e^{\sigma\delta}p_s(x, y) = e^{-s\delta}P_\gamma(x \mid y, \delta),$$

and that all the further terms in the series vanish in the limit $\sigma \to \infty$.

One now "tames" the integral in (A17) by the same substitutions as used in Sec. II, replacing the variables x, y by z and ζ. Expanding the integrand in powers of $(\gamma\delta)^{\frac{1}{3}}$, using the expansions (5), (6), (7) and (11) for ν_0, l, s, and the eigenfunction $H_0(z)$, one finds up to order $(\gamma\delta)^{\frac{1}{3}}$

$$\bar{n}_2(\delta^+; l) = \frac{1}{2\delta l}\left[1 - \tfrac{3}{2}l_1(\gamma\delta)^{\frac{1}{3}}\right.$$

$$\left. + \tfrac{3}{2}(\gamma\delta)^{\frac{1}{3}}\int_{-\infty}^{+\infty} dz\, z^2H_0^{(0)}(z)^2\right], \tag{A18}$$

where also Eqs. (15) and (16) have been used. It is easy to verify that, to order $(\gamma\delta)^{\frac{1}{3}}$, (A18) is equivalent to Eq. (43).

On the van der Waals Theory of the Vapor–Liquid Equilibrium. IV. The Pair Correlation Function and Equation of State for Long-Range Forces

P. C. HEMMER*
Institutt for Teoretisk Fysikk, Norges Tekniske Høgskole, Trondheim, Norway
(Received 30 July 1963)

A fluid where the pair interaction potential between the particles consists of a hard core and a weak attraction of long range is considered. The pair correlation function and the equation of state in the one-phase region are calculated by a perturbation method with the ratio of the volume of the hard core to the interaction volume as perturbation parameter. The technique used is to expand the well-known density expansion in this perturbation parameter, and finally resum to all orders in the density. To lowest order, the equation of state is of the van der Waals type, and the asymptotic behavior of the pair correlation function near the critical point is that described by the Ornstein–Zernike theory. Higher approximations to these results are given.

1. INTRODUCTION

THE n-particle distribution function for a nonideal gas or liquid in equilibrium expresses the probabilities of observing different configurations of a set of n molecules. As is well-known, the two-particle or pair distribution function is particularly useful, because it determines completely the thermodynamic quantities and the angular distribution of molecular light scattering. General expansion theorems for these distribution functions for a classical homogeneous gas in powers of the fugacity or in powers of the density were first developed by Mayer and Montroll,[1] and their results are conveniently formulated in terms of linear graphs.[2]

If one now considers intermolecular potentials that contain a hard repulsive core and a weak attraction of long range (see Fig. 1), one is able to go further. In this case, a separation of the effects of the repulsive and the attractive forces is possible, and in particular, the derivation[3] of a van der Waals-like equation of state is based upon this type of potential. The attractive potential we assume to have the form

$$\Phi^{\text{attr}}(r) = -\gamma^3\varphi(\gamma r), \tag{1}$$

where γ is a small parameter, whose inverse characterizes the range of the attractive forces.[4] This gives a natural small parameter of expansion, viz.

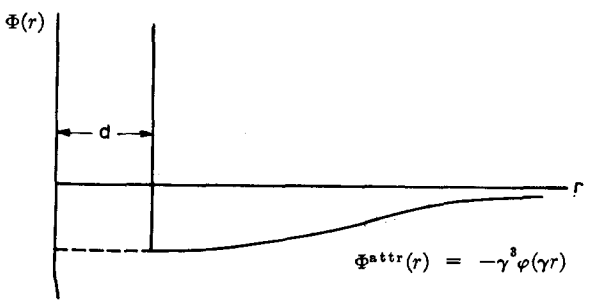


FIG. 1. The intermolecular potential.

$$\xi = (\gamma d)^3 = \frac{\text{volume of the hard core,}}{\text{interaction volume}} \tag{2}$$

and the object of this article is to give the pair correlation function as a power series expansion in ξ.

This idea was applied by Kac, Uhlenbeck, and Hemmer[5–7] to a one-dimensional gas model for which Kac's integral equation method[8] allows all calculations to be carried through explicitly. Because this expansion yields a condensation already in zeroth approximation (even in one dimension!), it is

* A major part of this work was performed at the Rockefeller Institute, New York, New York.

[1] J. E. Mayer and E. W. Montroll, J. Chem. Phys. 9, 2 (1941).

[2] E. E. Salpeter, Ann. Phys. (N.Y.) 5, 183 (1958); E. Meron, Phys. Fluids 1, 139 (1958); J. Math. Phys. 1, 192 (1960); T. Morita and K. Hiroike, Progr. Theoret. Phys. (Kyoto) 23, 1003 (1960); J. M. J. Van Leeuwen, J. Groeneveld, and J. de Boer, Physica 25, 792 (1959); and especially G. E. Uhlenbeck and G. W. Ford, in *Studies in Statistical Mechanics*, (North-Holland Publishing Company, Amsterdam, 1962), Vol. I, Part B.

[3] J. D. van der Waals, Dissertation, Leiden, 1873. L. S. Ornstein, Dissertation, Leiden, 1908.

[4] The parameter γ is not precisely defined, but one can make it definite by choosing

$$\frac{1}{\gamma^2} = \frac{\int dr\, r^2 \Phi^{\text{attr}}}{\int dr\, \Phi^{\text{attr}}}.$$

It should be emphasized that the results are independent of the precise definition. The power three in Eq. (1) reflects the dimensionality. With this choice, the integral over the attractive potential is independent of γ.

[5] M. Kac, G. E. Uhlenbeck, and P. C. Hemmer, J. Math. Phys. 4, 216 (1963).

[6] G. E. Uhlenbeck, P. C. Hemmer, and M. Kac, J. Math. Phys. 4, 229 (1963).

[7] P. C. Hemmer, M. Kac, and G. E. Uhlenbeck, J. Math. Phys. 5, 60 (1964) (previous paper).

[8] M. Kac, Phys. Fluids 2, 8 (1959).

hoped that it offers a better description of a real fluid than expansions starting with the ideal gas.

The method used here is to select the graphs in the density expansion of the pair correlation function which are of the same order in the parameter ξ, and to sum these infinite subclasses of graphs. This technique does not depend upon the dimensionality. We must, however, limit the discussion to a *single* gas or liquid phase, because the basic density expansion is not valid in the two-phase region.

In Sec. 2 the basic graphological theorem on the density expansion of distribution functions is stated. This expansion is modified in order to separate the effects of the repulsion and the attraction, and graphs with two types of lines are introduced. In Sec. 3 is considered the selection of graphs according to powers of the parameter ξ.

In discussing the pair correlation function, one must distinguish the case when the distance between the two particles is of the order of the hard-core diameter d (the near range) from the case when this distance is of the order of the range γ^{-1} of the attractive force (the far range). In the near range, the weak attractive forces only produce small corrections to the pair correlation function for a gas of hard spheres, and in Sec. 4 these are given up to first order in ξ. In Sec. 5 the more interesting far-range behavior is calculated up to second order in ξ. The result interpolates between two limiting cases: At low densities, the pair distribution function is given by the Boltzmann factor and thus reflects the detailed form of the force, while near the critical point the Ornstein–Zernike idea[9] of a universal form, determined by a few characteristics of the intermolecular forces only, is confirmed and generalized. In lowest order is obtained the obvious generalization to three dimensions of Uhlenbeck, Hemmer, and Kac's result[6] for one dimension. Lebowitz and Percus[10] have also obtained the same result by a different and interesting method, in which the so-called direct correlation function plays a crucial role. However, the first method is limited to one dimension, by the second method one encounters difficulties when one tries to go beyond the lowest order, and the present work is motivated by the need for a general and systematic approach.

In Sec. 6 the connection with the equation of state is discussed. Via the well-known virial and fluctuation theorems, the expansion of the pair correlation function implies an expansion in ξ of the equation of state, starting in zeroth order with a van der Waals-like equation. As a valuable check it is shown that the equation of state obtained in this way coincides with the ξ expansion obtained directly from the virial development of the pressure.

2. THE VIRIAL EXPANSION OF CLUSTER FUNCTIONS IN TERMS OF COMPOSITE GRAPHS

We consider a fluid of identical particles in thermal equilibrium at temperature T in an infinite volume such that the volume per particle is $v \equiv 1/\rho$. The intermolecular forces are assumed additative with an interaction potential $\Phi(r)$ between two molecules. As usual, let

$$f_{ij} = f(|\mathbf{r}_i - \mathbf{r}_j|) = e^{-\Phi(r_{ij})/\kappa T} - 1. \tag{3}$$

The canonical s-particle distribution function is defined such that $\bar{n}_s(\mathbf{r}_1, \mathbf{r}_2 \cdots \mathbf{r}_s; \rho, T)\, d\mathbf{r}_1 \cdots d\mathbf{r}_s$ is the probability of finding any s particles in $d\mathbf{r}_1 \cdots d\mathbf{r}_s$. The distribution functions depend upon the relative distances only, and when all these become very large, $\bar{n}_s$ approaches the constant value ρ^s. The corresponding clusterfunctions defined by

$$\chi_2(\mathbf{r}_1, \mathbf{r}_2; \rho) = \bar{n}_2(\mathbf{r}_1, \mathbf{r}_2; \rho) - \rho^2,$$

$$\chi_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3; \rho) = \bar{n}_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3; \rho) - \rho\bar{n}_2(\mathbf{r}_1, \mathbf{r}_2; \rho) - \rho\bar{n}_2(\mathbf{r}_1, \mathbf{r}_3; \rho) - \rho\bar{n}_2(\mathbf{r}_2, \mathbf{r}_3; \rho) + 2\rho^3, \tag{4}$$

etc., vanish when the particles are separated in two or more noninteracting groups (the cluster property). Therefore the integrals over all the variables except one exist, and the resulting so-called fluctuation integrals[11]

$$\int \chi_2(\mathbf{r}_1, \mathbf{r}_2; \rho)\, d\mathbf{r}_1 = \kappa T \rho \frac{\partial \rho}{\partial p} - \rho, \tag{5a}$$

$$\iint \chi_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3; \rho)\, d\mathbf{r}_1\, d\mathbf{r}_2 = \kappa^2 T^2 \rho \frac{\partial}{\partial p}\left(\rho \frac{\partial \rho}{\partial p}\right) - 3\kappa T \rho \frac{\partial \rho}{\partial p} + 2\rho, \tag{5b}$$

etc., are expressed in thermodynamic variables. The clusterfunction χ_2 is usually called the pair correlation function, and we put for brevity

$$g(r_{12}) \equiv \chi_2(\mathbf{r}_1, \mathbf{r}_2; \rho). \tag{6}$$

It will be convenient to introduce another set of functions μ_s, viz.

[9] L. S. Ornstein and F. Zernike, Proc. Acad. Sci. Amsterdam 17, 793 (1914), Phys. Z. 19, 134 (1918).

[10] J. L. Lebowitz and J. K. Percus, J. Math. Phys. 4, 248 (1963).

[11] Equation (a) is contained in reference 9. The result (b) is given in reference 6.

$$\mu_2(\mathbf{r}_1, \mathbf{r}_2; \rho) = \chi_2(\mathbf{r}_1, \mathbf{r}_2; \rho) + \rho\delta(\mathbf{r}_1 - \mathbf{r}_2),$$

$$\begin{aligned}\mu_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3; \rho) &= \chi_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3; \rho) \\ &+ \delta(\mathbf{r}_{12})\chi_2(\mathbf{r}_1, \mathbf{r}_3; \rho) + \delta(\mathbf{r}_{13})\chi_2(\mathbf{r}_1, \mathbf{r}_2; \rho) \\ &+ \delta(\mathbf{r}_{23})\chi_2(\mathbf{r}_1, \mathbf{r}_2; \rho) + \rho\delta(\mathbf{r}_{12})\delta(\mathbf{r}_{13}),\end{aligned} \tag{7}$$

etc., which also possess the cluster property. From the fluctuation integrals (5) follow

$$\int \mu_2(\mathbf{r}_1, \mathbf{r}_2; \rho)\, d\mathbf{r}_1 = \rho\kappa T \frac{\partial\rho}{\partial p}, \tag{8}$$

$$\iint \mu_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3; \rho)\, d\mathbf{r}_1\, d\mathbf{r}_2 = \rho\kappa^2 T^2 \frac{\partial}{\partial p}\left(\rho\frac{\partial\rho}{\partial p}\right). \tag{9}$$

The expansion of the distribution functions or the cluster functions as power series in the number density ρ is well known.[1,2] The result can be expressed in terms of cluster integrals, to which linear graphs can be associated in the following way.

A graph occuring in the virial expansion of the clusterfunction $\chi_s(\mathbf{r}_1 \cdots \mathbf{r}_s; \rho)$ will consist of s *root points* labeled 1, 2, $\cdots$ s (represented by white circles) and k *field points* labeled $s + 1, \cdots, s + k$ (represented by black circles). k is any nonnegative integer. Between these $(k + s)$ points, single bonds may be drawn satisfying the following topological condition: The resulting graph is connected and becomes a star[12] if additional lines are inserted connecting those root points which are not already directly connected. Such graphs are called *irreducible*.

The contribution corresponding to a particular graph is the following integral: Each point represents a factor ρ, and a bond between two points i and j represents a factor f_{ij}, given by Eq. (3). Furthermore a factor $1/s$ is associated with the graph, where s, the symmetry number, denotes the number of permutations of the field points which do *not* lead to a new diagram. Finally, an integration is carried out over all the field point variables.

A graph that has s labeled root points and *unlabeled* field points is called a *generic s graph*.

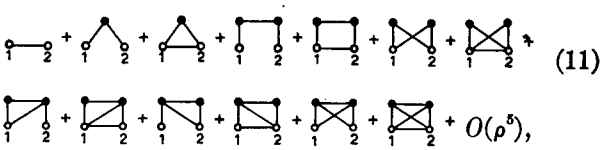
The well-known virial expansion of the cluster functions can then be stated as follows:

$$\chi_s(\mathbf{r}_1, \cdots \mathbf{r}_s; \rho) = \text{the sum of all distinct irreducible generic } s \text{ graphs.} \tag{10}$$

As an illustration, the first few terms of the expansion for $s = 2, 3$ are given below.

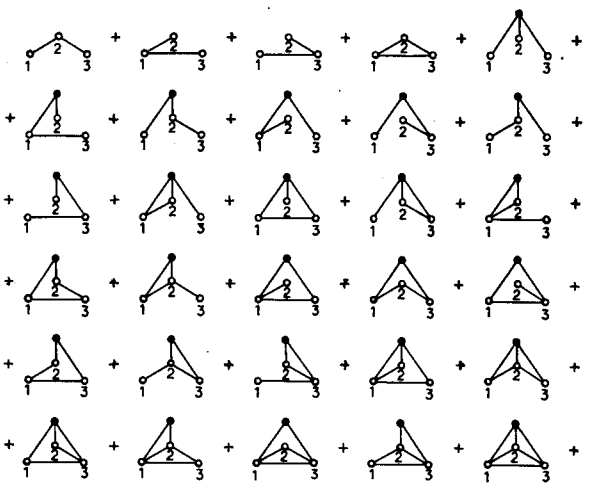
[12] A star is a connected graph with no articulation points. An articulation point has the property that if the graph is cut at such a point it becomes disconnected.

$\chi_2(\mathbf{r}_1, \mathbf{r}_2; \rho) =$



$+ O(\rho^5),$ (11)

$\chi_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3; \rho) =$



$+ O(\rho^5).$ (12)

Let us now consider a potential which consists of a hard repulsive core of diameter d and an attractive part of the form (1),

$$\Phi(r) = \Phi^{\text{hc}}(r) + \Phi^{\text{attr}}(r). \tag{13}$$

Then Mayer f functions of the form (3) can be introduced for the repulsive and the attractive part separately, and one sees immediately that

$$f_{ij} = f_{ij}^{\text{hc}} + f_{ij}^{\text{attr}} + f_{ij}^{\text{hc}} f_{ij}^{\text{attr}}. \tag{14}$$

Furthermore, it is clear that, because of the weakness of the attraction, $f^{\text{attr}}(r) \approx -\Phi^{\text{attr}}(r)/\kappa T$, and an expansion in powers of the attractive potential is natural:

$$f_{ij} = f_{ij}^{\text{hc}} + (1 + f_{ij}^{\text{hc}}) \sum_{r=1}^{\infty} v_{ij}^r \frac{1}{r!}, \tag{15}$$

with

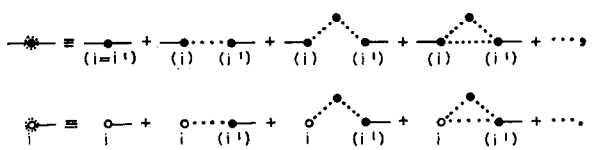
$$v_{ij} = -\Phi^{\text{attr}}(r_{ij})/\kappa T. \tag{16}$$

Thus we are led to introduce two different types of bonds: A factor f_{ij}^{hc} is represented by a dotted line (a *short-range bond*), and a factor v_{ij} is represented by a solid line (a *long-range bond*). Equation (15) shows that a given f bond in a graph can be represented by any nonnegative number of long-range bonds with or without the simultaneous presence of a short-range bond between the two

points. The resulting graphs we call *composite* in distinction from the original *simple* graphs. If the two types of bonds are considered distinct in the definition of the symmetry number of a graph, and if r long-range bonds between two points i, j are associated with a factor $v_a^r/r!$, then the expansion (10) of the clusterfunctions can be written

$$\chi_s(\mathbf{r}_1, \cdots \mathbf{r}_s; \rho) = \text{the sum of all distinct composite irreducible generic } s \text{ graphs.} \qquad (17)$$

Finally, we make the following convention. A graph with a dotted circle surrounding a field point with two incident long-range bonds (or a root point with one long-range bond) stands for a sum of graphs, viz. the original graph plus all graphs resulting from cutting the graph at this point and inserting all possible short-range graphs occurring in the χ_2 expansion (11). As an illustration, the first few terms are shown in Eq. (18).



(18)

It is clear that, if the diagram without the dotted circle is irreducible, then the diagrams with short bonds inserted are also irreducible. One also sees easily that if a reducible (i.e., disconnected or containing an articulation point) short-range graph is inserted, the resulting graph is reducible.

The contribution from a graph where *all* field points of degree two (and possibly some root points) are marked with dotted circles is found as before, except that each such point is now given two labels i, i', and instead of just a factor ρ, the point is associated with the function

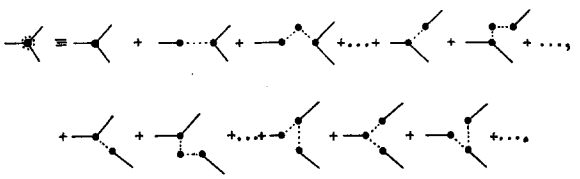
$$\rho\delta(\mathbf{r}_i - \mathbf{r}_{i'}) + \chi_2^{hc}(\mathbf{r}_i, \mathbf{r}_{i'}; \rho) \equiv \mu_2^{hc}(\mathbf{r}_i, \mathbf{r}_{i'}; \rho), \qquad (19)$$

by the definition (7) of μ_2. The first term in (19) reproduces the original graph, and the second term represents graphs with insertions. In case of a field point, one integrates over both variables; in case of a root point, only over the variable which does not carry the original root label. We omit the proof, which merely consists of verifying that the symmetry numbers are correctly given, because the types of graphs to be considered later make a direct verification easy.

Similarly, a graph with a dotted circle surrounding a field point of degree 3 (or a root point of degree 2) stands for a sum of graphs. These are found by opening the graph at this point and between the three ends i, k, l inserting all short-range graphs represented by the function μ_3^{hc}, Eq. (7):

$$\mu_3^{hc}(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k; \rho) = \rho\delta(\mathbf{r}_{ij})\delta(\mathbf{r}_{jk}) + \delta(\mathbf{r}_{ij})\chi_2^{hc}(\mathbf{r}_i, \mathbf{r}_k; \rho) + \delta(\mathbf{r}_{ik})\chi_2^{hc}(\mathbf{r}_i, \mathbf{r}_j; \rho) + \delta(\mathbf{r}_{jk}) \times \chi_2^{hc}(\mathbf{r}_i, \mathbf{r}_j; \rho) + \chi_3^{hc}(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k; \rho). \qquad (20)$$

A δ function symbolizes that the two points contained in its argument are identical. A function χ_s^{hc} indicates insertion of all short-range graphs occurring in the χ_s^{hc} expansion (10) between the s points occurring in its argument. In Eq. (21) is given all graphs with less than three short-range bonds.



(21)

Again it is straightforward to prove that these insertions between the three points are the only ones which conserve irreducibility of the complete graph.

3. SELECTION OF GRAPHS

In order to clarify how graphs are selected according to their order in ξ, consider for a moment graphs with only long-range bonds. Because of the weakness of the forces, each such bond is of $O(\xi)$, and because of the long range, each volume integration reduces the order in ξ by one (Introduce new integration variables $\mathbf{r}_i^* = \gamma\mathbf{r}_i$!) The contribution of a graph with α long-range bonds and β field points is therefore of $O(\xi^{\alpha-\beta})$. The leading terms in the expansion of $g(\mathbf{r}_{12})$ are of $O(\xi)$, and these graphs are chain graphs between the root points 1 and 2 [Fig. 2(a)].

When short-range bonds are present, the order in ξ of a graph must, of course, still be equal to or less than the number α of long-range bonds. If there exists a path of short-range bonds from every field point to a root point, the graph *is* of $O(\xi^\alpha)$, because the vanishing of $f^{hc}(r)$ for $r > d$ reduces the region in which the integrand is nonzero to a region of $O(1)$. If such a path does not exist, the order is reduced. Imagine for a moment all long-range bonds in a composite graph erased, so that the graph disintegrates into disconnected parts

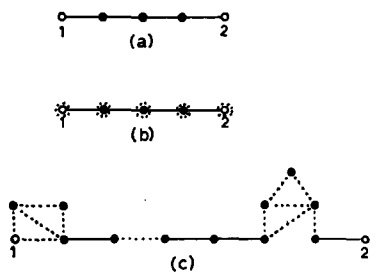


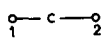
FIG. 2. A composite chain graph (b) generated from the simple chain graph (a). One actual graph contained in the collection (b) is shown below (c).

(graphs or points). Let β be the number of these parts that do not contain root points. By considering these parts as entities in the original graph, all connections are long-range bonds. It is easy to extend the result for graphs with only long-range bonds to show that the order in ξ is reduced by β. The graph is therefore of $O(\xi^{\alpha-\beta})$. As a consequence, all the composite graphs we can derive by adding dotted circles around points of a given solid line graph are of the same order as the latter. As an example, the *composite chain graphs* generated [Fig. 2(b)] from the simple chain graphs [Fig. 2(a)] are all of $O(\xi)$.

The infinite sum of all possible composite chains will occur so frequently in the following that it is advantageous to introduce the graphical notation

(22)

This superbond will be called a *chain bond*, and it is the following function of the end-point coordinates:



$$= \rho^2 C(r_{12};\rho) = \rho^2 v(r_{12}) + \rho^2 \int v(r_{13})\, d\mathbf{r}_3 \times \mu_2^{hc}(r_{34})\, d\mathbf{r}_4 v(r_{42}) + \rho^2 \int v(r_{13})\, d\mathbf{r}_3 \mu_2^{hc}(r_{34})\, d\mathbf{r}_4 \times v(r_{45})\, d\mathbf{r}_5 \mu_2^{hc}(r_{56})\, d\mathbf{r}_6 v(r_{62}) + \cdots . \quad (23)$$

The symmetry number for a graph in (22) is obviously the product of the symmetry numbers for the short-range groups, and therefore correctly given by Eq. (23).

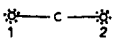
By use of Fourier transforms one evaluates the sum (23) over convolution integrals. Defining the three-dimensional Fourier transform by

$$\tilde{v}(\mathbf{k}) = \int d\mathbf{r}\, e^{i\mathbf{k}\mathbf{r}\gamma} v(\mathbf{r}), \qquad v(\mathbf{r}) = (\gamma/2\pi)^3 \int d\mathbf{k}\, e^{-i\mathbf{k}\mathbf{r}\gamma} \tilde{v}(\mathbf{k}), \quad (24)$$

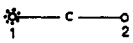
the chain bond is given by the compact expression

$$\tilde{C}(k) = \tilde{v}(k)\{1 + \tilde{v}(k)\tilde{\mu}_2^{hc}(k) + [\tilde{v}(k)\tilde{\mu}_2^{hc}(k)]^2 + \cdots\} = \tilde{v}(k)/[1 - \tilde{v}(k)\tilde{\mu}_2^{hc}(k)]. \quad (25)$$

Similarly we evaluate the sum of all composite chain graphs with (or without) short-range groups inserted at one or both ends:



$$= D(r_{12};\rho) = \int \mu_2^{hc}(r_{13})\, d\mathbf{r}_3 C(r_{34})\, d\mathbf{r}_4 \mu_2^{hc}(r_{42}), \quad (26a)$$



$$= E(r_{12};\rho) = \rho \int \mu_2^{hc}(r_{13})\, d\mathbf{r}_3 C(r_{32}). \quad (26b)$$

We get

$$\tilde{D}(k) = \tilde{v}(k)[\tilde{\mu}_2^{hc}(k)]^2/[1 - \tilde{v}(k)\tilde{\mu}_2^{hc}(k)], \quad (27a)$$

$$\tilde{E}(k) = \tilde{v}(k)\rho\tilde{\mu}_2^{hc}(k)/[1 - \tilde{v}(k)\tilde{\mu}^{hc}(k)]. \quad (27b)$$

Notice that $\tilde{v}(k)$ does *not* contain the parameter γ. The hard-core function, however, can be expanded:

$$\tilde{\mu}_2^{hc}(k) = \int e^{i\mathbf{k}\mathbf{r}\gamma} \mu_2^{hc}(r)\, d\mathbf{r} = \tilde{\mu}_2^{hc}(0) - \tfrac{1}{6}\gamma^2 k^2 \int r^2 g^{hc}(r)\, d\mathbf{r} + O(\gamma^4). \quad (28)$$

By Eq. (9) we therefore have, to lowest order,

$$\tilde{\mu}_2^{hc}(k)/\kappa T = \eta \equiv \rho(\partial p^{hc}/\partial\rho)^{-1}. \quad (29)$$

Denoting the Fourier transform of the *attractive* potential by $\tilde{\Phi}(k)$, we have, to lowest order in γ,

$$D(r,\rho) = -\left(\frac{\gamma}{2\pi}\right)^3 \cdot \kappa T \eta^2 \times \int d\mathbf{k}\, e^{-i\mathbf{k}\mathbf{r}\gamma} \frac{\tilde{\Phi}(k)}{1 + \eta\tilde{\Phi}(k)}, \quad (30)$$

and

$$E(r,\rho) = (\rho/\kappa T\eta) D(r,\rho). \quad (31)$$

4. THE CORRELATION FUNCTION IN THE NEAR RANGE

Let us now investigate the correlation function $g(r)$ when $r = O(d)$. In zeroth approximation, $g(r) = g^{hc}(r)$, of course, corresponding to the sum of all irreducible 2 graphs with short-range bonds.

Which graphs contribute to the first approximation $O(\xi)$? Denote the number of long-range bonds by $\alpha \geq 1$. Erasing these, a graph of $O(\xi)$ contains $\alpha - 1$ disconnected groups of short-range graphs (or single field points) that are free of root points. Because the complete graph has to be connected and irreducible, the only way the α solid bonds

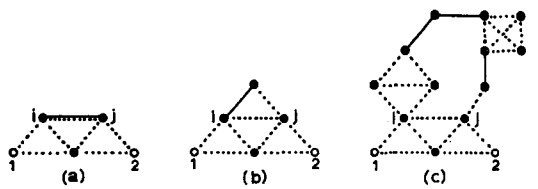


FIG. 3. Three examples of graphs whose connected root parts have zero (a), one (b), and two (c) articulation points, respectively.

can connect the $\alpha - 1$ groups is by forming a composite chain contained in (22).

Consider now the remaining *root part* of the graphs. Either it consists of (α) two disconnected parts each containing one root point, or (β) it is connected.

Possibility (α) occurs if and only if the complete graph is a graph occurring in the expansion of $D(r_{12}; \rho)$, Eq. (26). To lowest order, the contribution of these graphs is the constant,

$$D = -\left(\frac{\gamma}{2\pi}\right)^3 \kappa T \eta^2 \int d\mathbf{k} \frac{\tilde{\Phi}(k)}{1 + \eta\tilde{\Phi}(k)}. \qquad (32)$$

This follows from Eq. (30) and the assumption that $r = O(d)$.

Now consider possibility (β) of a connected root part. It can have 0, 1, or 2 articulation points, as illustrated in Fig. 4. It is preferable to cut the graphs at this (these) articulation point(s) and associate the part(s) which does (do) not contain root points with the composite chain. Thus the complete graph is divided in a connected irreducible root part and a chain graph with short-range parts at the ends.

All these graphs are contained in the following prescription: Start with any graph G occurring in the expansion of $g^{\text{hc}}(r_{12}; \rho)$. Take any graph D in the expansion of $D(r_{ij}; \rho)$, Eq. (26), and hang it onto G by letting the points i and j coincide with any pair of points in G. Let the root points 1, 2 of G be the root points of the complete graph. Take the sum of all graphs resulting from different choices for G, D, and the pair of common points (denoted by i, j in Fig. 3).

Consider first one G graph, G_n, with n points, whose contribution $W(G_n)$ is proportional to ρ^n. Since the distance between the end points i, j of the D graphs is necessarily of $O(d)$, it follows from Eq. (32) that the total contribution of the graphs constructed by hanging all possible D graphs on a given pair of points of G_n is $\rho^{-2}DW(G_n)$. The symmetry number of each combined graph is hereby given as the product of the symmetry numbers of the D and G graphs, which is clearly correct. The factor ρ^{-2} avoids counting the common pair of points twice. There are $\frac{1}{2}n(n-1)$ ways of hanging each D graph on G_n. Hence the sum of all graphs resulting from hanging all possible D graphs in all possible ways onto G_n equals

$$\tfrac{1}{2}n(n-1)\rho^{-2}DW(G_n) = \tfrac{1}{2}D(\partial^2/\partial\rho^2)W(G_n). \qquad (33)$$

Summing over all possible G graphs, we obtain the contribution

$$\tfrac{1}{2}D(\partial^2/\partial\rho^2)g^{\text{hc}}(r_{12}; \rho). \qquad (34)$$

Adding the first class of graphs, Eq. (32), to this, and using Eq. (4), we find for the correlation function up to $O(\xi)$:

$$g(r_{12}; \rho) = g^{\text{hc}}(r_{12}; \rho) + \tfrac{1}{2}D(\partial^2/\partial\rho^2)\bar{n}_2^{\text{hc}}(r_{12}; \rho). \qquad (35)$$

As a check we make a comparison with the exact solution[6] for a one-dimensional gas with a hard core δ and the exponential attraction

$$\Phi^{\text{attr}}(x)/\kappa T = -\nu_0\gamma e^{-\gamma|x|}. \qquad (36)$$

With the obvious modifications because of dimensionality, Eq. (32) gives

$$D = \gamma\nu_0\rho(1 - \rho\delta)^4/B, \qquad (37)$$

with

$$B = \rho^{-1}[1 - 2\nu_0\rho(1 - \rho\delta)^2]^{\frac{1}{2}}. \qquad (38)$$

Tonks' equation of state,[13]

$$p^{\text{h.c.}} = \kappa T\rho/(1 - \rho\delta), \qquad (39)$$

has been used. The resulting expression (35) for the distribution function can now be written

$$\bar{n}_2(x; \rho) = [1 + \tfrac{1}{2}\gamma\nu_0 B^{-1}\rho(1 - \rho\delta)^4 (\partial^2/\partial\rho^2)]\bar{n}_2^{\text{hc}}(x; \rho), \qquad (40)$$

which agrees with Eq. (44) of reference 6.

5. THE CORRELATION FUNCTION IN THE FAR RANGE

In this section we study the correlation between two particles separated a distance of the order of the range γ^{-1} of the attractive force, or greater. It is essentially this far-range correlation that can be measured by x-ray scattering experiments.

The first question arises immediately: what can be said about the hard-sphere correlation function for such large distances? It is of course so that *if* $g^{\text{hc}}(r)$ should happen to decay like r^{-n}, say, then $g^{\text{hc}}(r) = O(\xi^{\frac{1}{2}n})$ when $r\gamma = O(1)$. From the fluctuation theorem (5) it follows[14] that $n > 3$,

[13] L. Tonks, Phys. Rev. 50, 955 (1936).

[14] This argument would fail for a density corresponding to the so-called Kirkwood transition, i.e., a phase transition in a dense hard-sphere gas, if such a transition exists.

excluding contribution to lowest order. In the absence of a proof, we make the *assumption* that all moments of $g^{hc}(r)$ exist, so that in the far range $g^{hc}(r)$ does not contribute to any order in ξ. This is true in one dimension, where the correlation function is explicitly known.[15] (As far as the calculation is carried here, the existence of the second moment is all that is needed.) Graphs, in which there exists a path of short-range bonds between the two root points, can consequently be excluded from consideration.

To find the lowest-order graphs, the considerations in the previous section still apply, but possibility (β) of a connected root part is now excluded. The contributing graphs are therefore simply those symbolized by

$$g^{(1)}(r_{12}) = D(r_{12};\rho) = \frac{\xi}{(2\pi d)^3}\int d\mathbf{k}\, e^{-i\mathbf{k}\mathbf{r}_{12}\gamma}\,\frac{\tilde{v}(k)[\tilde{\mu}_2^{hc}(k)]^2}{1-\tilde{v}(k)\tilde{\mu}_2^{hc}(k)}, \qquad (41)$$

using Eqs. (26a) and (27a). By $g^{(n)}$ we denote the sum of all graphs of $O(\xi^n)$. The discussion in Sec. 3 shows that n is integer. However, the *further* expansion of the function $g^{(n)}$ goes with fractional powers of ξ (but integer powers of $\gamma d = \xi^{\frac{1}{3}}$). In particular, the expansion (28) of $\tilde{\mu}_2^{hc}(k)$ implies [by Eq. (41)] the following expansion for $g^{(1)}$:

$$\tilde{g}^{(1)}(k) = -\kappa T\,\frac{\eta^2\tilde{\Phi}(k)}{1+\eta\tilde{\Phi}(k)} - \frac{k^2\gamma^2}{6}\{1-[1+\eta\tilde{\Phi}(k)]^{-2}\}\int r^2 g^{hc}(r)\,dr. \qquad (42)$$

The first, lowest-order term in Eq. (42) is exactly the result obtained in reference 6 (for one dimension), and by Lebowitz and Percus.[10] In one dimension, the second term in (42) is of $O[(\gamma\delta)^3]$, and is not calculated in reference 6.

Let us also calculate $g^{(2)}$. We recall that the graphs of $O(\xi^2)$ have two long-range bonds more than the number of root-free isolated parts remaining when all solid bonds are removed. One finds two classes of graphs of this order.

Class I consists of all actual graphs corresponding to the four chain-bond graphs in Fig. 4.

The contribution of all four chain diagrams can can be written as one integral[16]:

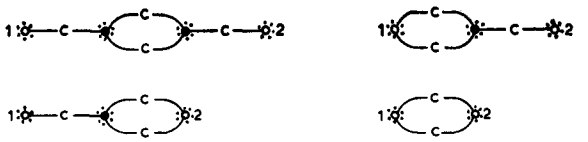


FIG. 4. The second-order graphs that form Class I.

$$g_{\mathrm{I}}^{(2)} = \frac{1}{2}\int\cdots\int[\rho^{-1}E(r_{13})+\delta(\mathbf{r}_{13})]\,d\mathbf{r}_3 \times\mu_3^{hc}(\mathbf{r}_3\mathbf{r}_4\mathbf{r}_5)\,d\mathbf{r}_4\,d\mathbf{r}_5 C(r_{46})C(r_{57})\,d\mathbf{r}_6\,d\mathbf{r}_7 \times\mu_3^{hc}(\mathbf{r}_6\mathbf{r}_7\mathbf{r}_8)\cdot d\mathbf{r}_8[\rho^{-1}E(r_{82})+\delta(\mathbf{r}_{82})], \qquad (43)$$

where $E(r)$ is given by Eq. (31). Introducing $\mathbf{r}^* = \gamma\mathbf{r}$ for a moment and using

$$\lim_{\gamma\to 0}\mu_3^{hc}(\mathbf{r}_i^*/\gamma,\mathbf{r}_j^*/\gamma,\mathbf{r}_k^*/\gamma) = \delta(\mathbf{r}_{ij})\delta(\mathbf{r}_{ik})\iint\mu_3^{hc}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)\,d\mathbf{r}_1\,d\mathbf{r}_2, \qquad (44)$$

we obtain to lowest order

$$g_{\mathrm{I}}^{(2)} = \tfrac{1}{2}(\kappa T)^4\left(\eta\frac{\partial\eta}{\partial\rho}\right)^2\iint[E(\mathbf{r}_{13})+\rho\delta(\mathbf{r}_{13})]\,d\mathbf{r}_3 \cdot C^2(r_{38})\,d\mathbf{r}_8[E(\mathbf{r}_{82})+\rho\delta(\mathbf{r}_{82})]. \qquad (45)$$

Equations (9) and (29) have been used. By insertion for $E(\mathbf{r})$, Eq. (31), we find for the Fourier transform

$$\tilde{g}_{\mathrm{I}}^{(2)}(k) = \tfrac{1}{2}[\eta(\partial\eta/\partial\rho)]^2(\kappa T)^4[1+\eta\tilde{\Phi}(k)]^{-2}\tilde{C}^2. \qquad (46)$$

The last factor is the Fourier transform of $C^2(\mathbf{r})$.

A typical graph in the second class is constructed as follows. Take one of the chain graphs in $g^{(1)}$. It consists of m solid bonds, say, alternating with $(m+1)$ short-range graphs (or single field points). Now attach the ends of a second chain graph to any two points of one of these short-range graphs, thereby making the complete graph of $O(\xi^2)$. In other words, we replace one short-range graph (occurring in g^{hc}) with a graph occurring in the $O(\xi)$ correction to g^{hc} in the near range. Appealing to the discussion in Sec. 4 leading to Eq. (34), we find that the sum $g_{\mathrm{II}}^{(2)}(r)$ of all graphs of this type is given by

$$\tilde{g}_{\mathrm{II}}^{(2)}(k) = \sum_{m=1}^{\infty}(m+1)[\tilde{v}(k)\mu_2^{hc}(k)]^m\,\frac{D}{2}\,\frac{\partial^2}{\partial\rho^2}\,\tilde{g}^{hc}(k). \qquad (47)$$

The factor $(m+1)$ occurs because the replacement just mentioned can be made at $(m+1)$ different places. Expanding the hard-core functions in γ, as in Eq. (28), we find to lowest order

$$\tilde{g}_{\mathrm{II}}^{(2)}(k) = \tfrac{1}{2}\kappa T D(\partial^2\eta/\partial\rho^2)\{[1+\eta\tilde{\Phi}(k)]^{-2}-1\}. \qquad (48)$$

The constant D is given by Eq. (32). The complete

[15] The result of F. Zernike and J. A. Prins, Z. Phys. **41**, 184 (1927), can be written

$$\int_0^\infty dx\, e^{-\sigma x}\tilde{n}_2^{hc}(x;\rho) = \{[\rho+\sigma(1-d\rho)]e^{d\sigma}-\rho\}^{-1},$$

where $1/\rho = L/N$ = the specific length.

[16] The factor $\frac{1}{2}$ is necessary because the integral counts all generic graphs twice, except (a) graphs with symmetry number 2, and (b) graphs where the "parallel" chain bonds stand for a double long-range bond between one pair of points i, j, represented by the factor $\frac{1}{2}v_{ij}^2$ [Eq. (15)].

second-order term is

$$g^{(2)} = g_{\mathrm{I}}^{(2)} + g_{\mathrm{II}}^{(2)}. \tag{49}$$

Also this result can be checked on the one-dimensional model of reference 6. With the notation of Sec. 4, we find in this case

$$\begin{aligned}\tilde{g}^{(2)}(k) &= 2\gamma\rho\nu_0^2(1-\rho\delta)^6 B^{-1}(k^2+B^2\rho^2)^{-2} \\ &\times [(1-3\rho\delta)^2(k^2+1)^2(k^2+4B^2\rho^2)^{-1} \\ &+ \rho\delta(3\rho\delta-2)(2k^2+1+B^2\rho^2)],\end{aligned} \tag{50}$$

which is precisely the Fourier transform of the result obtained in reference 6.

We conclude this section with a remark on the corresponding expansion of the higher distribution functions. All graphs in the χ_s expansion (10) are connected, and as a result $\chi_s = O(\xi^{s-1})$, when *all* relative distances between the s particles are of order of the range γ^{-1} of the attraction. It is then easy to show that, in this range, Kirkwood's superposition principle

$$\begin{aligned}\rho^3\bar{n}_3(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3;\rho) &= \bar{n}_2(\mathbf{r}_1,\mathbf{r}_2;\rho) \\ &\times \bar{n}_2(\mathbf{r}_1,\mathbf{r}_3;\rho)\bar{n}_2(\mathbf{r}_2,\mathbf{r}_3;\rho)\end{aligned} \tag{51}$$

is satisfied to $O(\xi)$. By means of the relations (4), both sides of Eq. (51) can be expressed in terms of cluster functions:

$$\begin{aligned}\rho^3\bar{n}_3(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3) &= \rho^6 + \rho^4[\chi_2(r_{12}) + \chi_2(r_{13}) \\ &+ \chi_2(r_{23})] + \rho^3\chi_3(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3),\end{aligned}$$

and

$$\begin{aligned}&\bar{n}_2(r_{12})\bar{n}_2(r_{13})\bar{n}_2(r_{23}) \\ &\quad= \rho^6 + \rho^4[\chi_2(r_{12}) + \chi_2(r_{13}) + \chi_2(r_{23})] \\ &\quad+ \rho^2[\chi_2(r_{12})\chi_2(r_{13}) + \chi_2(r_{12})\chi_2(r_{23}) \\ &\quad+ \chi_2(r_{13})\chi_2(r_{23})] + \chi_2(r_{12})\chi_2(r_{13})\chi_2(r_{23}).\end{aligned}$$

The right-hand sides of these identities differ only by terms of $O(\xi^2)$, and this proves the assertion.

6. THE EQUATION OF STATE

As is well known, the equation of state can be calculated from the pair distribution function in two ways, either through the fluctuation theorem (5a) or through the virial theorem

$$\begin{aligned}p &= \kappa T\rho + \frac{2\pi d^3}{3}\kappa T\bar{n}_2(d^+;\rho) \\ &- \frac{1}{6}\int d(\mathbf{r})r\frac{d\Phi^{\mathrm{attr}}(r)}{dr}\bar{n}_2(r;\rho).\end{aligned} \tag{52}$$

By insertion of the results (35) and (49) one finds that, to $O(\xi)$, the two theorems give the same expansion

$$p = p^{(0)} + p^{(1)} + \cdots \tag{53}$$

of the pressure in powers of ξ. To lowest order, the *van der Waals-like* equation[6]

$$\begin{aligned}p^{(0)} &= \kappa T\rho + \frac{2\pi d^3}{3}\kappa T\bar{n}_2^{\mathrm{hc}}(d^+;\rho) \\ &+ \frac{\rho^2}{2}\int d(\mathbf{r})\Phi^{\mathrm{attr}}(r),\end{aligned}$$

or

$$p^{(0)} = p^{\mathrm{hc}} + \tfrac{1}{2}\tilde{\Phi}(0)\rho^2 \tag{54}$$

is obtained. To next order, the length of the calculation contrasts with the simplicity of the result:

$$\frac{p^{(1)}}{\kappa T} = \frac{1}{2}\left(\frac{\gamma}{2\pi}\right)^3\left(\rho\frac{\partial}{\partial\rho} - 1\right)\int d\mathbf{k}\ln[1+\eta\tilde{\Phi}(k)], \tag{55}$$

where η is given by Eq. (29).

General proofs of the consistency between the virial expansions of the pair correlation function, and of the pressure through the fluctuation and the virial theorems are available.[17] Still it may be instructive to see how the result (55) comes about.

The virial expansion of the pressure can be formulated in terms of graphs (with field points only) as follows[18]:

$$\frac{p}{\kappa T} = \rho + \left(1 - \rho\frac{\partial}{\partial\rho}\right)\sum_{\nu=2}^{\infty} A_\nu, \tag{56}$$

where A_ν is the sum of all generic stars with ν points. The integration is performed over all the variables, except one. (The inverse symmetry number is again associated with the graphs). One can again introduce composite graphs and select graphs by the same method as in Sec. 3.

The zero-order terms include all hard-core graphs plus the graph that consists of a single long-range bond:

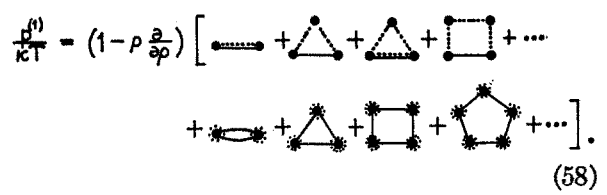
$$\begin{aligned}\frac{p^{(0)}}{\kappa T} &= \rho + \left(1 - \rho\frac{\partial}{\partial\rho}\right)\left[\sum_{\nu=2}^{\infty} A_\nu^{\mathrm{hc}} + \bullet\!\!-\!\!\bullet\right] \\ &= \frac{p^{\mathrm{hc}}}{\kappa T} - \tfrac{1}{2}\tilde{v}(0)\rho^2,\end{aligned} \tag{57}$$

in accordance with Eq. (54). The following graphs

[17] J. de Boer, Rept. Progr. Phys. **12**, 305 (1949). G. S. Rushbrooke and H. I. Scoins, Proc. Roy. Soc. (London) **A216**, 203 (1953).

[18] J. E. Mayer and M. G. Mayer, *Statistical Mechanics* (John Wiley & Sons, Inc., New York, 1940), Chap. 13.

give the next-order pressure correction:



(58)

The graphs in the upper line are obtained from all graphs in χ_2^{hc} by joining the root points with a solid bond, simultaneously turning root points into field points. The lower line consists of all *composite* chain graphs. We may write

$$\frac{p^{(1)}}{\kappa T} = \left(1 - \rho\frac{\partial}{\partial\rho}\right)\sum_{n=1}^{\infty}\frac{1}{2n}\int \mu_2^{hc}(r_{12})\,d\mathbf{r}_2 v(r_{23})\,d\mathbf{r}_3 \times \mu_2^{hc}(r_{34})\cdots\mu_2^{hc}(r_{2n-1,2n})\,d\mathbf{r}_{2n}v(r_{2n,1}), \quad (59)$$

where the term $n = 1$ represents all graphs of the first category. We omit the complete analysis leading to the factor $1/2n$, and only notice that the symmetry number for a ring graph with n long-range bonds is precisely $2n$. Such a ring graph with an additional hard-core graph inserted at one point will be produced n times by the integral (59), resulting in a symmetry number 2 as it should be, and so on.

By standard Fourier transform methods we obtain

$$\frac{p^{(1)}}{\kappa T} = \left(1 - \rho\frac{\partial}{\partial\rho}\right)\sum_{n=1}^{\infty}\frac{1}{2n}\left(\frac{\gamma}{2\pi}\right)^3\int d\mathbf{k}[\tilde{v}(k)\tilde{\mu}_2^{hc}(k)]^n$$

$$= \frac{1}{2}\left(\frac{\gamma}{2\pi}\right)^3\left(\rho\frac{\partial}{\partial\rho} - 1\right)\int d\mathbf{k}\ln[1 - \tilde{v}(k)\tilde{\mu}_2^{hc}(k)]. \quad (60)$$

Expanding the hard-core function according to Eq. (29), we obtain the expected result (55).

7. CONCLUDING REMARKS

While the pair correlation function $g(r)$ in the far range [Eqs. (42) and (49)] is a complicated function of the distance, considerable simplifications are obtained in two limiting cases.

For *low densities*, $\eta \simeq \rho/\kappa T$ is very small, and from (42) and (49) we obtain to $O(\rho^2)$

$$g(r) \simeq -\rho^2\Phi^{attr}(r)/\kappa T + \tfrac{1}{2}\rho^2[\Phi^{attr}(r)/\kappa T]^2, \quad (61)$$

in agreement with the first terms in the density expansion

$$g(r) = \rho^2 f(r) = \rho^2\sum_{n=1}^{\infty}(-\Phi(r)/\kappa T)^n/n!. \quad (62)$$

Near *the critical point*, $\partial p/\partial\rho$ is very small. To zeroth order in ξ, Eq. (54), together with the definition (29), gives

$$\frac{1}{\rho}\frac{\partial p^{(0)}}{\partial\rho} = \frac{1 + \eta\tilde{\Phi}(0)}{\eta}, \quad (63)$$

and we will now restrict ourselves to those densities and temperatures where $\partial p^{(0)}/\partial\rho$ is small,

$$\frac{\partial p^{(0)}}{\partial\rho}\Big/\frac{\partial p^{hc}}{\partial\rho} = \frac{\eta}{\rho}\frac{\partial p^{(0)}}{\partial\rho} = 1 + \eta\tilde{\Phi}(0) = \epsilon^2 \ll 1. \quad (64)$$

For such large values of the compressibility, the Fourier transform (42), (49) of the correlation function has two poles near $k = 0$, approximately given by

$$1 + \eta\tilde{\Phi}(k) \simeq \epsilon^2 + \tfrac{1}{2}\eta\tilde{\Phi}''(0)k^2 \simeq \epsilon^2 - k^2\tilde{\Phi}''(0)/2\tilde{\Phi}(0) = 0. \quad (65)$$

The asymptotic behavior of $g(r)$ for very large r will be determined by this pair of poles. The convenient definition of the range $1/\gamma$ of the attractive force, viz.

$$\frac{1}{\gamma^2} = \frac{\int r^2\Phi^{attr}(r)\,d\mathbf{r}}{\int\Phi^{attr}(r)\,d\mathbf{r}} \equiv -\frac{3\tilde{\Phi}''(0)}{\gamma^2\tilde{\Phi}(0)}, \quad (66)$$

will give the results a more transparent form. Using this in Eq. (65) we find that the poles are located at $k = \pm i(6)^{\frac{1}{2}}\epsilon$, approximately.

To lowest order in ξ and ϵ, Eq. (42) yields

$$\tilde{g}^{(1)}(k) \simeq \frac{6\kappa T}{|\tilde{\Phi}(0)|}\frac{1}{6\epsilon^2 + k^2}, \quad (67)$$

or

$$g^{(1)}(r) \simeq \frac{\kappa T}{4\pi}\frac{6\gamma^2}{|\tilde{\Phi}(0)|}\frac{e^{-\epsilon(6)^{\frac{1}{2}}\gamma r}}{r}. \quad (68)$$

This asymptotic form is precisely that given by the Ornstein–Zernike theory.[9,6,10] The range of the correlation is now much larger than the range of the force, by a factor ϵ^{-1}.

The result (68) is, of course, only the leading term of an expansion in the two parameters ξ and ϵ. The next terms can be determined from (42) and (49). A straightforward calculation yields the following correction terms (the hard-core functions must be taken at critical density and temperature):

$$\tilde{g}^{(1)}(k) \simeq (67) + \frac{\gamma^2}{6^3}\int r^2 g^{hc}(r)\,d\mathbf{r}\cdot\frac{k^2}{(k^2 + 6\epsilon^2)^2}, \quad (69)$$

$$\tilde{g}^{(2)}(k) \simeq \left(9\kappa T\frac{d\eta}{d\rho}\right)^2\frac{2\arctan[k/2(6^{\frac{1}{2}})\epsilon]}{\pi k(k^2 + 6\epsilon^2)^2}\gamma^3. \quad (70)$$

From these equations, one sees that the dominating terms depend upon the expansion parameters ξ and ϵ as follows:

$$g^{(1)}(r) = \xi\epsilon f_1(\gamma\epsilon r); \qquad g^{(2)}(r) = \xi^2\epsilon^{-2}f_2(\gamma\epsilon r). \quad (71)$$

Since $g^{(2)}/g^{(1)} = O(\xi/\epsilon^3)$, the expansion used in this article is certainly only meaningful when the compressibility is not too large, i.e.,

$$\epsilon \gg \xi^{\frac{1}{3}}. \tag{72}$$

What happens in the small region where (72) is violated is outside the scope of this article. In one dimension, this region corresponds to the small *critical region*,[7] in which the transition between the one-phase region and the two-phase region occurs rapidly, but not discontinuously. The mere occurrence of the limitation (72) in three dimensions is perhaps an indication that the existence of a critical region, rather than a critical point, is a general feature.

It should be emphasized that we have tacitly *assumed* convergence without investigating the conditions for this. The range of validity of the expansions used is therefore unknown. One may, however, optimistically hope that the results are valid in *the whole one-phase region*, because it is so in one dimension.[6]

The results obtained have, of course, experimental consequences.[19] In view of the assumed potential form, the application to the inert gases seems especially appropiate. For these the cubed ratio (2) of the ranges of the repulsive and the attractive potential is of the order of 0.1, which should make a usable expansion parameter.

In a comparison with experimental data, the equation of state of a hard-sphere gas enters the formulas and is therefore needed. One has to be satisfied with the five-term virial series[20] or the machine calculations[21] for p^{hc}. In this connection, it should be mentioned that "short-range repulsion" clearly can be substituted for "hard core" in all the present results.

ACKNOWLEDGMENT

I wish to express my sincere appreciation to Dr. G. E. Uhlenbeck and Dr. M. Kac for several helpful discussions.

[19] Near the critical point, x-ray scattering experiments which determine $\tilde{g}(k)$ show deviations from the Ornstein–Zernike theory. While this theory predicts $1/\tilde{g}(k)$ to vary linearly with k^2 [see Eq. (67)], the plot of the experimental values for $1/\tilde{g}(k)$ vs k^2 shows a straight line for large values of k^2 bending down for k^2 small (negative second derivative). A fuller discussion is given in Part III of this series,[7] where it is also shown that the pair correlation function for the previously mentioned, one-dimensional model is in qualitative agreement with these experimental deviations. In the interesting region where ξ is very small, but ξ/ϵ^3 cannot be neglected compared to unity, one may deduce from Eqs. (69) and (70) that, in three dimensions, the qualitative agreement between the theoretical and experimental corrections to Ornstein and Zernike's result persists.

[20] J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids* (John Wiley & Sons, Inc., New York, 1954), p. 157.

[21] W. W. Wood and J. D. Jacobson, J. Chem. Phys. **27**, 1207 (1957). T. E. Wainright and B. J. Alder, J. Chem. Phys. **27**, 1209 (1957); **31**, 459 (1959).

Generalized Master Equations for the Density Matrix*

ROBERT L. PETERSON† AND PAUL M. QUAY, S. J.‡
Case Institute of Technology, Cleveland, Ohio
(Received 14 August 1963)

Generalized master equations determining the time dependence of both diagonal and off-diagonal density matrix elements are derived. The equations are valid for systems of arbitrary size and initial conditions, and have all the generality of the von Neumann equation for the density matrix. They are general enough for use in determining all higher-order contributions to transport coefficients. Their advantage over the von Neumann equation is that they are in forms suited to reduction by time-independent many-body approximation techniques.

1. INTRODUCTION

GENERALIZED master equations have been derived by Van Hove[1] and Janner[2] for systems with an infinite number of degrees of freedom, making use of perturbation theory and the diagonal singularity properties[1] which a large class of perturbations possess. These equations determine the time dependence of the quantities $[U^\dagger(t)]_{ij}[U(t)]_{jk}$ for all i, j, and k. [$U(t) = \exp(-iHt/\hbar)$ is the time-evolution operator, and H is the complete Hamiltonian. Subscript indices denote matrix elements in an unperturbed-energy representation.] Swenson[3] has shown that equations formally identical to those of Van Hove and Janner can be derived for finite systems, without recourse to perturbation theory, and without any special assumptions about the properties of the perturbation.

These generalized master equations (GME's) may be considered as alternate ways of writing the Heisenberg equation of motion for diagonal operators, in forms more suited to application to the nonequilibrium behavior of many-body systems. That they are useful and tractable has already been demonstrated.[4,5]

Other investigators (for example, Zwanzig,[6] and Prigogine and collaborators[7]) have also derived very general rate equations, usually with special initial conditions. Fujita[8] has discussed the connection between the Van Hove and Prigogine equations.

In this paper we derive, in the manner of Swenson, GME's from which the time dependence of the four-index quantity $[U^\dagger(t)]_{ij}[U(t)]_{kn}$ and the density matrix are obtained. The four-index quantity generally is necessary for a complete description of nonequilibrium phenomena. For example, the formal solution to the von Neumann equation

$$d\rho(t)/dt = (i/\hbar)[\rho(t), H] \tag{1}$$

is

$$\rho(t) = U(t)\rho(0)U^\dagger(t). \tag{2}$$

$\rho(t)$ is the density matrix. In terms of matrix elements, Eq. (2) is

$$\rho_{kj}(t) = \sum_{i,n} U^\dagger_{ij}U_{kn}\rho_{ni}(0). \tag{3}$$

The diagonal elements $\rho_{ii}(t)$ are thus determined by the $U^\dagger_{ij}U_{in}$, which were considered by Van Hove, Janner, and Swenson. However, the off-diagonal elements require the four-index quantity unless $\rho(0)$ is taken to be diagonal, an assumption which, in the general case, is not possible.[9] The evaluation of higher-order contributions to transport coefficients likewise requires the use of the four-index quantity.[10]

Our results are simply rewritings of the von Neumann equation in forms more suited to reduction by time-independent many-body approximation

* Supported by the U. S. Atomic Energy Commission and the D. C. Miller Research Fund.
† Present Address: The National Bureau of Standards, Boulder, Colorado.
‡ Present Address: Seminaire des Missions, St.-Martin-d'Ablois (Marne), France.
[1] L. Van Hove, Physica **23**, 441 (1957); also his lecture in *La théorie des gaz neutres et ionisés*, edited by C. DeWitt and J. G. Detoeuf (John Wiley & Sons, Inc., New York, 1959).
[2] A. Janner, Helv. Phys. Acta **35**, 1 (1962).
[3] R. J. Swenson, J. Math. Phys. **3**, 1017 (1962).
[4] J. A. McLennan, Jr. and R. J. Swenson, J. Math. Phys. **4**, 1527 (1963).
[5] L. Van Hove and E. Verboven, Physica **27**, 418 (1961); A. Janner, L. Van Hove, and E. Verboven, Physica **28**, 1341 (1962).
[6] R. Zwanzig, J. Chem. Phys. **33**, 1338 (1960).
[7] I. Prigogine and P. Résibois, Physica **27**, 629 (1961), and earlier papers by the Prigogine group.
[8] S. Fujita, Physica **28**, 281 (1962).
[9] R. L. Peterson, Phys. Rev. **130**, 612 (1963).
[10] Transport coefficients can be expressed in terms of correlation functions of the form tr $\{e^{-\beta H}S^\mu U\dagger(t)S^\nu U(t)\}$, where S^μ and S^ν are components of flux operators for energy, electric charge, etc. If these flux operators are not diagonal, which will be the case if (for example) potential energy of interaction is included in the energy flux, the four-index quantity $U_{ij}\dagger U_{kn}$ must be used. [The correlation-function formulation of transport theory has been developed by several authors, notably, M. S. Green, J. Chem. Phys. **20**, 1281 (1952), **22**, 398 (1954); R. Kubo, J. Phys. Soc. Japan **12**, 570 (1957); and J. A. McLennan, Jr., Phys. Rev. **115**, 1405 (1959), Phys. Fluids **3**, 493 (1960), and Phys. Fluids **4**, 1319 (1961).]

techniques. They are exact and are valid for systems of arbitrary size and arbitrary initial conditions.

It is evident by our derivation that more than one GME can be obtained by means of various definitions of a certain auxiliary quantity. We present three such equations. The equation given in Sec. 2 has the simplest form. However, when specialized to the two- or three-index cases, it does not become identical to the equations given by Swenson. In Sec. 3, we derive two other GME's, one of which does so reduce. Some points of comparison among the equations are also discussed.

In the Appendix we reproduce in our notation some of the relations derived previously by Van Hove, Janner, and Swenson.

2. DERIVATION

The time-evolution operator $U(t)$ may be expressed as

$$U(t) = -(2\pi i)^{-1} \int_C dz\, R(z) \exp(-izt/\hbar), \tag{4}$$

where $R(z)$ is the resolvent operator

$$R(z) = (H - z)^{-1}. \tag{5}$$

The contour C encloses all the poles of the resolvent. Hence,

$$U^{\dagger}_{ij}U_{kn} = -(2\pi)^{-2} \iint dz_1\, dz_2\, \mathfrak{R}^{12}_{ijkn} \times \exp[i(z_1 - z_2)t/\hbar], \tag{6}$$

where

$$\mathfrak{R}^{12}_{ijkn} \equiv R^1_{ij}R^2_{kn} \equiv R_{ij}(z_1)R_{kn}(z_2). \tag{7}$$

We shall use superscripts 1 or 2 to denote the arguments z_1 or z_2.

The Hamiltonian is written as the sum of a part H_0 whose eigenvalues can be found, and a remaining part H_1. In the basis representation of the complete set of eigenstates of H_0, the resolvent is written as the sum of its diagonal part D, and nondiagonal part DND,

$$R = D + DND. \tag{8}$$

Then D can be written (see Appendix)

$$D = [H_0 + G(z) - z]^{-1}, \tag{9}$$

where $G(z)$ is diagonal. Let d^{12}, g^{12}, and $\mathfrak{D}^{12}$ be diagonal operators whose diagonal elements are

$$d^{12}_k = D^1_k - D^2_k,\ g^{12}_k = G^1_k - G^2_k,\ \mathfrak{D}^{12}_k = D^1_kD^2_k. \tag{10}$$

Let $\mathfrak{N}^{12}$ be the operator whose elements are

$$\mathfrak{N}^{12}_{ij} = N^1_{ji}N^2_{ij}. \tag{11}$$

One can show (see Appendix) that

$$d^{12}_k = (z_1 - z_2 - g^{12}_k)\mathfrak{D}^{12}_k, \tag{12}$$

and

$$g^{12}_k = -\sum_m d^{12}_m W^{12}_{km}. \tag{13}$$

W^{12} is defined below [Eq. (17)].

The crux of the derivation consists in establishing a suitable identity for $\mathfrak{R}^{12}_{ijkn}$. We write, using Eqs. (7) and (8),

$$\mathfrak{R}^{12}_{ijkn} = \mathfrak{D}^{12}_k Q^{12}_{ijkn}, \tag{14}$$

where

$$Q^{12}_{ijkn} = (D^1_i/D^1_k)[\delta_{ij}\delta_{kn} + D^1_iN^1_{ij}\delta_{kn} + D^2_nN^2_{kn}\delta_{ij} + D^1_iD^2_nN^1_{ij}N^2_{kn}]. \tag{15}$$

It is convenient to introduce an operator notation, which brings out the simplicity of the derivation, and to suppress the superscripts. From here through Eq. (22), every operator is understood to have superscript 12. Let $\mathfrak{R}_{ij}$, Q_{ij}, and V_{ij} be operators whose kn matrix elements are $\mathfrak{R}_{ijkn}$, Q_{ijkn}, and V_{ijkn}, respectively.

Equation (14) becomes

$$\mathfrak{R}_{ij} = \mathfrak{D}Q_{ij}. \tag{16}$$

Define W and V_{ij} by

$$\mathfrak{N} = W(I + \mathfrak{D}\mathfrak{N}), \tag{17}$$

$$Q_{ij} = (I + \mathfrak{N}\mathfrak{D})V_{ij}. \tag{18}$$

The relation $\mathfrak{N}\mathfrak{D}V_{ij} = W\mathfrak{R}_{ij}$ is deduced by multiplying Eq. (18) from the left by $\mathfrak{D}$ and using Eq. (16), then multiplying Eq. (17) from the right by $\mathfrak{D}V_{ij}$. Equation (18) can thus be written[11]

$$Q_{ij} = V_{ij} + W\mathfrak{R}_{ij}. \tag{19}$$

Multiply Eq. (16) from the left by $z_1 - z_2 - g$, and use Eqs. (12) and (19) to obtain

$$(z_1 - z_2)\mathfrak{R}_{ij} = dV_{ij} + dW\mathfrak{R}_{ij} + g\mathfrak{R}_{ij}. \tag{20}$$

This is the desired identity from which the GME is obtained.[12] In terms of matrix elements it may be written, with the use of Eq. (13), as

$$(z_1 - z_2)\mathfrak{R}_{ijkn} = d_kV_{ijkn} + \sum_m [d_kW_{km}\mathfrak{R}_{ijmn} - d_mW_{km}\mathfrak{R}_{ijkn}]. \tag{21}$$

[11] The derivation can be made even shorter by defining V_{ij} by Eq. (19) instead of Eq. (18). However, V_{ij} defined by Eq. (18) is the more general in the sense that Eq. (18) cannot be derived from Eq. (19) unless the existence of the inverses in Eq. (22) is assumed.

[12] The definition of $\mathfrak{N}$ by Eq. (11) is not necessary for establishing Eq. (20). However, Eq. (13), and hence Eq. (21) and subsequent equations, depend upon Eq. (11).

Van Hove and Janner assume from the start the existence of the operator

$$\mathfrak{N}(I + \mathfrak{D}\mathfrak{N})^{-1} = \mathfrak{N} - \mathfrak{N}\mathfrak{D}\mathfrak{N} + \cdots = (I + \mathfrak{N}\mathfrak{D})^{-1}\mathfrak{N} \quad (22)$$

(more exactly, the existence of its equivalent for an infinitely large system with a perturbation having the diagonal singularity property). Swenson does also, though merely for the purpose of obtaining equations formally identical to those of Van Hove and Janner. When operator (22) exists, it is equal to W, by Eq. (17). Then since $\mathfrak{N}_{ij}^{12} = \mathfrak{N}_{ji}^{21}$ by Eq. (11), it follows that $W_{ij}^{12} = W_{ji}^{21}$. In turn, Eq. (13) can be written

$$g_k^{12} = -\sum_m d_m^{12} W_{mk}^{12}. \quad (23)$$

With the definition

$$\tilde{W}_{km}^{12} = i d_k^{12} W_{km}^{12}, \quad (24)$$

Eq. (21) takes the form

$$(z_1 - z_2)\mathfrak{R}_{ijkn}^{12} = d_k^{12} V_{ijkn}^{12} - i \sum_m [\tilde{W}_{km}^{12}\mathfrak{R}_{ijmn}^{12} - \tilde{W}_{mk}^{12}\mathfrak{R}_{ijkn}^{12}], \quad (25)$$

which is quite similar in form to the Van Hove, Janner, and Swenson equations, though expressed in terms of the four-index quantities.

Equation (17) defining W^{12} is the same as Swenson's. However, the three-index V_{ijjn}^{12}, obtained as a special case of Eq. (18), is not the same as that of Swenson. Further, Eq. (25) specialized to $\mathfrak{R}_{ijjn}^{12}$ differs not only through V_{ijjn}^{12} but also by the four-index $\mathfrak{R}_{ijmn}^{12}$ in the second term on the right, rather than the three-index $\mathfrak{R}_{imnn}^{12}$ as in Janner's and Swenson's equations. We shall return to these points in the following section.

Equations (21) or (25) are transformed into time-dependent equations in the usual way. For fuller detail, see reference 3. By introducing new variables and deforming the paths of integration, one can write for $t > 0$

$$U_{ij}^{\dagger} U_{kn} = \int_{-\infty}^{\infty} dE P_E(t; ijkn), \quad (26)$$

where

$$P_E(t; ijkn) \equiv (1/2\pi^2) \times \int_{-\infty}^{\infty} dE' \mathfrak{R}_{ijkn}^{+-} \exp[(E' - i\eta)2it/\hbar], \quad (27)$$

and

$$\mathfrak{R}_{ijkn}^{+-} \equiv R_{ij}(E + E' - i\eta)R_{kn}(E - E' + i\eta). \quad (28)$$

E, E', and η are real and η is positive. Hereafter, the superscript $+$ will refer to the argument $E + (E' - i\eta)$, and the superscript $-$ will refer to the argument $E - (E' - i\eta)$. Defining

$$f_E(t; ijkn) \equiv (i/2\pi^2\hbar) \times \int_{-\infty}^{\infty} dE'\, d_k^{+-} V_{ijkn}^{+-} \exp[E' - i\eta)2it/\hbar], \quad (29)$$

and

$$w_E(t; j, km) \equiv (i/2\pi^2\hbar^2) \times \int_{-\infty}^{\infty} dE'\, d_j^{+-} W_{km}^{+-} \exp[(E' - i\eta)2it/\hbar], \quad (30)$$

one obtains, from Eqs. (21) and (27)

$$dP_E(t; ijkn)/dt = f_E(t; ijkn) + 2\pi \sum_m \int_0^t dt' \times [w_E(t - t'; k, km)P_E(t'; ijmn) - w_E(t - t'; m, km)P_E(t'; ijkn)]. \quad (31)$$

Equation (31) is the desired four-index GME from which $U_{ij}^{\dagger} U_{kn}$ may be obtained. Note that it is derived from Eq. (21) and thus does not depend upon the assumption of the existence of the operator (22). If Eq. (25) were used, the second w_E on the right in Eq. (31) could also be written $w_E(t-t'; m, mk)$.

Equation (31) can be transformed easily into a GME for the density matrix $\rho(t)$. Defining

$$\rho_E(t; kj) \equiv \sum_{i,n} P_E(t; ijkn)\rho_{ni}(0), \quad (32)$$

and using Eqs. (3) and (26), we have

$$\rho_{ki}(t) = \int_{-\infty}^{\infty} dE \rho_E(t; kj). \quad (33)$$

By Eqs. (31) and (32), the GME determining $\rho_E(t; kj)$ is

$$d\rho_E(t; kj)/dt = \rho_E'(t; kj) + 2\pi \sum_m \int_0^t dt' \times [w_E(t - t'; k, km)\rho_E(t'; jm) - w_E(t - t'; m, km)\rho_E(t'; jk)], \quad (34)$$

where

$$\rho_E'(t; kj) = \sum_{i,n} f_E(t; ijkn)\rho_{ni}(0). \quad (35)$$

Note that Eq. (34) expresses both diagonal and off-diagonal elements of $\rho_E(t; kj)$ in terms of the off-diagonal elements $\rho_E(t'; jm)$.

The inhomogeneous term of Eq. (34) describes the unperturbed motion, plus some corrections. To see this, we take the perturbation to vanish, or,

equivalently, suppose that we are working in the total-energy representation. Then since $\mathfrak{N} = 0$, we have $w_E = 0$ and

$$V^{+-}_{ijkn} = (D^+_j/D^+_k)\delta_{ij}\delta_{kn}, \tag{36}$$

by Eqs. (18) and (15). Using Eqs. (29), (33), and (35), and carrying out the integrations on E and E', one finds

$$d\rho_{kj}(t)/dt = (i/\hbar)(\epsilon_j - \epsilon_k)\rho_{kj}(0) \times \exp[i(\epsilon_j - \epsilon_k)t/\hbar], \tag{37}$$

where ϵ_j is the eigenvalue of H_0 in state j. Equation (37) integrates to

$$\rho_{kj}(t) = \rho_{kj}(0) \exp[i(\epsilon_j - \epsilon_k)t/\hbar], \tag{38}$$

which is just the unperturbed motion.

3. ALTERNATE GENERALIZED MASTER EQUATIONS: DISCUSSION

Several of the equations of the preceding section are not symmetric in that the third index is singled out for summation. A more symmetric set of equations is easily constructed and yields a different GME for the density matrix.

In addition to Eq. (14), we may write [suppressing the superscript 12 through Eq. (43)]

$$\mathfrak{R}_{ijkn} = \mathfrak{D}_i\hat{Q}_{ijkn}, \tag{39}$$

where

$$\hat{Q}_{ijkn} = (\mathfrak{D}_k/\mathfrak{D}_j)Q_{ijkn}. \tag{40}$$

We define V as in Eq. (18) and also a $\hat{V}$:

$$Q_{ijkn} = \sum_m (I + \mathfrak{N}\mathfrak{D})_{km}V_{ijmn}, \tag{18}$$

$$\hat{Q}_{ijkn} = \sum_m (I + \mathfrak{N}\mathfrak{D})_{jm}\hat{V}_{imkn}. \tag{41}$$

From these we deduce

$$Q_{ijkn} = V_{ijkn} + \sum_m W_{km}\mathfrak{R}_{ijmn}, \tag{19}$$

$$\hat{Q}_{ijkn} = \hat{V}_{ijkn} + \sum_m W_{jm}\mathfrak{R}_{imkn}. \tag{42}$$

Multiply Eq. (14) by $z_1 - z_2 - g_k$, Eq. (39) by $z_1 - z_2 - g_j$, and add to get

$$2(z_1 - z_2)\mathfrak{R}_{ijkn} = d_kV_{ijkn} + d_j\hat{V}_{ijkn} + \sum_m \{d_kW_{km}\mathfrak{R}_{ijmn} + d_jW_{jm}\mathfrak{R}_{imkn} - (d_mW_{km} + d_mW_{jm})\mathfrak{R}_{ijkn}\}. \tag{43}$$

Equation (43) is to be compared to Eq. (21). The corresponding GME for the "spectral component" $\rho_E(t; kj)$ of the density matrix is

$$d\rho_E(t; kj)/dt = \rho'_E(t; kj) + \pi \sum_m \int_0^t dt' \times \{w_E(t - t'; k, km)\rho_E(t'; mj) + w_E(t - t'; j, jm)\rho_E(t'; km) - [w_E(t - t'; m, km) + w_E(t - t'; m, jm)]\rho_E(t'; kj)\}, \tag{44}$$

where

$$\rho'_E(t; kj) = \sum_{i,n} \rho_{ni}(0)(i/4\pi^2\hbar) \int_{-\infty}^{\infty} dE'(d^{+-}_k V^{+-}_{ijkn} + d^{+-}_i \hat{V}^{+-}_{ijkn}) \exp[(E' - i\eta)2it/\hbar]. \tag{45}$$

It is also straightforward to construct a four-index equation which reduces to the Van Hove, Janner, or Swenson equations. For example, one may use Eq. (14) and a V' defined by

$$V'_{ijkn} = Q_{ijkn} - \tfrac{1}{2}\sum_m (W_{km} + W_{jm})\mathfrak{R}_{immn}, \tag{46}$$

to obtain

$$(z_1 - z_2)\mathfrak{R}_{ijkn} = d_kV'_{ijkn} + \sum_m [\tfrac{1}{2}d_k(W_{km} + W_{jm})\mathfrak{R}_{immn} - d_mW_{km}\mathfrak{R}_{ijkn}]. \tag{47}$$

[The superscript 12 is suppressed in Eqs. (46) and (47).] The resulting GME for $P_E(t; ijkn)$ is

$$dP_E(t; ijkn)/dt = \hat{f}_E(t; ijkn) + 2\pi \sum_m \int_0^t dt' \{\tfrac{1}{2}[w_E(t - t'; k, km) + w_E(t - t'; k, jm)]P_E(t; immn) - w_E(t - t'; m, km)P_E(t', ijkn)\}, \tag{48}$$

with

$$\hat{f}_E(t; ijkn) = (i/2\pi^2\hbar) \int_{-\infty}^{\infty} dE'\, d^{+-}_k V'^{+-}_{ijkn} \times \exp[(E' - i\eta)2it/\hbar]. \tag{49}$$

Equations (46)–(49) are identical to those of Swenson for $k = j$. The corresponding GME for the density matrix spectral component is

$$d\rho_E(t; kj)/dt = \hat{\rho}_E(t; kj) + 2\pi \sum_m \int_0^t dt' \{\tfrac{1}{2}[w_E(t - t'; k, km) + w_E(t - t'; k, jm)]\rho_E(t'; mm) - w_E(t - t'; m, km)\rho_E(t'; kj)\}, \tag{50}$$

with

$$\hat{\rho}_E(t; kj) = \sum_{i,n} \hat{f}_E(t; ijkn)\rho_{ni}(0). \tag{51}$$

Other GME's can also be derived. Which one is the most useful will depend upon the context in which it is to be used. We may note, however, that Eq. (50) appears to have an advantage over Eqs. (34) or (44), in that both diagonal and off-diagonal elements of $\rho_E(t; kj)$ are expressed in terms of the diagonal elements $\rho_E(t'; mm)$.

APPENDIX

Since $R = (H - z)^{-1}$, its diagonal part D clearly will have the form $[H_0 + G(z) - z]^{-1}$, where $G(z)$ is diagonal. $G(z)$ may be found by writing Eq. (8) as

$$(D + DND)(H_0 + H_1 - z) = I, \tag{A1}$$

and taking the diagonal part:

$$D[(H_0 + H_1 - z)_d + (NDH_1)_d] = I. \tag{A2}$$

The subscript d means "diagonal part of". By Eq. (A2) then,

$$G(z) = (H_1)_d + (NDH_1)_d. \tag{A3}$$

The nondiagonal (nd) part of Eq. (A1) gives, with $V \equiv (H_1)_{nd}$,

$$N = -V - (NDV)_{nd} + [ND(NDV)_d]_{nd}. \tag{A4}$$

Here we have used the expression $D(H_0 - z) = I - DG$, and the form (A3) for G. If desired, Eq. (A4) may be iterated to give a series expansion in V. This series is *not* irreducible; that is, intermediate states *are* occasionally equal to the initial or final states, or to each other. It follows that an expansion of W^{12} [Eq. (22)] is not expressible as an irreducible product of two expansions in V, contrary to Swenson's assertion [his Eq. (40)], except for an infinite system.

Since $(R^2)^{-1} - (R^1)^{-1} = z_1 - z_2$, one can write

$$R^1 - R^2 = (z_1 - z_2)R^2R^1, \tag{A5}$$

the diagonal part of which gives

$$d_k^{12} = (z_1 - z_2) \sum_i \mathfrak{R}_{ikki}^{12}. \tag{A6}$$

Similarly,

$$D^1 - D^2 = (z_1 - z_2 - G^1 + G^2)D^1D^2, \tag{A7}$$

the diagonal part of which gives

$$d_k^{12} = (z_1 - z_2 - g_k^{12})\mathfrak{D}_k^{12}. \tag{A8}$$

To establish Eq. (13), one sets $n = i$ and $j = k$ in Eq. (14), multiplies by $z_1 - z_2 - g_k^{12}$, and sums on i, using Eqs. (A6) and (A8) to obtain

$$d_k^{12} - g_k^{12} \sum_i \mathfrak{R}_{ikki}^{12} = d_k^{12} \sum_i Q_{ikki}^{12}. \tag{A9}$$

Using Eqs. (15), (17), and (A6), one sees that Eq. (A9) reduces to Eq. (13).

ACKNOWLEDGMENTS

One of us (PMQ) wishes to take this opportunity to thank the Case Department of Physics for their warm hospitality during the year of his stay with them and for making available to him such ideal conditions for his research.

Applicability of the Pfaffian Method to Combinatorial Problems on a Lattice

C. A. HURST
University of Adelaide, Adelaide, South Australia
(Received 10 April 1963)

For the solution of certain problems in statistical mechanics, a formulation in terms of the combinatorial problem of counting open or closed polygons drawn on a lattice has been found useful. The solution of such combinatorial problems has been given as a generating function which can be expressed as a Pfaffian. Necessary and sufficient conditions are given here to determine when such a Pfaffian representation is possible. If certain consistency conditions are satisfied, then the solution to the combinatorial problem can be written down immediately. The method makes some use of the ideas of emission and absorption operators for fermions.

1. INTRODUCTION

IN a number of recent papers, a new mathematical method has been found useful for solving some combinatorial problems occurring in statistical mechanics. This mathematical method uses the concept of a Pfaffian, which is defined as the square root of an antisymmetric determinant of even order. Such quantities were first used in quantum field theory by Caianiello,[1] and later were shown to give a simple solution of the Ising problem for the two-dimensional rectangular lattice by Hurst and Green.[2] Since then, the same technique has been used to solve, for the first time, the problem of the statistical mechanics of dimers on a plane lattice.[3-5]

In all the papers on statistical mechanics quoted, the partition function of the system was expressed as a Pfaffian which was derived by considering an associated combinatorial problem of which the Pfaffian was the appropriate generating function. In order to demonstrate the correctness of the Pfaffian solution so obtained, it was necessary to show first of all that the individual terms of the Pfaffian, when it was expanded according to the rules for evaluating such expressions, counted correctly all the terms of the corresponding combinatorial problem, and secondly that all these terms appeared with the correct sign. Both of these questions were fully discussed in the treatment of the dimer problem, but for the Ising problem only the first one was considered. The question of the correct sign of the individual terms is, in general, a deep one, and is connected with the occurrence of "crossed bonds". This aspect of the Ising problem will not be considered here, as an exhaustive treatment will be given elsewhere;[6] an alternative treatment has already been given by Sherman.[7]

In this paper, a general solution will be given to the question of when can a Pfaffian be used to solve those combinatorial problems which may be represented as the enumeration of polygons drawn on a two-dimensional lattice. Only the counting aspect of this question will be considered here; a few general remarks on the correctness or otherwise of the signs of the individual terms will be made at the end of the paper. These polygons are constructed by joining different lattice points by lines, which will be called *bonds*, and a problem is specified by stating which lattice points may be directly connected by bonds. More than one bond may terminate at a lattice point, and the permitted sets of bonds which may be simultaneously connected to a lattice point specify the structure of the lattice-point connections. It will be shown that if the structure of the lattice-point connections for each lattice point is known, and also the weight factors, if any, associated with each bond, then an explicit expression can be written down for the generating functions of all polygons which can be drawn on the lattice, consistent with the given structure. Such an explicit expression, in the form of a Pfaffian, is however only possible if certain consistency conditions are satisfied. These consistency conditions restrict the variety of possible lattice-point structures which permit a Pfaffian solution. If the conditions are not satisfied, a Pfaffian representation cannot be found. Simple examples will be given illustrating the success and failure of the method. In another

[1] E. R. Caianiello, Nuovo Cimento **10**, 1634 (1953).
[2] C. A. Hurst and H. S. Green, J. Chem. Phys. **33**, 1059 (1960).
[3] H. N. V. Temperley and M. E. Fisher, Phil. Mag. **68**, 1061 (1961).
[4] M. E. Fisher, Phys. Rev. **124**, 1664 (1961).
[5] P. W. Kasteleyn, Physica **27**, 1209 (1961).
[6] H. S. Green and C. A. Hurst, *Order–Disorder Phenomena* (Interscience Publishers, Inc., New York, to be published), Chap. 4.
[7] S. Sherman, J. Math. Phys. **1**, 202 (1960).

paper[8] it will be shown how all the known soluble Ising models can be simply treated by this technique.

Suppose the lattice points are numbered systematically 1, 2, $\cdots$, N, where N is the number of points in the lattice. Then the approach used here will be to represent a bond drawn on the lattice from point j to point k as a "particle" travelling from j to k (or from k to j). Such a "particle" may be imagined as emitted at j and absorbed at k, and such a process is described mathematically by applying an emission operator $a^\dagger$ to a state vector Ω and then subsequently applying the corresponding absorption operators a. To every bond there is a distinct pair of emission and absorption operators. For a bond joining the lattice points j and k there will be an emission operator $a^\dagger$ and an absorption operator a. It is to some extent arbitrary which lattice point is associated with a and which with $a^\dagger$. The convention chosen is that $a^\dagger$ is associated with j and a with k when $j < k$ and conversely. Let there be n_j different types of bonds connected to lattice point j: such bonds may or may not be connected simultaneously. Suppose it is compatible with the lattice-point structure that a set $\nu_j \leq n_j$ of the bonds can be connected simultaneously to j. Such a situation can be represented by a product of $\nu_{j'}$ emission operators and $\nu_{j''}$ absorption operators, with $\nu_j = \nu_{j'} + \nu_{j''}$, applied to Ω.[9] Then the total set of connections to j is described by a sum of operator products, multiplied by appropriate weight factors. Each term in the sum corresponds to a permitted set of connections to j, and omitted terms correspond to forbidden connections. The order of the resulting polynomial $P_j(a)$ will be $\leq n_j$. (We do not consider the case when a bond may be described more than once, as such a situation may be replaced by one in which two or more bonds connecting the same pair of lattice points are regarded as distinct). For example, the Ising model for the two-dimensional rectangular lattice considers polynomials constructed out of horizontal and vertical bonds joining neighboring lattice points only. Each horizontal bond has weight $x = \tanh \beta J$ where J is the interaction energy of neighboring lattice points in the same row, and each vertical bond has weight $y = \tanh \beta J'$ where J' is the interaction energy of neighboring lattice points in the same column. As only an even number of bonds may be connected to any lattice point, and the total number of bonds connected to a lattice point may be 0, 2, or 4, the polynomial $P_j(a)$ in this case has the form

$$P_j(a) = (1 + a_{j-m}^{(2)} a_{j-1}^{(1)} + x a_j^{(1)\dagger} a_{j-1}^{(2)} + x a_j^{(1)\dagger} a_{j-m}^{(2)} + y a_j^{(2)\dagger} a_{j-1}^{(1)} + y a_j^{(2)\dagger} a_{j-m}^{(2)} + xy a_j^{(2)\dagger} a_j^{(1)+} + xy a_j^{(2)\dagger} a_j^{(1)\dagger} a_{j-m}^{(2)} a_{j-1}^{(1)}),$$

which is the form assumed later [Eq. (36)] with the weight factors explicitly inserted here. Here $n_j = 4$, and m is the number of points in a row. By inspection it is evident that $P_j(a)$ describes the only connections compatible with the conditions of this Ising problem. A formal proof of this statement is given elsewhere.[6] Further, it turns out to be essential for this method that the "particles" all be fermions, so that a and $a^\dagger$ satisfy the anticommutation relation

$$\{a, a^\dagger\} = 1. \tag{1}$$

Operators corresponding to different "particles" (bonds) will be assumed to anticommute. For the counting of graphs on a lattice, it does not seem essential to choose between fermions and say, bosons. However the consistency conditions derived later in Secs. 3 and 4 make considerable use of expansion theorems for Pfaffians which would not be valid for the mathematical quantities ("hafnians") associated with bosons. Furthermore, even for the simple case of the rectangular Ising lattice discussed in Sec. 5(a), the consistency condition cannot be satisfied for bosons. In that case, if $c_0 = 3$ and all the other coefficients remained the same, then bosons would have to be used; but then the simple expansion and multiplication theorems for Pfaffians and determinants required for the explicit evaluation of the partition function would not be available. So even if bosons could be used, they would be quite valueless for an explicit solution. Furthermore, the state Ω in many cases may be taken to be the vacuum state of the system, so that it satisfies

$$a\Omega = 0, \tag{2}$$

for all absorption operators a. There will therefore be N operator polynomials, one for each lattice point, and these are to be multiplied together in order 1, $\cdots$, N reading from right to left. If this polynomial is applied to Ω, a new state Ω_G is produced, and the scalar product (Ω, Ω_G) will be a sum of terms each of which corresponds to a *saturated polygon* drawn on the lattice. By "saturated" is meant here that all the bonds drawn terminate

[8] C. A. Hurst, J. Chem. Phys. **38**, 2558 (1963).

[9] Other authors[3-5] have gone directly from the combinatorial problem to the Pfaffian solution without the intermediate step of introducing the emission and absorption operators. However in this author's opinion, the latter point of view is somewhat more perspicuous.

at lattice points in distinction to "unsaturated polygons" defined later. If the number of bonds connected to a lattice point is odd, then such a lattice point will belong to what is usually called an open polygon. To see this, suppose that Πa is a product of emission and absorption operators which is a typical term in the expansion of the above polynomial. Then (Ω, Ω_G) will contain a term $(\Omega, \Pi a \Omega)$. Because of Eq. (2), such a term will be zero unless for every absorption operator appearing in Πa, there is also the corresponding emission operator $a^\dagger$ present. Hence the only terms which contribute to (Ω, Ω_G) will be those in which every bond drawn on the lattice terminates on two points of the lattice. If the lattice-point structure is such that only even-order terms occur in every $P_i(a)$, this means that each lattice point must have an even number of bonds connected to it, and in this case, the set of bonds corresponding to $(\Omega, \Pi a \Omega)$ is a set of closed polygons in the usual sense. If $P_i(a)$ contains odd-order terms, then *unsaturated polygons* may occur. If Ω is replaced by a state Ω_1, which is not a vacuum state but may have one or more particles of various types (bonds) present, then $(\Omega_1, \Pi a \Omega_1)$ will correspond to graphs drawn on the lattice with some lines leading out of the lattice. Even more general graphs can be formed by taking matrix elements between different states: $(\Omega_1, \Pi a \Omega_2)$.

Because of the use of anticommuting fermion operators, the product (Ω, Ω_G) may contain explicit and unwanted negative signs which do not appear in the desired generating function. Such explicit negative signs will invalidate this method, but, as mentioned earlier, they can be shown to be connected with the appearance of "crossed bonds," i.e. bonds which intersect elsewhere than at a lattice point. As many useful types of problems do not lead to such graphs, it is with them that this paper is concerned.

From the foregoing remarks it is clear that the point of view adopted here is a "local" one, in that attention is concentrated on the structure of the connections to a single lattice point. The object is to replace an inhomogeneous operator polynomial $P_i(a)$ by a product of factors, each one of which is homogeneous, of degree unity, in the fermion operators already introduced, together with auxiliary fermion operators called *internal operators*. The original operators are called *external operators*. This process will be called *factorization*.

Apart from the difficulties associated with crossed bonds, there is another basic limitation to which this method is subject. This method is applicable only to problems in which the lattice-point structure admits connections all of the same parity only, i.e., only an even number or only an odd number of bonds may be connected to a lattice point. The parity of connections may vary from point to point, but must be fixed for a given point. This restriction prevents the immediate application of this method to the mixed monomer–dimer problem[10] and to the Ising problem in an external magnetic field.

If the parity of the structure of a lattice point j is fixed, there is a further subdivision into two cases:

(i) the maximum number of bonds which can be simultaneously connected to j is n_j;

(ii) the maximum number $\nu_j < n_j$. ν_j may or may not have the same parity as n_j.

These two cases will be considered in Secs. 3 and 4 of this paper, respectively. Sec. 2 will contain a demonstration of the representation of the vacuum expectation value of a product of linear fermion operators as a Pfaffian.

In Sec. 5, simple examples of the use of this method will be given, and the connection between the failure of the factorization method and the appearance of crossed bonds will be discussed.

Some results on Pfaffians required for these arguments are given in an appendix.

2. REPRESENTATION OF THE VACUUM EXPECTATION VALUE OF A PRODUCT OF LINEAR FERMION OPERATORS AS A PFAFFIAN

In Sec. 1 it was shown that the enumeration of closed polygons drawn on a lattice could be expressed as the problem of evaluating the vacuum expectation value of a product of operator polynomials $[\Omega, \Pi_i P_i(a)\Omega]$. The method proposed for this evaluation is to replace $P_i(a)$ by a product of terms, each a linear homogeneous expression in fermion operators. So in this section, the reduction of such a product to Pfaffian form will be given. Alternative proofs of the Pfaffian representation have been given elsewhere,[4,5] but they are rather less direct than the one to be given here. A linear fermion operator A is defined as a linear combination $\Sigma\ \alpha_i a_i$ of emission and absorption fermion operators a_i.

The main result which is required for this demonstration is Wick's theorem[11] on the expression

[10] A possible solution to this difficulty would be to introduce an external lattice point which can be joined to *any* point of the lattice. Such a procedure, which was essentially used by Fisher (reference 4), is, however, likely to introduce explicit negative signs.

[11] G. C. Wick, Phys. Rev. **80**, 268 (1950).

of a product of linear fermion operators as a series of normal products:

$$A_1A_2\cdots A_r = :A_1A_2\cdots A_r: + \Sigma :A_1^{\cdot}A_2\cdots A_p^{\cdot}\cdots A_r: + \Sigma\Sigma :A_1^{\cdot}A_2^{\cdot}\cdots A_p^{\cdot}\cdots A_q^{\cdot}\cdots A_r: + \cdots . \quad (3)$$

Pairing and normal products for emission and absorption are defined as follows:

$$a_ka_l^{\dagger} = -a_l^{\dagger}a_k + \{a_k, a_l^{\dagger}\} = -:a_ka_l^{\dagger}: + a_k^{\cdot}a_l^{\dagger\cdot},$$
$$a_ka_l = :a_ka_l:, \qquad a_k^{\dagger}a_l^{\dagger} = :a_k^{\dagger}a_l^{\dagger}: . \quad (4)$$

A normal product with pairing $:A_1^{\cdot}A_2\cdots A_i^{\cdot}\cdots A_r:$ is defined by

$$:A_1^{\cdot}A_2\cdots A_i^{\cdot}\cdots A_r: = (-1)^P(A_1^{\cdot}A_i^{\cdot}) :A_2\cdots A_r:, \quad (5)$$

where P is the parity of the permutation required to go from the ordering $1, 2\cdots i\cdots r$ to $1, i, 2\cdots r$. Analogously, we define a normal product with two pairings by

$$:A_1^{\cdot}A_2^{\cdot\cdot}A_3\cdots A_i^{\cdot}\cdots A_j^{\cdot\cdot}\cdots A_r: = (-1)^P(A_1^{\cdot}A_i^{\cdot})(A_2^{\cdot\cdot}A_j^{\cdot\cdot}) :A_3\cdots A_r:, \quad (6)$$

where P is the parity of the permutation required to produce the indicated change in ordering. The summations in (3) are taken over single, double, triple, etc. pairings.

As the vacuum expectation value of a normal product is zero, we have from (3)

$$\begin{aligned}(\Omega, A_1A_2\cdots A_r\Omega) &= (\Omega, \Sigma\cdots\Sigma :A_1^{\cdot}\cdots A_r^{\cdot}: \Omega),\\ &= \Sigma\cdots\Sigma :A_1^{\cdot}\cdots A_r^{\cdot}: (\Omega, \Omega),\\ &= \Sigma\cdots\Sigma(-1)^P(A_1^{\cdot}A_i^{\cdot})(A_2^{\cdot}A_j^{\cdot})\cdots\\ &= \begin{vmatrix} A_1^{\cdot}A_2^{\cdot} & A_1^{\cdot}A_3^{\cdot} & \cdots & A_1^{\cdot}A_r^{\cdot}\\ & A_2^{\cdot}A_3^{\cdot} & & A_2^{\cdot}A_r^{\cdot}\\ & & & A_{r-1}^{\cdot}A_r^{\cdot}\end{vmatrix}, \end{aligned} \quad (7)$$

by the definition of a Pfaffian.[12]

Now a slight generalization of (7) is required for the case where the vacuum expectation value is only partial. This means that the expectation value is taken with respect to some, but not all of the particles in the product (7). We suppose that this product of r operators can be decomposed into s subproducts such that the jth product contains σ_j terms (so that $r = \sum_{j=1}^{s}\sigma_j$). Each subproduct is to be such that if any absorption operator and its corresponding emission operator occurs in it, then this pair of operators does not appear in any other subproduct. Such operators will be called internal operators, and the remaining operators external operators. We relabel A_j as A_{ji} with $1 \le j \le s$, $1 \le i \le \sigma_j$, and write

$$A_{ji} = A_{ji}^{(e)} + A_{ji}^{(i)}. \quad (8)$$

Within each subproduct, the external operators are further assumed to be in normal order. This last assumption is not essential for the results of this section, but it is a natural requirement in the context of the later sections of this paper.

Then if $\Omega^{(i)}$ is the vacuum with respect to the internal operators only, we put

$$\Omega^{(i)} = \prod_{j=1}^{s}{}^{\cdot}\,\Omega_j^{(i)}, \quad (9)$$

where $\Omega_j^{(i)}$ is the vacuum state for the internal operators which appear in the partial product $\prod_{i=1}^{\sigma_j}{}^{\cdot}\,A_{ji}$. Then we have

$$\left[\Omega^{(i)}, \prod_{j=1}^{s}{}^{\cdot}\prod_{i=1}^{\sigma_j}{}^{\cdot}\,A_{ji}\Omega^{(i)}\right] = \prod_{j=1}^{s}{}^{\cdot}\left[\Omega_j^{(i)}, \prod_{i=1}^{\sigma_j}{}^{\cdot}\,A_{ji}\Omega_j^{(i)}\right]. \quad (10)$$

Now

$$\begin{aligned}&\left(\Omega_j^{(i)}\prod_{i=1}^{\sigma_j}{}^{\cdot}\,A_{ji}\Omega_j^{(i)}\right)\\ &= (\Omega_j^{(i)}, (:A_{j1}\cdots A_{j\sigma_j}: + \Sigma :A_{j1}^{\cdot}\cdots A_{j\sigma_j}: + .)\Omega_j^{(i)})\\ &= :A_{j1}^{(e)}\cdots A_{j\sigma_j}^{(e)} + \cdots\Sigma :A_{j1}^{(i)}\cdots A_{j\sigma_j}^{(e)}:\\ &\quad + \Sigma\Sigma :A_{j1}^{(i)\cdot}A_{j2}^{(i)\cdot}\cdots A_{j\sigma_j}^{(e)}: + \cdots, \end{aligned} \quad (11)$$

from (3). Here we have used the condition that all the internal operators must be paired to give a nonzero contribution, and all the external operators have zero pairings because they are in normal order.

Any normal product of fermion operators $:A_1\cdots A_r:$ can be written symbolically as a Pfaffian:

$$:A_1A_2\cdots A_r: = :\begin{vmatrix} A_1A_2 & A_1A_3 & \cdots & A_1A_r\\ & A_2A_3 & \cdots & A_2A_r\\ & & & \cdots\\ & & & A_{r-1}A_r\end{vmatrix}: \quad (12)$$

if the Pfaffian is evaluated *as if the operators commuted*. This is because each term in the expansion of the Pfaffian contains each operator once and

[12] G. Kowaleski, *Einfuhring in die Determinantentheorie* (Chelsea Publishing Company, New York, 1948) 3rd ed., Sec. 61.

once only, and so taking out the common factor $:A_1A_2 \cdots A_r:$, the coefficient is the expansion of a Pfaffian, all of whose elements are unity. Such a Pfaffian has the value unity, as can be easily verified.

Hence each normal product of external operators in (11) can be written as a Pfaffian with an explicit sign given by the appropriate permutation required to remove the paired internal operators. But this is just the expression which is obtained if we evaluate the symbolic expression

$$\left(\Omega_j^{(i)}, \prod_{i=1}^{\sigma_j} A_{ji}\Omega_j^{(i)}\right) = :\begin{vmatrix} A_{j1}^{(i)\cdot}A_{j2}^{(i)\cdot} + A_{j1}^{(e)}A_{j2}^{(e)} & A_{j1}^{(i)\cdot}A_{j3}^{(i)\cdot} + A_{j1}^{(e)}A_{j3}^{(e)} \cdots \\ A_{j2}^{(i)\cdot}A_{j3}^{(i)\cdot} + A_{j2}^{(e)}A_{j3}^{(e)} \cdots \\ \cdots \end{vmatrix}:, \tag{13}$$

with all products evaluated as if the elements commuted among themselves, and the operator products are written in normal form. So (13) is an operator Pfaffian.

3. VALIDITY OF THE FACTORIZATION METHOD WHEN $\nu_j = n_j$.

It has been shown in the last section that the vacuum expectation value of a product of linear fermion operators can be written as a Pfaffian with numerical elements [Eq. (7)]. Once such a representation has been achieved, the further evaluation becomes a matter of evaluating an antisymmetric determinant (obtained by squaring the Pfaffian) and this is generally practicable if the determinant has a simple enough structure. Furthermore, it was shown in the last section that an operator polynomial in external operators, which can be written as an operator Pfaffian (13), is the vacuum expectation value of a product of linear fermion operators. Hence, if each term of a product of polynomials of the type discussed in Sec. 1 can be written as an operator Pfaffian, by introducing appropriate internal operators, $[\Omega, \Pi_j P_j(a)\Omega]$ can be written as a vacuum expectation value of a product of linear fermion operators and hence as a Pfaffian. So the question is to find what operator polynomials can be written in the form (13).

The process of factorization can be represented graphically by replacing each lattice point to which more than one bond is connected by a set of terminals, to each of which may be connected one and only one bond. Such a bond may be connected to a terminal of the same or another lattice point. Bonds joining terminals to the same lattice point are called *internal bonds*, and the corresponding operators internal operators, while bonds connecting terminals of different lattice points are *external bonds*, and the operators external operators. This accords with the usage of the previous section.

It is clear that this process of replacing a single lattice point, to which may be connected various numbers of bonds, by a cluster of terminals each of which has a fixed number of bonds (one) connected to it, is possible only if the parities of the various numbers of bonds which may be connected to the lattice point are the same in all circumstances, e.g., if only an even (including zero) number of bonds can be connected to it. The procedure of taking the vacuum expectation value of the internal operators will pair off internal terminals, and so each such pairing decreases the number of external operators by two. This is the reason for the failure of this method to apply directly to the monomer–dimer problem mentioned in Sec. 1.

If μ_j is the number of terminals ($\mu_j = \sigma_j$), the number of linear fermion factors, n_j the number of external operators, and ν_j the maximum number which can be simultaneously connected, then, from the foregoing remarks it is clear that μ_j and ν_j must have the same parity, but n_j can have a different parity. Hence there are four possibilities:

(i) μ_j, ν_j even n_j even,

(ii) μ_j, ν_j even n_j odd,

(iii) μ_j, ν_j odd n_j even,

(iv) μ_j, ν_j odd n_j odd.

In cases (i) and (iv), $\nu_j \leq n_j$, but in (ii) and (iii), $\nu_j < n_j$. In this section we will consider the case $\nu_j = n_j$, so that (ii) and (iii) are excluded. For simplicity, only case (i) will be considered in detail.

A typical lattice-point polynomial will be

$$P(a) = \sum_{r=0}^{\nu} \sum_{c_r} c_{i_1 i_2 \cdots i_r} a_{i_1} a_{i_2} \cdots a_{i_r}. \tag{14}$$

The second summation is over all permitted selections of r of the external operators $a_1, a_2, \cdots, a_n$ (emission or absorption), and only even-order terms appear. From now on the subscript j referring to the particular lattice point will be dropped. Each

operator product is in normal order, and if an emission (absorption) operator occurs in $P(a)$, the corresponding absorption (emission) operator does not. So the problem is now: given the coefficients $c_{i_1 \cdots i_\nu}$, write (14) in the form

$$P(a) = \left[\Omega^{(i)} \prod_{i=1}^{\nu} \left(\sum_{i'=1}^{\nu} b_{ii'} + x_i a_i \right) \Omega^{(i)} \right], \quad (15)$$

where a_i is an external operator, x_i is a numerical weight factor, $b_{ii'}$ is an internal operator, emission or absorption, and $\Omega^{(i)}$ is the vacuum state with respect to $b_{ii'}$, i.e.,

$$b_{ii'} \Omega^{(i)} = 0, \quad (16)$$

if $b_{ii'}$ is an absorption operator.

The operator $a_{i'}$ could be a linear fermion operator, as in the dimer problem. Such an operator corresponds to the case in which the appearance of one type of external bond excludes the appearance of another type. However such a linear fermion operator can be replaced by a product of factors with each factor containing only a single external emission or absorption operator. This can be done as follows:

$$a_1 + a_2 = [\Omega^{(i)}, (a_2 + c_{13})(a_2 + c_{23})(c_{13}^\dagger + c_{23}^\dagger)\Omega^{(i)}],$$

with

$$c_{13}^{\cdot} c_{13}^{\dagger \cdot} = -1, \qquad c_{23}^{\cdot} c_{23}^{\dagger \cdot} = 1,$$

$$a_1 + a_2 + a_3 = [\Omega^{(i)}, (a_1 + c_{13})(a_3 + c_{23} + c_{25}) \times (c_{13}^\dagger + c_{23}^\dagger)(a_2 + c_{45})(c_{25}^\dagger + c_{45}^\dagger)\Omega^{(i)}],$$

with

$$c_{13}^{\cdot} c_{13}^{\dagger \cdot} = c_{25}^{\cdot} c_{25}^{\dagger \cdot} = 1, \qquad c_{45}^{\cdot} c_{45}^{\dagger \cdot} = -1 = c_{23}^{\cdot} c_{23}^{\dagger \cdot}.$$

If we absorb the numerical factors x_i into the operator a_i, then, using (13), (15) can be written

$$P(a) = :\!\left| \Delta_{ii'} + a_i a_{i'} \right|\!:, \quad (17)$$

with

$$\Delta_{ii'} = b_{ii'}^{\cdot} b_{ii'}^{\dagger \cdot} (b_{ii'}^\dagger = b_{ii'} \quad \text{for} \quad i < i').$$

As $P(a)$ can be multiplied by a numerical factor without altering the nature of the problem, $c_{12\cdots\nu}$ can be chosen to be unity, and then comparing coefficients of the highest-order terms in (14) and (15) [or (17)], we see that they agree.

Now we compare coefficients of terms in (14) and (17) of order $\nu - 2$. We find

$$c_{i_1 i_2 \cdots i_{\nu-2}} = (-1)^{i_{\nu-1} + i_\nu - 1} \Delta_{i_{\nu-1} i_\nu}, \quad (18)$$

for each of the $\binom{\nu}{2}$ coefficients. Such a coefficient is obtained from (17) by striking out the rows and columns which contain the indices $i_1, \cdots i_{\nu-2}$ with $i_1 < i_2 < \cdots < i_{\nu-2}$ and $i_{\nu-1} < i_\nu$. The factor $(-1)^{i_{\nu-1} + i_\nu - 1}$ comes from the antisymmetric character of a Pfaffian.

So for $n = \nu$, the elements $\Delta_{ii'}$ of $P(a)$ are completely determined by the coefficients of order $\nu - 2$. Hence if (14) can be factorized, it is necessary that all the remaining coefficients of order $< \nu - 2$ satisfy certain relationships. This is a consequence of the fact that there are $2^{\nu-1} - 1$ coefficients in (14) but only $\binom{\nu}{2}$ independent elements $\Delta_{ii'}$ in (17). Conversely, if the lower-order coefficients bear the correct relation to the coefficients of order $\nu - 2$, then (14) can be written as a Pfaffian and accordingly be factorized.

For example, the coefficient $c_{i_1 i_2 \cdots i_{\nu-4}} (i_1 < i_2 < \cdots < i_{\nu-4})$ of a term of order $\nu - 4$ in (14) is obtained from (17) by striking out the rows and columns containing these indices, and will be a Pfaffian of order 4, whose elements are Δ_{st}, with $s < t$ and s, t selected from $i_{\nu-3}, i_{\nu-2}, i_{\nu-1}$, and i_ν. (The order of a Pfaffian will be taken as the order of the antisymmetric determinant of which it is the square root). If we substitute from (18) in this Pfaffian we find that

$$c_{i_1 i_2 \cdots i_{\nu-4}} = \begin{vmatrix} c_{i_1 i_2 \cdots i_{\nu-4} i_{\nu-3} i_{\nu-2}} & c_{i_1 i_2 \cdots i_{\nu-3} i_{\nu-1}} & c_{i_1 \cdots i_{\nu-3} i_\nu} \\ & c_{i_1 \cdots i_{\nu-2} i_{\nu-1}} & c_{i_1 \cdots i_{\nu-2} i_\nu} \\ & & c_{i_1 \cdots i_{\nu-1} i_\nu} \end{vmatrix}, \quad (19)$$

for all selections of $i_1 < i_2 < \cdots < i_{\nu-4}$. Similarly, by considering terms of order $\nu - 6$, we have

$$c_{i_1 i_2 \cdots i_{\nu-6}} = \begin{vmatrix} c_{i_1 \cdots i_{\nu-6} i_{\nu-5} i_{\nu-4}} & c_{i_1 \cdots i_{\nu-6} i_{\nu-5} i_{\nu-3}} & \cdots & c_{i_1 \cdots i_{\nu-6} i_{\nu-5} i_\nu} \\ & c_{i_1 \cdots i_{\nu-6} i_{\nu-4} i_{\nu-3}} & \cdots & c_{i_1 \cdots i_{\nu-6} i_{\nu-4} i_\nu} \\ & & & \cdots \\ & & & c_{i_1 \cdots i_{\nu-5} i_{\nu-4} i_\nu} \end{vmatrix}, \quad (20)$$

and so on.

So far we have considered the case $n = \nu = \mu$. Now we will consider the case where $n = \nu = \mu - 2$. This means that we replace a lattice point by a

cluster of terminals with two of the terminals dummies, such that no external bond emanates from them. In this case,

$$P(a) = :\begin{vmatrix} \Delta_{\mu-1\mu} & \Delta_{\mu-11} & \cdots & \cdots & \Delta_{\mu-1\nu} \\ & \Delta_{\mu 1} & \cdots & \cdots & \Delta_{\mu\nu} \\ & & \Delta_{12}+a_1a_2 & \cdots & \Delta_{1\nu}+a_1a_\nu \\ & & & & \vdots \\ & & & & \Delta_{\nu-1\nu}+a_{\nu-1}a_\nu \end{vmatrix} : . \tag{21}$$

The external operators $a_1, \cdots, a_{\mu-2} \equiv a_\nu$ now occur only in rows 3, $\cdots$, μ, while the first two rows contain numerical elements which correspond to internal bonds connected to one or other of the dummy terminals.

The coefficient of the term of order ν in $P(a)$ is $\Delta_{\mu-1\mu}$ which is to be equal to $c_{12\cdots\nu}$. So if this coefficient has been normalized to unity, $\Delta_{\mu-1\mu} = 1$. If we consider next the terms of order $\nu-2(\equiv\mu-4)$, and compare coefficients, then, instead of (18), we have

$$c_{i_1 i_2 \cdots i_{\nu-2}} = (-1)^{i_{\nu-1}+i_{\nu-1}}(\Delta_{\mu-1\mu}\Delta_{i_{\nu-1}i_\nu} + \Delta_{\mu-1i_\nu}\Delta_{\mu i_{\nu-1}} - \Delta_{\mu-1i_{\nu-1}}\Delta_{\mu i_\nu}). \tag{22}$$

The right-hand side is the Pfaffian of order 4 obtained by striking out the rows and columns of (21) which contain the indices $i_1 < i_2 < \cdots < i_{\nu-2}$, together with the explicit signs of the cofactor. The indices $i_{\nu-1}$ and i_ν are the two indices remaining from the set $1, \cdots \nu$ (which label the rows $3, \cdots, \mu$). The Δ's cannot be determined explicitly from (22), but if $\Delta_{\mu-1\mu} \neq 0$, this is not necessary. For if we apply the reduction theorem (A7) to (21), with $a_{12} = \Delta_{\mu-1\mu}$, $a_{2i} = \Delta_{\mu i}$, and $a_{ij} = \Delta_{ij} + a_i a_j$, then (with the Pfaffian specified by the set of initial elements in each row),

$$\diagdown a_{12}a_{2i}a_{ij}| = \diagdown \Delta_{\mu-1\mu}\Delta_{\mu i}\Delta_{ij}| + \Delta_{\mu-1\mu}a_i a_j,$$

and (21) becomes

$$P(a) = \Delta_{\mu-1\mu}^{-\frac{1}{2}\mu+2} :\diagdown\diagdown \Delta_{\mu-1\mu}\Delta_{\mu i}\Delta_{ij}| + \Delta_{\mu-1\mu}a_i a_j|:. \tag{23}$$

The coefficient of the term of order ν is $\Delta_{\mu-1\mu}^{-\frac{1}{2}\mu+2} \times \Delta_{\mu-1\mu}^{\frac{1}{2}\mu-1} = \Delta_{\mu-1\mu}$, as already stated. The coefficient of a term of order $\nu - 2$, namely (22), can be written as

$$c_{i_1\cdots i_{\nu-2}} = (-1)^{i_{\nu-1}+i_{\nu-1}}\diagdown \Delta_{\mu-1\mu}\Delta_{\mu i_{\nu-1}}\Delta_{i_{\nu-1}i_\nu}|.$$

The coefficient of a term of order $\nu - 4$, using the expansion (A3), has the form

$$c_{i_1\cdots i_{\nu-4}} = (-1)^P \Delta_{\mu-1\mu}\Delta_{\mu-1}^{-1} \begin{vmatrix} \diagdown \Delta_{\mu-1\mu}\Delta_{\mu i_{\nu-3}}\Delta_{i_{\nu-3}i_{\nu-2}}| & \diagdown \Delta_{\mu-1\mu}\Delta_{\mu i_{\nu-3}}\Delta_{i_{\nu-3}i_{\nu-1}}| & \diagdown \Delta_{\mu-1\mu}\Delta_{\mu i_{\nu-3}}\Delta_{i_{\nu-3}i_\nu}| \\ \diagdown \Delta_{\mu-1\mu}\Delta_{\mu i_{\nu-2}}\Delta_{i_{\nu-2}i_{\nu-1}}| & \diagdown \Delta_{\mu-1\mu}\Delta_{\mu i_{\nu-2}}\Delta_{i_{\nu-2}i_\nu}| & \\ \diagdown \Delta_{\mu-1\mu}\Delta_{\mu i_{\nu-1}}\Delta_{i_{\nu-1}i_\nu}| & & \end{vmatrix}, \tag{25}$$

where $i_{\nu-3}$, $i_{\nu-2}$, $i_{\nu-1}$ and i_ν are the indices remaining from the set $1, \cdots, \nu$ after the indices $i_1, \cdots i_{\nu-4}$ have been deleted. $(-1)^P$ is the appropriate explicit sign. If we substitute (24) in (25), we find

$$c_{i_1\cdots i_{\nu-4}} = c_{12\cdots\nu}^{-1} \times \begin{vmatrix} c_{i_1\cdots i_{\nu-4}i_{\nu-1}i_\nu} & c_{i_1\cdots i_{\nu-3}i_\nu} & c_{i_1\cdots i_{\nu-3}i_{\nu-1}} \\ & c_{i_1\cdots i_{\nu-3}i_\nu} & c_{i_1\cdots i_{\nu-3}i_{\nu-1}} \\ & & c_{i_1\cdots i_{\nu-3}i_{\nu-2}} \end{vmatrix}. \tag{26}$$

With the change of notation

$$c'_{i_1\cdots i_{\nu-4}} = c_{i_1\cdots i_{\nu-4}}/c_{1\cdots\nu},$$

$$c'_{i_1\cdots i_{\nu-1}i_\nu} = c_{i_1\cdots i_{\nu-1}i_\nu}/c_{1\cdots\nu},$$

Eq. (26) has the same form as (19). If this process is continued, the same relations now hold between the dashed coefficients as originally held between the undashed coefficients when $n = \nu = \mu$. Hence if there are *n external bonds, and graphs with n bonds to a lattice point can occur, and if a solution can be found with no dummy terminals* ($\nu = \mu$) *at the lattice point, then it can be found with additional dummy terminals.* Conversely, if there is *no solution without dummy terminals,* there is *no solution with dummy terminals.* So the introduction of dummy terminals does not improve the chance of solving the factorization problem.

The solution found in this section has been worked out for the case where n, ν, and μ are even, and $n = \nu$. The treatment for the case where these three indices are odd follows the same lines with only the modification that the first line of $P(a)$ in (17) is now $a_1, a_2, \cdots a_\nu$, while the rest of the Pfaffian has the same structure as before. Such a Pfaffian, which is of order $(\nu + 1)$, is the vacuum expectation value with respect to the internal bonds of a product of an *odd number* of linear fermion operators. For

operator Pfaffians, therefore, it is possible to have an odd number of linear fermion operators in contrast to numerical Pfaffians where only an even number of fermion operators may be present. This is simply because every fermion operator present does not act on its vacuum state, and so its corresponding Hermitian conjugate need not be present. Equations (21) and (23) need now to be appropriately modified.

4. FACTORIZATION METHOD WHEN $\nu_j < n_j$

Now we will consider how to treat the case when not all of the external bonds can be simultaneously connected to a lattice point, and for definiteness we will confine ourselves to the case where ν, n, and μ are all even, and $\nu = n - 2$. This will only be possible if there are dummy terminals present as an inspection of (17) reveals. From the last section we also conclude that $\mu = n + 2$, because $\mu > n + 2$ will not give anything new. We must have $c_{12\cdots n} = \Delta_{\mu-1\mu} = 0$, and (26) then imposes conditions on the coefficients of order ν, so that they cannot be chosen arbitrarily. In order to simplify some of the formulas, we write $c_{i_1\cdots i_\nu} = c_{i_{n-1}i_n}$, and then (26) may be written as

$$0 = c_{n-1n}c_{ij} + c_{in}c_{jn-1} - c_{in-1}c_{jn},$$

or

$$c_{ij} = (c_{in-1}c_{jn} - c_{in}c_{jn-1})/c_{n-1n}, \tag{28}$$

when $i_{n-1} = n - 1$, $i_n = n$. (If $c_{n-1n} = 0$, a corresponding relation can be written in terms of some other coefficient, since by assumption not all of them are zero. In this section, the subscript j must not be confused with the lattice-point index j of earlier sections.) Equation (28) means that for all $0 < i, j \leq n - 2 = \nu$, c_{ij} can be expressed in terms of c_{in-1} and c_{in}, of which there are only $2n-4$ independent quantities. From (22) we see that

$$\begin{aligned} c_{in-1} &= (-1)^i(\Delta_{\mu-1i}\Delta_{\mu n-1} - \Delta_{\mu-1n-1}\Delta_{\mu i}), \\ c_{in} &= (-1)^{i+1}(\Delta_{\mu-1i}\Delta_{\mu n} - \Delta_{\mu-1n}\Delta_{\mu i}). \end{aligned} \tag{29}$$

The sign factors have been given on the assumption that n is even. Equations (29) can be solved for $\Delta_{\mu-1i}$ and $\Delta_{\mu i}$ in terms of c_{in-1}, c_{in}, $\Delta_{\mu-1n-1}$, etc. The solutions are

$$\begin{aligned} \Delta_{\mu-1i} &= (-1)^{i+1}(\Delta_{\mu-1n}c_{in-1} + \Delta_{\mu-1n-1}c_{in})/c_{n-1n} \\ \Delta_{\mu i} &= (-1)^{i+1}(\Delta_{\mu n}c_{in-1} + \Delta_{\mu n-1}c_{in})/c_{n-1n}. \end{aligned} \tag{30}$$

The elements $\Delta_{\mu-1n-1}$, etc. can be chosen arbitrarily so long as (30) is satisfied. Equation (30) only determines the weight factors $\Delta_{\mu-1i}$ and $\Delta_{\mu i}$ for $1 \leq i \leq \nu$, and the remaining coefficients must be determined by appeal to the next-lower-order coefficients.

With the same notational conventions, we have

$$c_{ijkl} = (-1)^{i+j+k+l} \begin{vmatrix} \Delta_{\mu-1\mu}\Delta_{\mu-1i}\Delta_{\mu-1j}\Delta_{\mu-1k}\Delta_{\mu-1l} & & & & \\ \Delta_{\mu i} & \Delta_{\mu j} & \Delta_{\mu k} & \Delta_{\mu l} & \\ \Delta_{ij} & \Delta_{ik} & \Delta_{il} & & \\ \Delta_{jk} & \Delta_{jl} & & & \\ \Delta_{kl} & & & & \end{vmatrix}, \tag{31}$$

for $1 \leq i, j, k, l, \leq n$. Using (22), this can be expanded out to give

$$\begin{aligned} c_{ijkl} = {} & (-1)^{k+l-1}c_{ij}\Delta_{kl} - (-1)^{j+l-1}c_{ik}\Delta_{jl} \\ & + (-1)^{j+k-1}c_{il}\Delta_{jk} + (-1)^{i+l-1}c_{jk}\Delta_{il} \\ & - (-1)^{i+k-1}c_{jl}\Delta_{ik} + (-1)^{i+j-1}c_{kl}\Delta_{ij}, \end{aligned} \tag{32}$$

and if we put $k = n - 1$, $l = n$, (32) enables Δ_{ij}, for all i, j $(1 \leq i, j \leq \nu)$, to be expressed in terms of c_{ijn-1n}, c_{ij}, c_{in-1}, $\cdots$, Δ_{in-1}, Δ_{in}, Δ_{jn-1}, and Δ_{jn}. Then the remaining equations obtained from (32) by putting $l = n - 1$ or n, with $1 \leq i, j, k \leq \nu$, or $1 \leq i, j, k, l \leq \nu$ give nothing new. This follows from relations between coefficients similar to (26) with lower-order coefficients. For example it can be shown that

$$c_{ijklmn} = c_0^{-2} \begin{vmatrix} c_{ij} & c_{ik} & c_{il} & c_{im} & c_{in} \\ & c_{jk} & c_{jl} & c_{jm} & c_{jn} \\ & & c_{kl} & c_{km} & c_{kn} \\ & & & c_{lm} & c_{ln} \\ & & & & c_{ln} \end{vmatrix}, \tag{33}$$

which, on expanding out and using (26) and (27), gives

$$\begin{aligned} 0 = {} & c_{ij}c_{klmn} - c_{ik}c_{jlmn} + c_{il}c_{jkmn} \\ & - c_{im}c_{jkln} + c_{in}c_{jklm}. \end{aligned} \tag{34}$$

When the values of Δ_{ij}, etc. obtained from (32) for the special values $k = n - 1$, $l = n$ are substituted back in (32) with general values of i, j, k, l instead, then the identities (34) are obtained. All the lower-order coefficients can be expressed in terms of c_{ijkl}, and hence if these relations are satisfied by the coefficients (14), the values of the weight factors Δ_{ij} obtained above will be consistent with (17).

This process can be continued although the algebraic details would become very complicated. At each stage, new identities necessarily holding between the coefficients of (14), if the Pfaffian formalism is to be applicable, reduce the number of independent equations available for determining the weight factors. So the fewer the number of nonzero higher-order terms, the greater the restrictions on

the remaining coefficients. The elements Δ_{ij} can then be chosen with a greater degree of arbitrariness, but the variety of polynomials which can be given Pfaffian expression is severely reduced.

5. SIMPLE APPLICATIONS OF THE FACTORIZATION METHOD

In this section we will consider some simple applications of the factorization method to show its power and limitations. As mentioned earlier, a more detailed investigation of its application to the solution of a wide variety of Ising models will be given elsewhere.[8] First of all, we will consider two cases where $n = \nu$, namely the rectangular Ising model without crossed bonds and the Ising model with crossed bonds.

(*a*) *Two-dimensional Ising problem for a rectangular lattice.* In this problem each lattice point can have four bonds incident upon it and any combination of zero, two, or four bonds may occur. If the weight factors associated with each bond are absorbed into the emission operators, then, if $a_i^{(1)}$, $a_i^{(2)}$ are the operators absorbing horizontal and vertical bonds, respectively, we have the pairings

$$a_i^{(1)} \cdot a_i^{(1)\dagger} \cdot = x, \qquad a_i^{(2)} \cdot a_i^{(2)\dagger} \cdot = y. \tag{35}$$

The operator polynomial $P(a)$ in this case will be

$$\begin{aligned} P_i(a) = [1 &+ a_{i-m}^{(2)} a_{i-1}^{(1)} + a_i^{(1)\dagger} a_{i-1}^{(1)} \\ &+ a_i^{(1)\dagger} a_{i-m}^{(2)} + a_i^{(2)\dagger} a_{i-1}^{(1)} + a_i^{(2)\dagger} a_{i-m}^{(2)} \\ &+ a_i^{(2)\dagger} a_i^{(1)\dagger} + a_i^{(2)\dagger} a_i^{(1)\dagger} a_{i-m}^{(2)} a_{i-1}^{(1)}], \end{aligned} \tag{36}$$

where m is the number of lattice points in a row. If we write (36) in the simpler notation

$$\begin{aligned} P(a) = :(1 &+ a_1a_2 + a_1a_3 + a_1a_4 + a_2a_3 \\ &+ a_2a_4 + a_3a_4 + a_1a_2a_3a_4):, \end{aligned} \tag{37}$$

then

$$c_0 = c_{12} = c_{13} = c_{14} = c_{23} = c_{24} = c_{34} = c_{1234} = 1.$$

Equation (18) then gives immediately

$$\Delta_{12} = c_{34} = 1 = \Delta_{14} = \Delta_{23} = \Delta_{34},$$

and

$$\Delta_{13} = -c_{23} = -1 = \Delta_{24}.$$

As $c_0 = 1$, Eq. (19) imposes the condition

$$c_0 = 1 = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & \\ 1 & & \end{vmatrix},$$

which is valid. So the factorization is possible, and it can be verified that (37) may be written as

$$P(a) = :\begin{vmatrix} 1 + a_1a_2 & -1 + a_1a_3 & 1 + a_1a_4 \\ 1 + a_2a_3 & -1 + a_2a_4 & \\ 1 + a_3a_4 & & \end{vmatrix},$$

which was the result given by Hurst and Green.[2]

(*b*) *Crossed bonds.* It has long been known that all the lattices for which the Ising problem cannot be solved in closed form are characterized by the appearance of crossed bonds.[13] These are bonds which intersect at points other than lattice points. One approach which suggests itself to replace such a crossing point by a new lattice point at which a cross over condition[7] is applied so that bonds may only go straight through the new lattice point and not turn through 90°. Such a new lattice point would be described mathematically by introducing a new operator polynomial appropriate to the restrictions imposed. With a similar notation to (37), such a polynomial would be

$$P(a) = :(1 + a_1a_3 + a_2a_4 + a_1a_2a_3a_4):. \tag{38}$$

Now we have

$$c_0 = c_{13} = c_{24} = c_{1234} = 1,$$

and

$$c_{12} = c_{14} = c_{23} = c_{34} = 0.$$

The consistency condition (19) then requires

$$c_0 = 1 \neq \begin{vmatrix} 0 & 1 & 0 \\ 0 & 1 & \\ 0 & & \end{vmatrix} = -1. \tag{39}$$

So factorization is not possible and the Pfaffian method is no longer directly applicable. It is thus suggested that the *appearance of crossed bonds* is *equivalent to a breakdown of the factorization method,* and what is required is a generalization of this method to factors which though homogenous are not linear. It is simple to verify that (38) can be factorized into a pair of homogenous second-order fermion operators, but the mathematical structure of such expressions is still unknown. As the lattice points of all the Ising models so far found intractable—such as the three-dimensional and two-dimensional next-nearest-neighbor models—are themselves factorizable, the exact evaluation of the partition depends, from this point of view at any rate, on the understanding of (38). So the replacement of a crossing point in the original lattice by a new lattice point with a cross-over condition means that the new lattice so obtained is free from crossing points, and hence no longer contains crossed bonds. At first sight, it would appear now that this extended lattice could be solved by the Pfaffian method because, as mentioned in the introduction, the absence of crossed bonds is necessary and sufficient

[13] G. F. Newell and E. W. Montroll, Rev. Mod. Phys. **25**, 353 (1953).

for the absence of unwanted negative signs in the Pfaffian expression for the desired partition function. However there now appears a new difficulty, that of the breakdown of the factorization method. Whether the converse is true, namely that an extension of the factorization method is sufficient to cope with the cross-over condition, is still unknown.

(c) *A type of rectangular two-dimensional lattice on which there are two sets of points, with each point of one set in the center of a rectangle formed by four points of the other set.* Closed polygons are to be *of one set in the center of a rectangle formed by four points of the other set.* Closed polygons are to be drawn connecting points of one set with themselves in such a way that polygons from different sets cannot intersect. One method of tackling this problem would be to introduce new lattice points at the points where the two lattices intersect, and to choose the corresponding operator polynomials so that only lines belonging to one lattice or the other are simultaneously incident at such a lattice point.[14] Such an operator would be

$$P(a) = :(1 + a_1a_2 + a_1'a_2'):,$$

with a_1, a_2 belonging to the first lattice, and a_1', a_2' to the second. The coefficients are now

$$c_0 = c_{12} = c_{34} = 1,$$

and

$$c_{13} = c_{14} = c_{23} = c_{24} = c_{1234} = 0.$$

The consisting condition (28) then requires

$$0 \neq \begin{vmatrix} 1 & 0 & 0 \\ & 0 & 0 \\ & & 1 \end{vmatrix} = 1.$$

Once again the method will not work.

This problem is a special case of $n = 4$, $\mu = 6$, and $\nu = 2$. We can use the consistency condition (28) to find for what problems non self-intersecting graphs can be constructed in accordance with the factorization method. Such non self-intersecting graphs may have applications to the discussion of self-avoiding walks, although the problem of obtaining the correct explicit sign for the terms of the Pfaffian expansion always has to be watched. Suppose we choose $c_{34} = 1$ and then put c_{i3}, $c_{i4} = 1$ or 0, and find the elements Δ_{ij} from (30) and (31). This means that we will determine what polynomials $(1 + c_{ij}a_ia_j)$ can be factorized. The results are listed in tabular form in Table I. The four bonds connected to a lattice point are numbered 1, 2, 3, 4, and pairs of bonds which can be connected are labeled (12), (13), etc. In the table are given the values of c_{ij} normalized to 1 or 0 except in some cases where the value -1 must be taken.

[14] I am indebted to Dr. D. Ruelle for suggesting this problem which provided the stimulus for this whole undertaking.

TABLE I. Possible two-bond connection to a factorizable lattice point.

(34)	(13)	(14)	(23)	(24)	(12)	(34)	(13)	(14)	(23)	(24)	(12)
1	1	1	1	1	0	1	0	1	1	1	−1
1	1	1	1	0	−1	1	0	1	1	0	−1
1	1	0	1	1	1	1	0	0	1	1	0
1	1	0	1	0	0	1	0	0	1	0	0
1	1	1	0	1	1	1	0	1	0	1	0
1	1	1	0	0	0	1	0	1	0	0	0
1	1	0	0	1	1	1	0	0	0	1	0
1	1	0	0	0	0	1	0	0	0	0	0

ACKNOWLEDGMENTS

This work was initiated while I was a guest of the Tait Institute of Mathematical Physics, and I would like to thank Professor N. Kemmer, F.R.S., and members of the Institute for their hospitality during my stay there. I would also like to thank the Royal Society of London for the grant of a Nuffield Bursary.

APPENDIX

Pfaffians have some properties which are closely analogous to those of determinants, and in this appendix, two simple results which closely imitate well known determinantal theorems will be given.

(a) *Addition of rows and columns.* If $D = |a_{ij}|$ is a skew-symmetric determinant, then it is unchanged in value, and is still skew symmetric when the pth column is added to the p'th column and the pth row to the p'th row. So the corresponding Pfaffians are also equal. (Strictly speaking, this argument only requires the corresponding Pfaffians to be equal in magnitude, but we can choose the sign of the square root appropriately.)

So if

$$D' = |a_{ij} + \delta_{ip'}a_{pj} - \delta_{p'j}a_{pi}| \equiv D,$$

then

$$P = \diagdown a_{ij}| = \diagdown a_{ij} + \delta_{ip'}a_{pj} - \delta_{p'j}a_{pi}| = P' \qquad (i < j). \tag{A1}$$

An alternative proof could be given by considering the definition of P as the expansion

$$P = \Sigma\,(-1)^P a_{i_1i_2}a_{i_3i_4}\cdots,$$

with $i_1 < i_2$, $i_3 < i_4$, and $(-1)^P$ is the parity of the permutation $1, 2 \cdots n \to i_1i_2i_3 \cdots$.

(b) *Reduction of a Pfaffian.* If the second row and

column is multiplied by $c_3, c_4, \cdots, c_n$ and added to the third, fourth, $\cdots$ rows and columns, respectively, the value of P is unchanged. These constants can be chosen so that $a'_{p1} = -a'_{p1} = 0$ for $p > 3$. This will be so if

$$a_{1p} + c_p a_{12} = 0, \quad \text{or} \quad c_p = -a_{1p}/a_{12}.$$

Then, from (A1), the general element of P' will be

$$\begin{aligned} a'_{ij} &= a_{ij} + \sum_{p=3}^{n} c_p(\delta_{ip}a_{2j} - \delta_{pj}a_{2i}), \\ &= a_{ij} - (a_{1i}a_{2j}/a_{12}) + (a_{1j}a_{2i}/a_{12}), \\ &= a_{12}^{-1} \begin{vmatrix} a_{12} & a_{1i} & a_{1j} \\ & a_{2i} & a_{2j} \\ & & a_{ij} \end{vmatrix}. \end{aligned} \tag{A2}$$

The element a'_{ij} is a Pfaffian of order 4 which is obtained from P by taking the elements whose indices contain two of the four suffices 1, 2, i, j without repetition. In P', the indices range over $3 \leq i < j \leq n$. Hence,

$$P = \diagdown a_{ij}| = a_{12}^{-\frac{1}{2}n+2} \diagdown\diagdown a_{12}\ a_{2i}\ a_{ij}|| = P'. \tag{A3}$$

This is the Pfaffian analogue of the determinantal theorem

$$a_{11}^{n-2}\ |a_{11}a_{22} \cdots a_{nn}| = ||a_{11}a_{22}|\ |a_{11}a_{33}| \cdots |a_{11}a_{nn}||.$$

This process of reduction can be continued to give

$$P = P'' = a_{12}^{-\frac{1}{2}n+2} a_{34}'^{-\frac{1}{2}n+4} \diagdown\diagdown a'_{34}a'_{4i}a'_{ij}||. \tag{A4}$$

From (A3) for the case $n = 6$, we have

$$\diagdown a'_{34}a'_{4i}a'_{ij}| = a_{12} \diagdown a_{12}a_{23}a_{34}a_{4i}a_{ij}|. \tag{A5}$$

If (A5) is substituted in (A4), it becomes

$$P'' = (\diagdown a_{12}a_{23}a_{34}|)^{-\frac{1}{2}n+4} \diagdown\diagdown a_{12}a_{23}a_{34}a_{4i}a_{ij}||. \tag{A6}$$

The general result, which can be proved by induction, is

$$\begin{aligned} &(\diagdown a_{12}a_{23}a_{34} \cdots a_{r-1\,r}|)^{\frac{1}{2}n-r} \diagdown\diagdown a_{12}a_{23} \cdots a_{n-1}a_n| \\ &\quad = \diagdown\diagdown a_{12}a_{23} \cdots a_{ri}a_{ij}||, \qquad (r < i < j). \end{aligned}$$

The Pfaffian on the right-hand side is of order $n - r$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 1 JANUARY 1964

Integral Representations for Vertex Functions*

C. FRONSDAL AND R. E. NORTON
Department of Physics, University of California, Los Angeles, California
(Received 30 July 1963)

The third-order Feynman graph is studied as a function of the three external masses squared for arbitrary real values of the internal masses. Single and double "dispersion" relations are derived which, for arbitrary real values of the undispersed variable(s), involve integrations only over real contours. Tables list the spectral functions for both the single and double integral formulas. In several cases, a non–Landau singularity (on the forward scattering curve) appears on the "physical sheet", but not as a singularity of the physical boundary value.

INTRODUCTION

ATTEMPTS to evaluate production amplitudes in strong interactions have taken adequate account of only two special types of contributions, the pole diagrams [Fig. 1(a)][1,2] and two-body final-state interactions [Fig. 1(b)].[3]

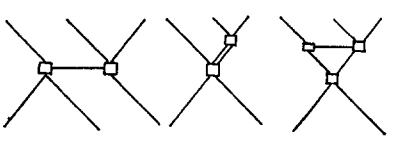


FIG. 1. (a) Pole diagram; (b) final-state interaction; and (c) triangle diagram, as contributions to a production amplitude.

Because much more detailed experimental information will soon become available, it is important to develop methods for evaluating the contributions of diagrams of the next order of complexity. (Whether such diagrams also represent the next order of

* Supported in part by the National Science Foundation.
[1] G. F. Chew and F. Low, Phys. Rev. **113**, 1640 (1959). Many applications of their theory have been published. See, for example, D. D. Carmony and R. T. Van de Walle, Phys. Rev. **127**, 959 (1962).
[2] R. Ascoli [Nuovo Cimento **18**, 744 (1960)] and Y. S. Kim [Phys. Rev. **125**, 1771 (1962)] have considered pole diagrams of a different nature from those of Ref. 1.
[3] See, for example, M. Jacob, G. Mahoux, and R. Omnes, Nuovo Cimento **23**, 838 (1962).

column is multiplied by $c_3, c_4, \cdots, c_n$ and added to the third, fourth, $\cdots$ rows and columns, respectively, the value of P is unchanged. These constants can be chosen so that $a'_{p1} = -a'_{p1} = 0$ for $p > 3$. This will be so if

$$a_{1p} + c_p a_{12} = 0, \quad \text{or} \quad c_p = -a_{1p}/a_{12}.$$

Then, from (A1), the general element of P' will be

$$\begin{aligned} a'_{ij} &= a_{ij} + \sum_{p=3}^{n} c_p(\delta_{ip}a_{2j} - \delta_{pj}a_{2i}), \\ &= a_{ij} - (a_{1i}a_{2j}/a_{12}) + (a_{1j}a_{2i}/a_{12}), \\ &= a_{12}^{-1} \begin{vmatrix} a_{12} & a_{1i} & a_{1j} \\ & a_{2i} & a_{2j} \\ & & a_{ij} \end{vmatrix}. \end{aligned} \tag{A2}$$

The element a'_{ij} is a Pfaffian of order 4 which is obtained from P by taking the elements whose indices contain two of the four suffices 1, 2, i, j without repetition. In P', the indices range over $3 \leq i < j \leq n$. Hence,

$$P = \diagdown a_{ij}| = a_{12}^{-\frac{1}{2}n+2} \diagdown\diagdown a_{12}\ a_{2i}\ a_{ij}|| = P'. \tag{A3}$$

This is the Pfaffian analogue of the determinantal theorem

$$a_{11}^{n-2}\ |a_{11}a_{22} \cdots a_{nn}| = ||a_{11}a_{22}|\ |a_{11}a_{33}| \cdots |a_{11}a_{nn}||.$$

This process of reduction can be continued to give

$$P = P'' = a_{12}^{-\frac{1}{2}n+2} a_{34}'^{-\frac{1}{2}n+4} \diagdown\diagdown a'_{34}a'_{4i}a'_{ij}||. \tag{A4}$$

From (A3) for the case $n = 6$, we have

$$\diagdown a'_{34}a'_{4i}a'_{ij}| = a_{12} \diagdown a_{12}a_{23}a_{34}a_{4i}a_{ij}|. \tag{A5}$$

If (A5) is substituted in (A4), it becomes

$$P'' = (\diagdown a_{12}a_{23}a_{34}|)^{-\frac{1}{2}n+4} \diagdown\diagdown a_{12}a_{23}a_{34}a_{4i}a_{ij}||. \tag{A6}$$

The general result, which can be proved by induction, is

$$\begin{aligned} &(\diagdown a_{12}a_{23}a_{34} \cdots a_{r-1\,r}|)^{\frac{1}{2}n-r} \diagdown\diagdown a_{12}a_{23} \cdots a_{n-1}a_n| \\ &\quad = \diagdown\diagdown a_{12}a_{23} \cdots a_{ri}a_{ij}||, \qquad (r < i < j). \end{aligned}$$

The Pfaffian on the right-hand side is of order $n - r$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 1 JANUARY 1964

Integral Representations for Vertex Functions*

C. FRONSDAL AND R. E. NORTON
Department of Physics, University of California, Los Angeles, California
(Received 30 July 1963)

The third-order Feynman graph is studied as a function of the three external masses squared for arbitrary real values of the internal masses. Single and double "dispersion" relations are derived which, for arbitrary real values of the undispersed variable(s), involve integrations only over real contours. Tables list the spectral functions for both the single and double integral formulas. In several cases, a non–Landau singularity (on the forward scattering curve) appears on the "physical sheet", but not as a singularity of the physical boundary value.

INTRODUCTION

ATTEMPTS to evaluate production amplitudes in strong interactions have taken adequate account of only two special types of contributions, the pole diagrams [Fig. 1(a)][1,2] and two-body final-state interactions [Fig. 1(b)].[3]

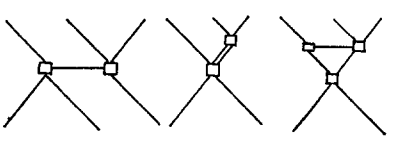


FIG. 1. (a) Pole diagram; (b) final-state interaction; and (c) triangle diagram, as contributions to a production amplitude.

Because much more detailed experimental information will soon become available, it is important to develop methods for evaluating the contributions of diagrams of the next order of complexity. (Whether such diagrams also represent the next order of

* Supported in part by the National Science Foundation.
[1] G. F. Chew and F. Low, Phys. Rev. **113**, 1640 (1959). Many applications of their theory have been published. See, for example, D. D. Carmony and R. T. Van de Walle, Phys. Rev. **127**, 959 (1962).
[2] R. Ascoli [Nuovo Cimento **18**, 744 (1960)] and Y. S. Kim [Phys. Rev. **125**, 1771 (1962)] have considered pole diagrams of a different nature from those of Ref. 1.
[3] See, for example, M. Jacob, G. Mahoux, and R. Omnes, Nuovo Cimento **23**, 838 (1962).

relevance can only be decided *a posteriori*.) In this paper we make some remarks that are relevant to the diagram of Fig. 1(c).

The straightforward evaluation of Feynman amplitudes associated with the triangle diagram of Fig. 1(c) predicates knowledge of the three boxes, i.e., form factors and scattering amplitudes, off the mass shell. It has been a very significant development of the last few years that off-shell knowledge is now seldom required. This is due to the introduction of methods based on the unitarity of the S-matrix and the analyticity of its elements; more particularly to the use of dispersion relations and the Mandelstam representation. Cutkosky[4] has announced a program for representing other "reduced" diagrams by integral representations, but up to now it remains unimplemented except for trivial special cases. The present work may be regarded as implementing Cutkosky's suggestions for the triangle diagram. We show that integral representations can be found that express the contribution of Fig. 1(c) in terms of the boxes on the mass shell only. These representations are approximate in the sense that they take adequate account of only those singularities that are explicit in Fig. 1(c), or more precisely, the singularities of the triangle diagram of Fig. 2. In actual applications the effect of singularities within the boxes will have to be considered, and represented by more complicated diagrams. Details of this program, together with physical applications, will be left to subsequent publications.

When the singularities of the boxes in Fig. 1(c) are ignored, the problem reduces to a study of the vertex diagram of Fig. 2. At present there are no general formulas for the corresponding amplitude that are suitable for numerical evaluation, except for special values of the variables. A byproduct of our investigation is to supply such formulas.

Considering the amplitude of Fig. 2. as a function of the six variables comprising the three external masses $(p_i^2)^{\frac{1}{2}}$ and the three internal masses m_i, we derive one-dimensional dispersion relations in one of the p_i^2 which is valid for arbitrary, fixed real values of the other five variables. Double integral representations are also obtained in two of the external masses for fixed real values of the other four masses. *In every case, the integration contours extend over real domains only,* although they are not always restricted to the normal physical cuts.

The methods used are essentially the same as in our treatment of the box diagram,[5] though the triangle proved considerably more complicated.

All the results are summarized in tables.

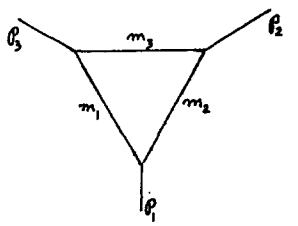


FIG. 2. Idealized triangle diagram.

DISPERSION RELATIONS

We treat spinless particles, normalize the states by $\langle p \mid p' \rangle = 2E(2\pi)^3\delta(p - p')$, and define the T matrix by $S = 1 - i(2\pi)^4 T\delta(\sum p)$. The diagram of Fig. 2 makes a contribution to the T matrix which we write

$$(\tfrac{1}{16})g_1g_2g_3F(p_1^2, p_2^2, p_3^2).$$

The function F satisfies a dispersion relation in p_1^2 if p_2^2, p_3^2 are restricted to the "physical region"

$$0 < p_2^2 < m_2^2 + m_3^2, \qquad 0 < p_3^2 < m_1^2 + m_3^2,$$
$$(p_2^2)^{\frac{1}{2}} + (p_3^2)^{\frac{1}{2}} < m_1 + m_2. \tag{1}$$

The dispersion relation actually has a larger range of validity; but it is convenient to begin by restricting the variables to (1), since there the unitarity of the S matrix determines the discontinuity unambiguously. The well known result is

$$F = \frac{1}{\pi}\int_1^\infty \frac{dx_1'}{x_1' - x_1} A(x_1', x_2, x_3), \tag{2}$$

$$A = (1/\pi)\Delta^{-\frac{1}{2}} \ln [(\alpha + \beta)/(\alpha - \beta)]. \tag{3}$$

The variables x_i are defined by $x_i = (p_i^2 - m_i^2 - m_{i+1}^2)/2m_im_{i+1}$ $(4 = 1)$, and the other new quantities in (2) are

$$\alpha = (p_1^2)^{-1}[p_1^2(p_1^2 - m_1^2 - m_2^2 + 2m_3^2)$$
$$+ p_2^2(m_2^2 - m_1^2 - p_1^2) + p_3^2(m_1^2 - m_2^2 - p_1^2)], \tag{4}$$

$$\beta = (p_1^2)^{-1}\{[p_1^2 - (m_1 + m_2)^2]$$
$$\times [p_1^2 - (m_1 - m_2)^2]\Delta\}^{\frac{1}{2}}, \tag{5}$$

$$\Delta = \sum_i p_i^4 - 2\sum_{i<j} p_i^2p_i^2. \tag{6}$$

It is well-known, and easily verified by writing down the "Feynman-parametrized" integral representation, that F has a unique extension into the product of the upper half complex planes of p_1^2, p_2^2, p_3^2.

[4] R. E. Cutkosky, J. Math. Phys. **1**. 429 (1960).

[5] C. Fronsdal, R. E. Norton, and K. T. Mahanthappa, J. Math. Phys. **4**, 859 (1963).

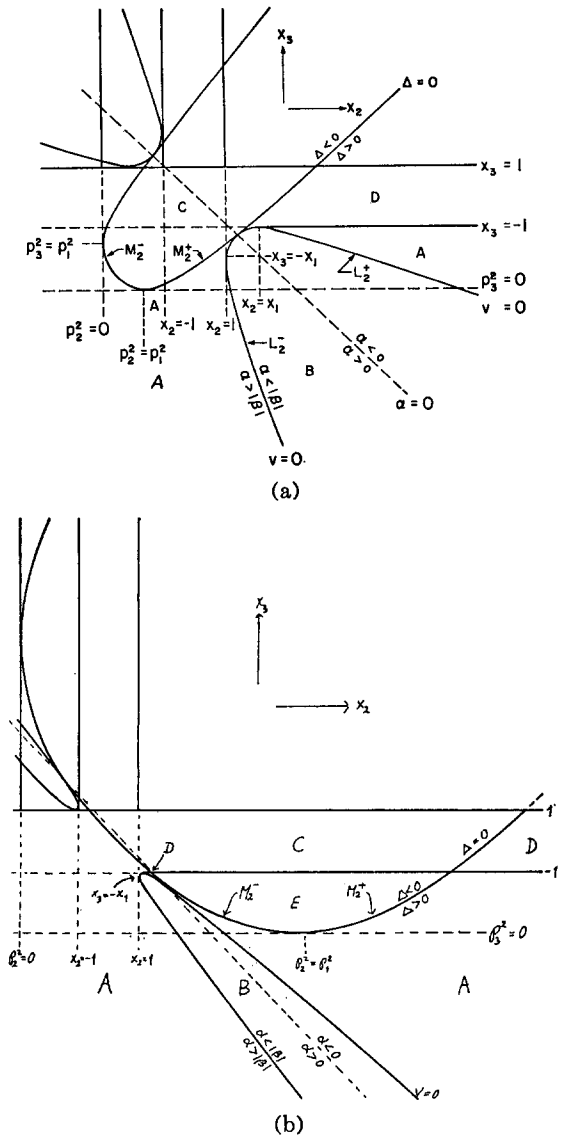


FIG. 3. (a) The x_2, x_3 plane for $x_1 > 1$ and $m_1 < m_3$ (choose $x_2 \geq x_3$); (b) x_2, x_3 plane where $x_1 \geq 1$ and $m_1 > m_3$ (choose $x_2 \geq x_3$).

Therefore, the object of our study may be taken to be defined by (2), (3) when p_2^2, p_3^2 are restricted by (1), and by analytic continuation elsewhere.

In order to effect the continuation of F, we study the singularities of A when $1 < x_1 < \infty$. The singularities of the logarithm in (2) are at[6]

$$\alpha^2 - \beta^2 \sim V(x_1, x_2, x_3) = 2x_1x_2x_3 + x_1^2 + x_2^2 + x_3^2 - 1 = 0. \quad (7)$$

The zeros of V are at

$$x_i = L_i^{\pm} \equiv -x_ix_k \pm [(1 - x_i^2)(1 - x_k^2)]^{\frac{1}{2}}. \quad (8)$$

Other possible singularities of A are the zeros of Δ; we identify these zeros by writing

$$\Delta \sim (x_i - M_i^-)(x_i - M_i^+). \quad (9)$$

In the absence of a statement to the contrary it will always be assumed that $L_i^+ > L_i^-$, and $M_i^+ > M_i^-$, whenever these quantities are real.

In Figs. 3(a) and (b) are shown the zeros of Δ, V, and α in the real x_2, x_3 plane, for $x_1 > 1$. Here, both in text and figure, it will be assumed that $m_3 > m_1$, and $m_3 > m_2$. In the summary this restriction has been lifted.

It is important to realize that the locus of all points $V(x_1, x_2, x_3) = 0$, $x_1 \geq 1$ includes the line $x_2 + x_3 = 0$; we shall begin with region $x_2 + x_3 < 0$, $x_2 < 1$, $x_3 < 1$. In this region, (2) and (3) are clearly valid wherever $\Delta > 0$. Because $\alpha > \beta$ when $\Delta = 0$ is approached, this curve is not singular, and (2) holds for $\Delta < 0$ as well, except that (3) must be replaced by

$$A = (2/\pi)\ |\Delta|^{-\frac{1}{2}} \tan^{-1}(\beta/\alpha). \quad (10)$$

When $x_3 < -1$ and x_2 is continued along the top of the real axis beyond 1, the singular points $L_1^{\pm}$ move in the x_1 plane as shown in Fig. 4, which means that (2) remains valid. Referring now to Fig. 3(a), we continue x_2 past L_2^- below the line $\alpha = 0$. Here $\alpha > \beta$ and the singular part of (3) is $-(1/\pi)\Delta^{-\frac{1}{2}}$ $\ln(L_2^- - x_2)$. Increasing x_2 with a positive imaginary part gives

$$A = (1/\pi)\ \Delta^{-\frac{1}{2}}\{\ln[(\beta + \alpha)/(\beta - \alpha)] + \pi i\}, \quad (11)$$

in the regions where $\beta > |\alpha|$. Similarly, continuing to $x_2 > L_2^+$ gives

$$A = \frac{1}{\pi}\Delta^{-\frac{1}{2}} \ln\left|\frac{+\alpha + |\beta|}{+\alpha - |\beta|}\right|. \quad (12)$$

The region that is labeled "D" in Fig. 3(a) may be

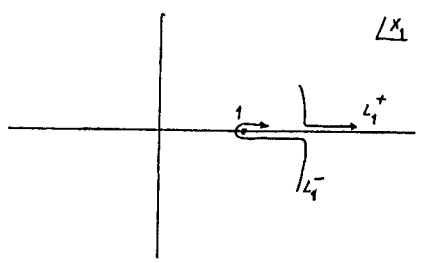


FIG. 4. The trajectories indicating how the singular points $L_1^{\pm}$ move for fixed $x_3 < -1$ as x_2 is increased just above $x_2 = 1$ and then past $x_2 = -x_3$. For $x_2 = 1$, the $L_1^{\pm}$ pinch the real axis, and as x_2 passes $-x_3$, L_1^- loops $x_1 = 1$ as shown.

[6] R. Karplus, C. M. Sommerfield, and E. H. Wichmann, Phys. Rev. 111, 1187 (1958), first discovered the anomalous singularities of the triangle diagram. Possible direct physical manifestations of these singularities have been discussed by P. V. Landshoff and S. B. Treiman, Nuovo Cimento 19, 1249 (1960), and by R. Aaron, Phys. Rev. Letters, 10, 32 (1962).

reached from the left. Here $\alpha < 0$, so that

$$A = \frac{1}{\pi i}|\Delta|^{-\frac{1}{2}} \ln \frac{-|\alpha| + i|\beta|}{-|\alpha| - i|\beta|}$$
$$= \frac{1}{\pi i}|\Delta|^{-\frac{1}{2}} \ln \left[\frac{|\alpha| - i|\beta|}{|\alpha| + i|\beta|} e^{2\pi i}\right]$$
$$\to \frac{1}{\pi}|\Delta|^{-\frac{1}{2}}\left[\ln\left|\frac{\alpha + |\beta|}{\alpha - |\beta|}\right| + 2\pi i\right]. \quad (13)$$

We now wish to continue $F(x_1, x_2, x_3)$ into the region $x_3^2 < 1$, $x_2 + x_3 > 0$, by holding x_3 fixed and increasing x_2 along the top of the real axis. Then the singular point L_1^+ of A migrates in the x_1 plane as shown in Fig. 5. The appearance of an anomalous threshold when $-x_3 < x_2 < 1$ is well-known. The representation (2) becomes

$$F = \frac{1}{\pi}\int_1^\infty \frac{dx_1'}{x_1' - x_1} A(x_1', x_2, x_3)$$
$$+ \frac{2}{\pi}\int_{L_1^+}^\infty \frac{dx_1'}{x_1' - x_1}|\Delta|^{-\frac{1}{2}}, \quad (14)$$

with A given by (12) when $x_2 < M_2^-$, by (13) when $x_2 > M_2^+$, and by (10) when $M_2^- < x_2 < M_2^+$. When $x_2 > 1$, L_1^+ is complex, and a one-dimensional dispersion relation with a real domain of integration exists only if it is possible to deform the contour in the manner shown in Fig. 5. We shall in fact prove that, in this case, the continuation gives

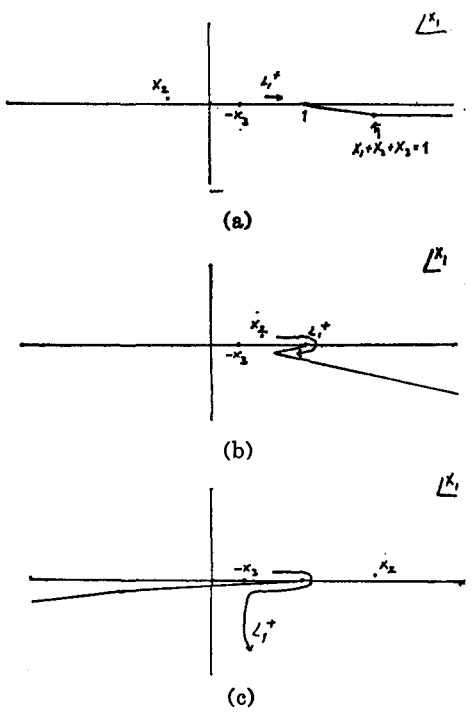


FIG. 5. The migration of the singularity L_1^+ of A as x_2 is moved through the upper half plane. The deformation of the cut is defined by Eqs. (16). In each case the cut becomes real in the limit Im $x_2 \to 0$.

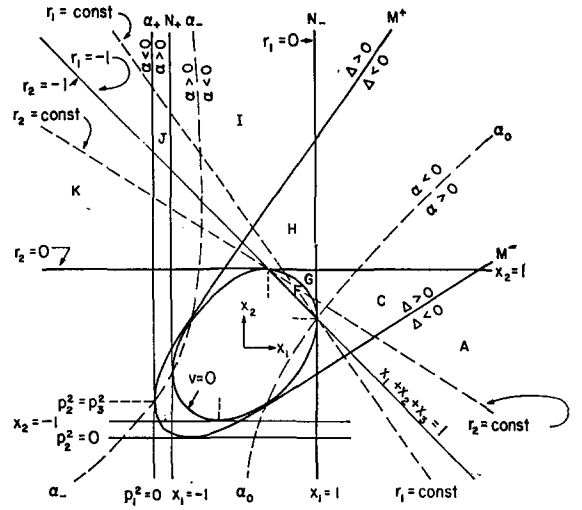


FIG. 6. The regions of integration in the single and double dispersion relations, for $x_3^2 < 1$ and $m_1 < m_2$.

$$F = -\frac{1}{\pi}\int_{-\infty}^{1} \frac{dx_1'}{x_1' - x_1} A(x_1', x_2, x_3). \quad (15)$$

The problem confronting us here is that of finding an integrating contour in the x_1 plane that satisfies the following conditions:

(i) With x_3 fixed and $x_3^2 < 1$, it goes from $x_1 = 1$ to infinity, along a path that may depend on x_2.

(ii) When $x_2 < -x_3$, it reduces to the real x_1 axis to the right of $x_1 = 1$.

(iii) When $x_2 > 1$, it reduces to a portion of the real x_1 axis.

(iv) When x_2 is in a strip $0 <$ Im $x_2 < \epsilon$, for some positive ϵ, it does not snag on any of the singularities of $A(x_1, x_2, x_3)$. In order to derive a double

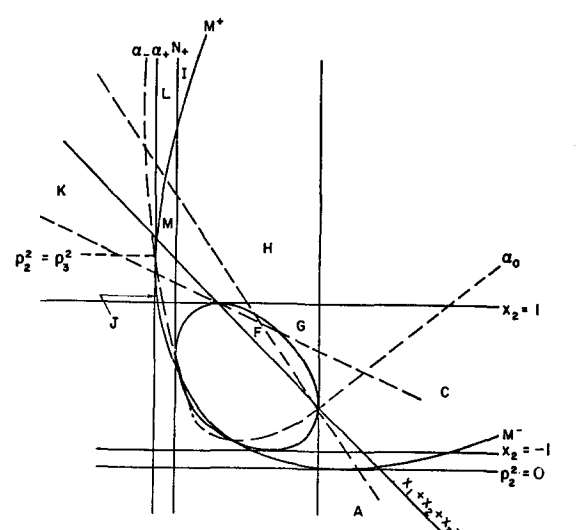


FIG. 7. Regions of integration in the single and double dispersion relations when $m_1 > m_2$ and $x_3^2 < 1$. The region J is empty if the point $p_2^2 = p_3^2$ lies below $x_2 = 1$.

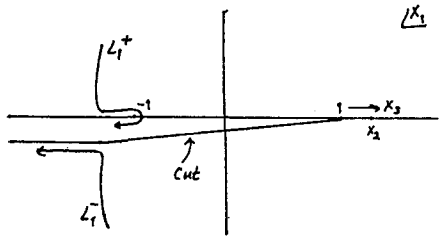


FIG. 8. The migration of $L_1^{\pm}$ as x_3 increases beyond 1, for $x_2 > 1$.

dispersion relation it is necessary to strengthen (iv).

As we know of no systematic way of solving this problem, we cannot do better than present a solution. Our choice of cut is the broken line parametrized by

$$r_1 = (x_1 - 1)/(x_2 + x_3), \qquad 0 > r_1 > -1, \tag{16}$$
$$r_2 = (x_2 - 1)/(x_1 + x_3), \qquad -1 < r_2 < 0.$$

This choice clearly satisfies the first three of the above conditions; its position in the x_1 plane is illustrated in Fig. 5. Therefore, if $x_3^2 < 1$ and $x_2 < -x_3$, the dispersion relation (2) may be written

$$F = \frac{1}{\pi}\int_0^{-1} \frac{dr_1'}{r_1' - r_1} A_1(r_1', x_2, x_3) + \frac{1}{\pi}\int_{-1}^{0} \frac{r_2}{r_2'}\frac{dr_2'}{r_2' - r_2} A_2(r_2', x_2, x_3), \tag{17}$$

where A_1 and A_2 are obtained from $A(x_1, x_2, x_3)$ by eliminating x_1 in favor of r_1 and r_2, respectively.

We have yet to test for condition (iv); in fact, a much stronger result is easily obtained.

In terms of r_1 or r_2, the expressions for α, β, V, and Δ have the form, up to irrelevant positive constant factors:

$$\alpha \sim -(x_2 - \alpha_0)(x_2 - \alpha_-)(x_2 - \alpha_+)^{-1},$$
$$\beta \sim \{(x_2 - N^-)(x_2 - N^+)\Delta\}^{\frac{1}{2}}(x_2 - \alpha_+)^{-1}, \tag{18}$$
$$\Delta \sim (x_2 - M^-)(x_2 - M^+),$$
$$V \sim -(x_2 - L)(x_2 - L')(x_2 - \alpha_+)^{-1}.$$

The possible singular points of A_1 or A_2 are α_+, N^*, M^*, L, and L', which of course depend on x_3 and either r_1 or r_2. All these points are real when $x_3^2 < 1$ and $-1 < r_i < 0$, which shows that *(17) is valid for $x_3^2 < 1$ and for all complex s_2 except for a cut along part of the real axis.* This not only shows that condition (iv) is satisfied, but provides the basis for a double integral representation.

The points $x_2 = \alpha_0$, $x_2 = \alpha_-$ may be represented in the real x_1, x_2 plane as the intersection of the curves r_i = constant and $\alpha = 0$. The other singular point may be similarly represented, and in Fig. 6 we have labeled various lines in such a way that, e.g., α_0 is that branch of the curve $\alpha = 0$, whose intersection with r_i = constant corresponds to $x_2 = \alpha_0$. This figure is qualitatively correct when $m_1 < m_2$. The extension to $m_1 > m_2$ is described in Fig. 7. With Fig. 6 giving the relative positions of the singular points along the real x_2 axis for fixed values of r_1 or r_2, A_1 and A_2 may be continued into the relevant regions on the basis of Eq. (18). As we have noted already, Eqs. (10) and (12) hold in the region $x_1 > 1$, which is relevant for $x_2 < 1$. The continuation into the regions F, G gives

$$A_1 = A_2 = \frac{i}{\pi}|\Delta|^{-\frac{1}{2}}\left[\ln\frac{|\beta| - \alpha}{|\beta| + \alpha} - i\pi\right] \text{ in } F, \tag{19a}$$

$$A_1 = \frac{i}{\pi}|\Delta|^{-\frac{1}{2}}\ln\frac{-\alpha + |\beta|}{-\alpha - |\beta|} \text{ in } G. \tag{19b}$$

$$A_2 = \frac{i}{\pi}|\Delta|^{-\frac{1}{2}}\left[\ln\frac{-\alpha + |\beta|}{-\alpha - |\beta|} - 2\pi i\right]$$

The integration over A_1 runs to the left from $x_1 = 1$

TABLE I. Single dispersion relations for $F(x_1, x_2, x_3)$, covering all real values of x_2, x_3. The absorbtive parts are listed in Table II. No generality is lost by taking $x_2 \geq x_3$ whenever either x_2 or x_3 is less than 1, and taking $m_1 < m_2$ when $x_2 > 1$, $x_3 > 1$.

If:	Dispersion relation	Integration domains	Remarks
$x_2 + x_3 < 0$ or $x_2 < -1$ or $x_3 < -1$	$F = \frac{1}{\pi}\int_1^{\infty} \frac{dx_1'}{x_1' - x_1} A(x_1', x_2, x_3)$	A–E	Choose $x_2 \geq x_3$. See Fig. 3(a) if $m_1 < m_3$, Fig. 3(b) if $m_1 > m_3$.
$x_2 + x_3 > 0$ and $x_2^2 < 1$ and $x_3^2 < 1$	F = above $+ \frac{2}{\pi}\int_{L_1^+}^{1} \frac{dx_1'}{x_1' - x_1}\|\Delta\|^{-\frac{1}{2}}$	A, C	$L_1^+ = -x_2x_3 + [(1 - x_2^2)(1 - x_3^2)]^{\frac{1}{2}}$
		H–K	$x_3 < 1$, $m_1 < m_2$. See Fig. 6.
$x_2 > 1$	$F = -\frac{1}{\pi}\int_{-\infty}^{1} \frac{dx_1'}{x_1' - x_1} A(x_1', x_2, x_3)$	H–M	$x_3 < 1$, $m_1 > m_2$. See Fig. 7.
$x_3 > -1$		A, H–N	$x_3 > 1$, choose $m_1 < m_2$. See Figs. 9, 10 and 11.

TABLE II. The absorbtive parts that appear in Table I are listed. In region G there are two alternatives, $A_1(r_1, x_2, x_3)$ and $A_2(r_2, x_2, x_3)$ representing different boundary values in this region; similarly in region O. A factor $\pi^{-1}|\Delta|^{-\frac{1}{2}}$ has been left out. The $\tan^{-1}$ takes its values between 0 and π.

A	$\ln\left\|\frac{\alpha+\|\beta\|}{\alpha-\|\beta\|}\right\| \equiv P\ln$
B	$P\ln + \pi i$
C	$2\tan^{-1}(\|\beta\|/\alpha)$
D	$P\ln + 2\pi i$
E	$-2\tan^{-1}(\|\beta\|/-\alpha)$
F	$-iP\ln + \pi$
G, A_1	$-iP\ln$
G, A_2	$-iP\ln + 2\pi$
H	$-iP\ln$
I	$2i\tan^{-1}(\|\beta\|/-\alpha)$
J	$-P\ln + 2\pi i$
K	$P\ln + \pi i$
L	$-P\ln$
M	$2\tan^{-1}(\|\beta\|/-\alpha)$
N	$-P\ln + \pi i$
O, A_1	$P\ln$
O, A_3	$P\ln + 2\pi i$

to $1 - x_2 - x_3$; that over A_2, to the right from $1 - x_2 - x_3$ to 1. Therefore only the difference contributes, and (17) reduces to (14), with $A = A_2 - A_1 = 2|\Delta|^{-\frac{1}{2}}$ in the anomalous region. When $x_2 > 1$, Eq. (15) is valid; the absorbtive parts are given in Tables I and II.

Finally we extend our results to the case $x_2 > 1$, $x_3 > 1$. For our present purpose, no generality would be lost by choosing $m_1 < m_2$, but for application to the next section we consider both cases.

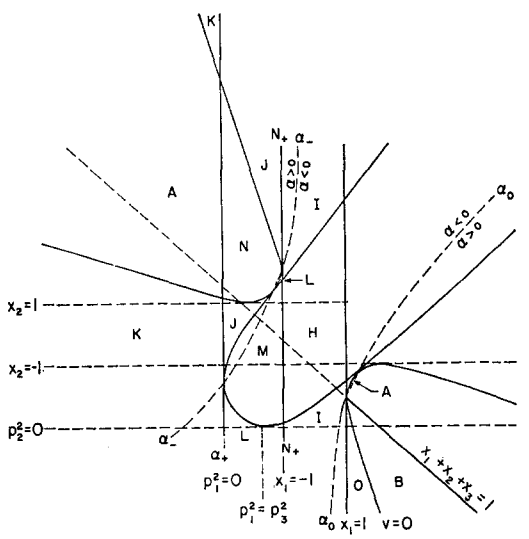


FIG. 9. The x_1, x_2 plane when $1 < x_3 < (m_1^2+m^2{}_2)/2m_1m_2$, and $m_3 < m_2 > m_1$. For $m_3 > m_2 < m_1$, see Fig. 10; only the latter inequality is of any consequence in the important region between $p_1^2 = 0$ and $x_1 = -1$. For $x_3 > (m_1^2 + m_2^2)/2m_1m_2$, see Fig. 11.

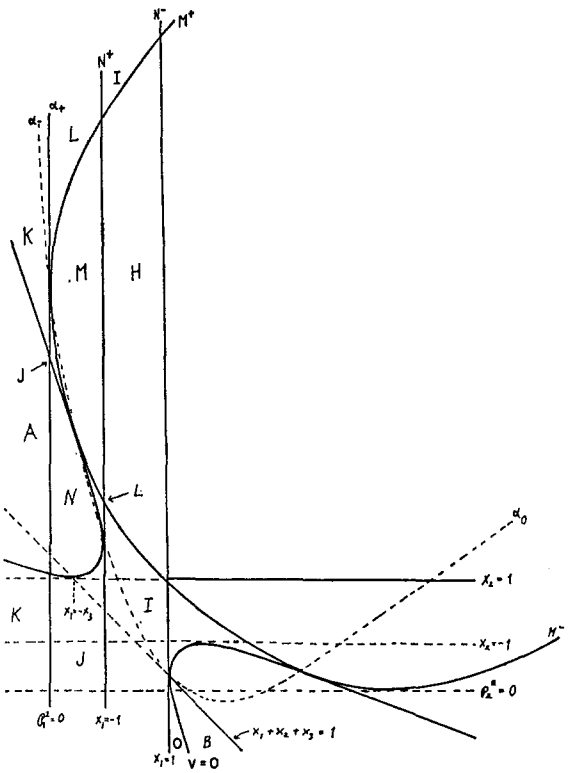


FIG. 10. The x_1, x_2 plane when $1 < x_3 < (m_1^2 + m_2^2)/2m_1m_2$, and $m_3 > m_2 < m_1$. The extension to $x_3 > (m_1^2 + m_2^2)/2m_1m_2$ is trivial and no special map is drawn up for this case. The condition $m_3 > m_2$ is not important for the relevant part of the Figure; for $m_1 < m_2$, see Figs. 9 and 11.

For a fixed value of x_2, greater than 1, the first thing that happens when x_3 is increased above 1 is shown in Fig. 8. The continuation is straightforward, and the results listed in Tables I and II and in Figs. 9–11. (For the present only the part $x_2 > 1$ of the figures is relevant.)

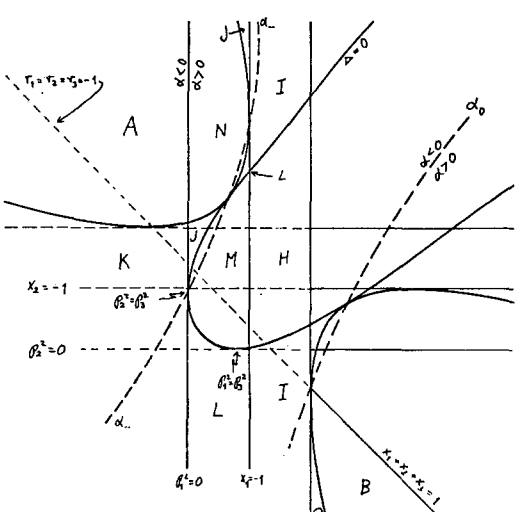


FIG. 11. The x_1, x_2 plane for $x_3 > (m_1^2 + m_2^2)/2m_1m_2$, and $m_1 < m_2 > m_3$.

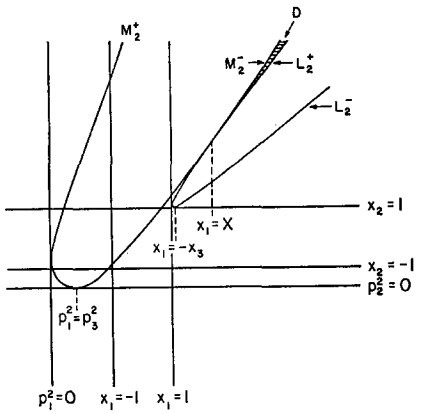


FIG. 12. The shaded region D is the "anomalous" integration region in Eq. (22), when $x_3 < -1$, $p_3^2 > 0$, and $m_1 > m_3$. When $m_1 < m_3$, the parabola $M_2^{\pm}$ lies to the right of the hyperbola $L_2^{\pm}$, and D is between M_2^+ and L_2^-.

DOUBLE DISPERSION RELATIONS

In order to derive double dispersion relations in x_1 and x_2 we study the analytic structure of the absorbtive parts in the whole x_2 plane. The simplest results obtain when $p_3^2 < 0$. Here the dispersion relation (2) holds, and it is sufficient to derive a dispersion relation for $A(x_1, x_2, x_3)$, as defined by (3), for $x_1 > 1$. Figure 3 shows that here $L_2^{\pm}$ are real and $M_2^{\pm}$ are complex. The latter are not singular because A is given by the principle branch of the logarithm. This may also be verified as follows. As x_1 varies over the integration region, M_2^+ (or M_2^-) varies along a curve which lies entirely in the upper half of the complex x_2 plane, and which has one finite complex end point, say, P. Because the continuation into the upper half plane must be unique, the function must be one-valued across this line. (Note the relevance of the fact that P is not a real point—both sides of the line are accessible from the upper half plane.) Therefore, A is regular at that zero of Δ which lies in the upper half plane. If the cuts connected to $L_2^{\pm}$ are chosen in such a way that A is a real analytic function, then the other zero is also a regular point.

This shows that, if $p_3^2 < 0$, A is regular in the x_2 plane except for the two real branchpoints $L_2^{\pm}$. A cut from L_2^- to L_2^+ along the real x_2 axis makes A a one-valued, real analytic function in the cut plane. The discontinuity of A is simply $2i$ Im $A = 2i\ |\Delta|^{-\frac{1}{2}}$. By substituting the dispersion relation for A into (2) we get the Mandelstam representation

$$F = \frac{1}{\pi^2}\int_1^{\infty}\frac{dx_1'}{x_1' - x_1} \times \int_{L_2^-}^{L_2^+}\frac{dx_2'}{x_2' - x_2}\frac{1}{[\Delta(x_1', x_2', x_3)]^{\frac{1}{2}}}, \quad p_3^2 < 0. \tag{20}$$

The Mandelstam representation may be extended to the domain $p_3^2 > 0$, $x_3 < -1$. For this purpose we must follow the zeros $M_2^{\pm}$ of Δ as p_3^2 is increased, and determine their paths relative to $L_2^{\pm}$. Without loss of generality we may take $m_1 > m_3$ until further notice.

Consider p_1^2 fixed and increase p_3^2 past zero. Then $M_2^{\pm}$ converge on the real x_2 axis at the point $p_2^2 = p_1^2$, to the right of $L_1^{\pm}$. As p_3^2 is increased further, M_2^- loops aroung L_2^+ and drags the contour with it to the right. The point at which M_2^- loops around L_2^+ is given by

$$p_1^2(p_3^2 - m_3^2 + m_1^2)^2 = p_3^2(p_1^2 - m_2^2 + m_1^2)^2, \tag{21}$$

and the corresponding value of x_2 by either of two equations obtained by cyclic permutations. (Only two of these three equations are independent.) When $x_1 > 1$, (21) has one and only one solution in the range $p_3^2 > 0$, $x_3 < -1$, and vice versa. Let this solution of (21) be

$$x_1 = X(p_3^2) > 1.$$

Then the Mandelstam representation (20) gets modified as follows:

F = "normal part" given by (20)

$$+ \frac{2}{\pi^2}\int_X^{\infty}\frac{dx_1'}{x_1' - x_1}\int_{L_2^+}^{M_2^-}\frac{dx_2'}{x_2' - x_2}\frac{1}{[\Delta(x_1', x_2', x_3)]^{\frac{1}{2}}}. \tag{22}$$

The "anomalous" region of integration is shown in Figs. 3 (b) and (12) (labeled D, $x_3 < -1$). When $m_1 < m_3$ the anomalous region is D, $x_3 < -1$ of Fig. 3(a). In either case, the double spectral function is Im $A = 2\ |\Delta|^{-\frac{1}{2}}$.

When x_3 is increased beyond -1, (22) fails to be a convenient representation because $L_2^{\pm}$ become complex. We have already noted that when $x_3 > -1$ the path of the x_1 integration must depend on x_2 (Fig. 5.) Therefore, the search for a double representation with real integration contours must go beyond the Mandelstam representation. Past experience[5] has indicated that the simplest way of solving this is to perform a conformal mapping

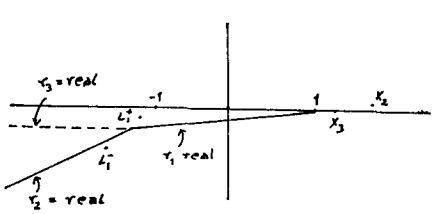


FIG. 13. The region between r_2 = real and r_3 = real is singularity free. For $x_3 > 1$, integrate along the latter.

which puts all the singularities on the real axis. In terms of the new variables, a Mandelstam representation can be written down which reduces to an integral formula of the Bergmann-Oka-Weil (B-O-W) type[7] when the old variables are replaced.

More precisely, let $r = r(x_1, x_2, x_3)$, with r real, be a parametrization of a cut beginning at $x_1 = 1$ and going to infinity. With a judicious choice of the function $r(x_1, x_2, x_3)$, it is possible to ensure that the singularities of A touch the cut only for real values of x_2. After eliminating x_1 in favor of r, a Mandelstam representation can be derived for F in terms of r and x_2, with x_3 fixed. If replacement of the original variables yields a real integration domain, a B-O-W representation is obtained with real integration contours.

A possible choice for the x_1 cut was given in Eq. (16), and the result of eliminating x_1 from $F(x_1, x_2, x_3)$ in Eq. (17). When $x_3^2 < 1$, this dispersion relation agrees with (2) when $x_2 < -x_3$, and is valid in the whole x_2 plane except on the real axis to the right of 1. This follows from the fact that all the singular points displayed in (18) are real. For all but M^{**} this is a trivial observation; that M^{**} are real is seen from the fact that the lines r_i = constant, $-1 < r_i < 0$, in the x_1, x_2 plane always have an intersection with the curve $\Delta = 0$, if $x_3^2 < 1$ (Figs. 6 and 7).

To write a dispersion relation in x_2 for the integrands in Eq. (17) we need to know $A_i(r_i, x_2 \pm i\epsilon, x_3)$ for $-1 < r_i < 0$, $x_3^2 < 1$ and x_2 real. The boundary value from above is given in Table II, and it turns out that $A_i(x_2 - i\epsilon) = A_i^*(x_2 + i\epsilon)$. Thus

$$A_i(r_i, x_2, x_3) = \frac{1}{\pi}\int_{-x_3}^{\infty} \frac{\operatorname{Im} A(r_i, x_2', x_3)}{x_2' - x_2}\, dx_2', \tag{23}$$

for $-1 < r_i < 0$, $x_3^2 < 1$. A double dispersion relation is obtained immediately by substituting (23) into (17) and replacing r_i by x_1 according to (16), and similarly r_i' by x_i'.

Before writing down the results, we introduce some notation which makes more apparent the B-O-W nature of the final expressions. In terms of x_i and x_i', the kernels in Eq. (17) can be written

$$\begin{aligned}
\frac{dr_1'}{r_1' - r_1} &= \frac{(x_2 + x_3)\, dx_1'}{(x_1' - x_1)(x_2 + x_3) - (x_2' - x_2)(x_1 - 1)} \\
&\equiv \frac{x_2 + x_3}{N_1}\, dx_1', \\
\frac{r_2}{r_2'}\frac{dr_2'}{r_2' - r_2} &= \frac{(x_2 - 1)\, dx_1'}{(x_1' - x_1)(x_2 - 1) - (x_2' - x_2)(x_1 + x_3)} \\
&\equiv \frac{1 - x_2}{N_2}\, dx_1.
\end{aligned} \tag{24}$$

The two terms of Eq. (17) contribute equal and opposite amounts in the regions F and G in Fig. 6. We add these contributions by combining the kernels:

$$\frac{1}{x_2' - x_2}\left(\frac{1 - x_2}{N_2} - \frac{x_2 + x_3}{N_1}\right) = -||q_{ij}||, \tag{25}$$

where

$$\begin{aligned}
q_{11} &\equiv (x_2 + x_3)/N_1, & q_{12} &\equiv (1 - x_1)/N_1, \\
q_{21} &\equiv (1 - x_2)/N_2, & q_{22} &\equiv (x_1 + x_3)/N_2
\end{aligned} \tag{26}$$

are B-O-W kernel functions for the two cuts parametrized in (16), and $||q_{ij}||$ is the B-O-W kernel for their intersection (a part of the distinguished boundary[7]).

The final result for $x_3^2 < 1$ is

$$\begin{aligned}
F = -\frac{1}{\pi^2}\Bigg\{ &\int_{-x_3}^{1} dx_1' \int_{1-x_1'-x_3}^{1} dx_2'\, ||q_{ij}|| \\
&+ \int_{-\infty}^{-x_3} dx_1' \int_{1}^{1-x_1'-x_3} dx_2' \frac{q_{21}}{x_2' - x_2} \\
&+ \int_{-\infty}^{1} dx_1' \int_{1-x_1'-x_3}^{\infty} dx_2' \frac{q_{11}}{x_2' - x_2}\Bigg\} \\
&\operatorname{Im} A(x_1', x_2', x_3),
\end{aligned} \tag{27}$$

where Im A is easily taken from Figs. 6, 7, and Table II.

There remains the case $x_3 > 1$. Unfortunately neither (15) nor (17) hold in the complex x_2 plane, and a new cut has to be chosen. Fig. 13 shows the position of the cut parametrized by r_1 and r_2 for the case x_3 just above 1 and real, x_2 larger than 1 with a small positive imaginary part. From Figs. 9 and 10 we see that there are no singularities in the x_1 plane between the left part of the cut (that parametrized by r_2), and the horizontal dotted line. Therefore (17) may be modified by introducing a new contour of integration parametrized by

$$\begin{aligned}
r_1 &= (x_1 - 1)/(x_2 + x_3), & 0 > r_1 > -1, \\
r_3 &= (x_3 - 1)/(x_1 + x_2), & -1 < r_3 < 0,
\end{aligned} \tag{28}$$

[7] For earlier applications of this representation and references to the original work, see C. Fronsdal, K. T. Mahanthappa, and R. E. Norton, Phys. Rev. **127**, 1847 (1962). We are grateful to Yu. A. Simonow for sending us a preprint of an article published in Zh. Eksperim. i Teor. Fiz., **43**, 2263, 1962 [English transl.: Soviet Phys.—JETP **16**, 1599 (1963)]. In that paper, a double integral representation of the B-O-W type is obtained for the case $m_1 = m_2 = m_3$. The aim seems to be quite different from ours, however, since no effort is made to achieve a real distinguished boundary.

TABLE III. Double integral respresentations for $F(x_1, x_2, x_3)$,covering all real values of x_3. The absorbtive parts are listed in Table II. The kernel functions are given in the text by (24), (26) for $x_3^2 < 1$, and by (33), (34) for $x_3 > 1$.

If:	Double dispersion relation	Integration domains	Remarks
$p_3^2 < 0$	$F = \frac{1}{\pi^2}\int_1^\infty \frac{dx_1'}{x_1' - x_1}\int_{L_2^-}^{L_2^+} \frac{dx_2'}{x_2' - x_2}\,\lvert\Delta(x_1', x_2', x_3)\rvert^{-\frac{1}{2}}$	B	Fig. 3(a) Fig. 3(b)
$p_3^2 > 0$ $x_3 < -1$	$F = \text{above} + \frac{2}{\pi^2}\int_X^\infty \frac{dx_1'}{x_1' - x_1}\int_{L_2^+}^{M_2^-} \frac{dx_2'}{x_2' - x_2}\,\lvert\Delta\rvert^{-\frac{1}{2}}$	B, D	Fig. 3(a) Fig. 3(b) Fig. 12
$x_3^2 < 1$	$F = -\frac{1}{\pi^2}\Big\{\int_{-x_3}^1 dx_2'\int_{1-x_3-x_2'}^1 dx_1'\,\lVert q_{ij}\rVert + \int_1^\infty dx_2'\int_{1-x_3-x_2'}^1 dx_1'\,\frac{q_{11}}{x_2' - x_2}$	F-K	$m_1 < m_2$ Fig. 6
	$+ \int_1^\infty dx_2'\int_{-\infty}^{1-x_3-x_2'} dx_1'\,\frac{q_{21}}{x_2' - x_2}\Big\}\,\text{Im}\,A(x_1', x_2', x_3)$	F-M	$m_1 > m_2$ Fig. 7
$x_3 > 1$	$F = -\frac{i}{\pi^2}\Big\{\int_1^\infty dx_1'\int_{-\infty}^{1-x_3-x_1'} dx_2'\,\lVert q_{ij}\rVert - \int_{-\infty}^{-1} dx_1'\int_{-\infty}^{1-x_3-x_1'} dx_2'\,\frac{q_{21}}{x_2' - x_2}$	B, J-O	Figs. 9, 10, 11
	$- \int_{-\infty}^{-1} dx_1'\int_{1-x_3-x_1'}^{\infty} dx_2'\,\frac{q_{11}}{x_2' - x_2}\Big\}\,\text{Re}\,A(x_1', x_2', x_3)$		

and (17) may be written

$$F = \frac{1}{\pi}\int_0^{-1} \frac{dr_1'}{r_1' - r_1}\,A_1(r_1', x_2, x_3) + \frac{1}{\pi}\int_{-1}^0 \frac{r_3}{r_3'}\frac{dr_3'}{r_3' - r_3}\,A_3(r_3', x_2, x_3), \qquad (29)$$

when $x_2 > 1$, $x_3 > 1$.

We already know that the singularities of $A_1(r_1, r_2, r_3)$ lie along the real x_2 axis when $x_3 < 1$. Figs. 9, 10, and 11 show that this situation remains true when x_3 increases past 1, except that the M^* eventually become complex if x_3 is increased sufficiently. To obtain a double representation with real integration contours, we must guarantee that the M^* do not introduce complex singularities in A_i. For x_2 in the upper half plane, $A_i(r_i, x_2, x_3)$ is defined by continuation from the real axis of the function tabulated in Table II. This function is not singular at an M_2 when it raises above the real axis. If the lower half x_2 plane is defined by continuation through the real x_2 axis in the domains H or I in Figs. (9), (10), and (11), then $A_{1,3}$ satisfies

$$A_i(r_i, x_2 - i\epsilon, x_3) = -A_i^*(r_i, x_2 + i\epsilon, x_3), \qquad i = 1, 3, \qquad (30)$$

and the A_i so defined are nonsingular in the lower half x_2 plane.

We can now write

$$A_i(r_i, x_2, x_3) = \frac{1}{\pi i}\left[\int_{-\infty}^N + \int_{N^+}^{+\infty}\right]\frac{dx_2'}{x_2' - x_2}\,A_i(r_i, x_2', x_3). \qquad (31)$$

We substitute these dispersion relations into (29) and eliminate x_1 as before. Both terms in Eq. (29) contribute to the region $x_1 > 1$, $x_1 + x_2 + x_3 < 1$. Here the x_1 integration runs from 1 to $1 - x_2 - x_3$ in the first term, and from $1 - x_2 - x_3$ in the second term. We combine these contributions by combining the kernels:

$$\frac{dx_2'}{x_2' - x_2}\left(-\frac{dr_1'}{r_1' - r_1} + \frac{r_3}{r_3'}\frac{dr_3'}{r_3' - r_3}\right) = \lVert q_{ij}\rVert\,dx_1'\,dx_2', \qquad (32)$$

where

$$q_{11} \equiv (x_2 + x_3)/N_1, \qquad q_{12} \equiv (1 - x_1)/N_1, \qquad q_{21} = 1/N_2, \qquad q_{22} = 1/N_2, \qquad (33)$$

$$N_1 = (x_1' - x_1)(x_2 + x_3) - (x_2' - x_2)(x_1 - 1), \qquad N_2 = x_1' - x_1 + x_2' - x_2. \qquad (34)$$

The final result is

$$F = \frac{1}{\pi^2 i}\Big\{\int_1^\infty dx_1'\int_{-\infty}^{1-x_1'-x_3} dx_2'\,\lVert q_{ij}\rVert - \int_{-\infty}^{-1} dx_1'\int_{-\infty}^{1-x_1'-x_3} dx_2'\,\frac{q_{21}}{x_2' - x_2} - \int_{-\infty}^{-1} dx_1'\int_{1-x_1'-x_3}^{\infty} dx_2'\,\frac{q_{11}}{x_2' - x_2}\Big\}\,\text{Re}\,A(x_1', x_2', x_3), \qquad (35)$$

where Re A can easily be taken from Figs. (9), (10), (11), and Table II.

A summary of the double dispersion relations is given in Table III.

Position and Polarization Operators in Relativistic and Nonrelativistic Mechanics

H. BACRY*
Faculté des Sciences, Université de Marseille, Marseille, France
(Received 9 August 1963)

The author shows how the position operator and the polarization operator are obtained in nonrelativistic quantum mechanics from their equivalents in relativistic quantum mechanics. This problem is investigated with the aid of group methods; it is furthermore shown that the two operators cannot be studied separately.

I

IN nonrelativistic quantum mechanics, it is customary to quantize separately the orbital angular momentum **L** and the spin momentum **S**, as two independent quantities. These angular momenta verify the following properties:

(a) **L** can be written as

$$\mathbf{L} = \mathbf{x} \times \mathbf{p}, \tag{1}$$

where **x** is the so-called position operator. The observable **L** obeys the equation

$$\mathbf{L} \times \mathbf{L} = i\mathbf{L}, \tag{2}$$

which can be obtained from the commutation rule,

$$[x^i, p^j] = i\delta^{ij}. \tag{3}$$

The eigenvalues of $\mathbf{L}^2$ can be written in the form $l(l+1)$, where l is an *integer*.

(b) Equation (2) is also obeyed by **S**, and the the eigenvalues of $\mathbf{S}^2$ are also of the form $s(s+1)$ but s is an integer or a *half integer*. The difference between $\mathbf{S}^2$ and $\mathbf{L}^2$ spectra is due to the fact that **S** cannot be written in a form similar to (1).

(c) The total angular momentum

$$\mathbf{M} = \mathbf{L} + \mathbf{S} \tag{4}$$

also obeys the relation

$$\mathbf{M} \times \mathbf{M} = i\mathbf{M}. \tag{5}$$

We note that, in nonrelativistic mechanics, we merely add orbital properties and spin properties since **L** is well defined by Eqs. (1) and (3).

II

In relativistic quantum mechanics, difficulties arise from the fact that the above scheme is no longer correct. There are three ways of investigating such a case:

(a) We can try to obtain some relativistic corrections to the above equations as was done by Thomas[1,2] in connection with the fine-structure spin–precession $\frac{1}{2}$ Thomas factor,[3–5] but it is better to find some rigorous method.

(b) We can use a relativistic equation such as Dirac's, but it is not a general method since the Dirac equation describes only spin-$\frac{1}{2}$ particles and, furthermore, many position and polarization operators can be obtained from Dirac operators by various transformations [Pryce–Foldy–Wouthuysen transformation;[6,7] Chakrabarti transformation[5,8]].

(c) We think that the best method consists in starting from an irreducible representation of the Poincaré group (with a mass $m \neq 0$) in order to define a covariant position operator, in the same manner as a covariant polarization has been defined by Wigner. We shall not use the operator[9]

$$x^\mu = -i(\partial/\partial p_\mu) \tag{6}$$

which obeys

$$[x^\mu, p^\nu] = -ig^{\mu\nu} \tag{7}$$

[a relation[10] which generalizes (3)], because x^μ does not commute with $p^\nu p_\nu$, a center operator. Such a difficulty is sometimes removed by replacing x^0 by a parameter (a c number) and p^0 by the operator $(p^2 + m^2)^{\frac{1}{2}}$, but in such a case, the method is not a covariant one and this objection has recently led us to publish a paper concerning a relativistic position operator for a nonzero mass particle.[11]

Let us first recall the main features of that paper. Consider an irreducible representation of the

* Present address: CERN, Geneva, Switzerland.
[1] L. H. Thomas, Nature **117**, 514 (1926).
[2] L. H. Thomas, Phil. Mag. **3**, 1 (1927).
[3] C. Möller, Tidsskr. Mat. B. 1950.
[4] H. Bacry, Nuovo Cimento **26**, 1164 (1962).
[5] H. Bacry, Thèse; Ann. Phys. (Paris) **8**, 197 (1963).
[6] M. H. L. Pryce, Proc. Roy. Soc. (London) **A195**, 62 (1948).
[7] L. Foldy and S. M. Wouthuysen, Phys. Rev. **96**, 1433 (1954).
[8] A. Chakrabarti (Preprint).
[9] See T. D. Newton and E. P. Wigner, Rev. Mod. Phys. **21**, 400 (1949).
[10] We use the metric defined by $g^{00} = 1$, and $g^{ii} = -1$.
[11] H. Bacry, Phys. Letters **5**, 37 (1963).

Poincaré group with $m \neq 0$, the generators of which are the P_μ's and the $M_{\mu\nu}$'s. If we decompose the tensor $M_{\mu\nu}$ into "electric" and "magnetic" parts with respect to the rest frame, we obtain[12]

$$M_{\mu\nu} = (X \wedge P)_{\mu\nu} + (P \wedge W)^*_{\mu\nu}/m^2, \tag{9}$$

where

$$W_\mu = M^*_{\mu\nu}P^\nu, \tag{10}$$

and

$$X_\mu = (1/2m^2)[M_{\mu\nu}, P^\nu]_+. \tag{11}$$

We can easily prove the relations

$$[X_\mu, X_\nu] = 0, \tag{12}$$

$$[X_\mu, P_\nu] = i[(P_\mu P_\nu/m^2) - g^{\mu\nu}], \tag{13}$$

$$[X_\mu, W_\nu] = i(W_\mu P_\nu/m^2), \tag{14}$$

$$[X_\mu, M_{\lambda\rho}] = i(g_{\mu\lambda}X_\rho - g_{\mu\rho}X_\lambda). \tag{15}$$

The operator W is the polarization operator and is orthogonal to P. The operator X which we have introduced can be considered as the position operator[13] *in the rest frame*, since we have

$$[X^\mu, P_\mu]_+ = 0. \tag{16}$$

III

We first intend to show how we can pass from the relativistic operator X to the nonrelativistic one x obeying (3). First, we note that we can define another position operator without modifying (9) by adding to X a vector operator colinear to P.[13] For instance, we can set

$$Y_\mu = X_\mu + \tfrac{1}{2}[\alpha, P_\mu]_+, \tag{17}$$

where α is a Lorentz scalar. Now, we make Y satisfy the condition

$$Y_\mu n^\mu = 0, \tag{18}$$

where n is a unitary timelike vector; we therefore obtain

$$Y_\mu = X_\mu - \tfrac{1}{2}[X_\alpha n^\alpha, P_\mu/P_\alpha n^\alpha]_+. \tag{19}$$

If we take n as a basis vector, we get

$$Y^0 = 0, \tag{20}$$

$$[Y^i, Y^j] \neq 0, \tag{21}$$

but

$$[\mathbf{Y}, P^0] = i(\mathbf{P}/P^0), \tag{22}$$

$$[Y^i, P^j] = i\delta^{ij}. \tag{23}$$

These are the relations obeyed by the nonrelativistic operators with P^0 as a Hamiltonian.[14]

Conversely, we have to consider that *the logically relativistic generalization of Eqs. (20)–(23) is provided by the relations (12) and (13).*

It is interesting to note that the "magnetic" parts

$$M^{ij}, \quad L^{ij} = X^iP^j - X^jP^i, \quad S^{ij} = (P \wedge W)^{*ij}$$

commute with P^0, and are consequently constants of motion. This is not the case in the Dirac equation, as is well known, except if we transform it under Pryce–Foldy–Wouthuysen or Chakrabarti transformations.[15]

IV

Let us now investigate the case of the polarization vector W. This vector is usually studied in the rest frame ("little group" study); we prefer to consider any frame and to define the vector W' obtained from W by the pure Lorentz transformation which applies P on a timelike vector n; we shall assume for the sake of simplicity, that $P_\mu P^\mu = n_\mu n^\mu = m^2 = 1$. The expression of W' is[5]

$$W'_\mu = W_\mu - [W^\alpha n_\alpha/(1 + P^\alpha n_\alpha)](n_\mu + P_\mu). \tag{24}$$

If n is a basis vector, we obtain the following relations:

$$W'_0 = 0, \tag{25}$$

$$\mathbf{W}' \times \mathbf{W}' = i\mathbf{W}'. \tag{26}$$

Note that this method is strictly equivalent to the little group method.

It is not difficult to obtain the converse of (24),

$$W_\mu = W'_\mu - [W'^\alpha P_\alpha/(1 + P^\alpha n_\alpha)](n_\mu + P_\mu), \tag{27}$$

---

[12] We adopt Einstein's convention for summations. Greek indices run from 0 to 3; Latin indices from 1 to 3. A star denotes a "dual" tensor as follows:

$$M_{\mu\nu}{}^* = \tfrac{1}{2}\epsilon_{\mu\nu\rho\lambda}M^{\rho\lambda}, \tag{8}$$

where $\epsilon_{\mu\nu\rho\lambda}$ equals 1(or -1) if $\mu\nu\rho\lambda$ is an even (or odd) permutation of 0123, and equals zero in all other cases.

[13] Note that according to its definition, it is a Hermitian operator. Equation (13) proves that it is not a "point" operator, but rather a space–time "straight line" operator. In fact, under a translation, X transforms as follows: $X_\mu' = \exp(-ia_\alpha P^\alpha)X_\mu \exp(ia_\alpha P^\alpha) = X_\mu + a_\mu - (a_\alpha P^\alpha)P_\mu$. Therefore, X_μ corresponds to any point of a given straight line, the direction of which is P. (This important remark results from an interesting discussion with Professor J. M. Souriau.)

[14] To prove the relations (20)–(23), we use the two following lemmas:

$$[[A, B]_+, C]_- = [[A, C]_-, B]_+ + [A, [B, C]_-]_+,$$

and

$$[[X^\mu, f(P)]_+, g(P)]_+ = 2[X^\mu, f(P)g(P)]_+,$$

where f and g are two arbitrary functions of P.

[15] See H. Bacry, thesis (reference 5), appendix B; note that the Dirac equation yields to

$$[\mathbf{x}, H] = i\boldsymbol{\alpha} \neq i(p/p^0),$$

where H is the Hamiltonian; consequently $\mathbf{x}$ denotes a point, the trajectory of which is not a straight line in space–time.

a relation which decompose into

$$W_0 = \mathbf{W}'\cdot\mathbf{P}, \tag{28}$$

$$\mathbf{W} = \mathbf{W}' + [\mathbf{W}'\cdot\mathbf{P}/(1 + P^0)]\mathbf{P}. \tag{29}$$

The pseudo vector M^{ij} then becomes, using (9), (19), (28), and (29),

$$\mathbf{M} = \mathbf{Y}\times\mathbf{P} + \mathbf{W}' - \{[\mathbf{P}\times(\mathbf{P}\times\mathbf{W}')]/(1 + P^0)\}. \tag{30}$$

This equation is essential; it proves, in fact, that the total angular momentum splits into two additive parts,[16] provided we neglect the second-order term $\{[\mathbf{P}\times(\mathbf{P}\times\mathbf{W}')]/(1 + P^0)\}$ which contains the well-known $\frac{1}{2}$ Thomas factor.[17]

CONCLUSION

From an irreducible representation of the Poincaré group with a mass $m \neq 0$, *two covariant* vectors can be defined [Eqs. (10) and (11)] which are orthogonal to P. They can each be transformed, in different ways, into vectors orthogonal to a given timelike vector n:

(a) The position operator X by considering the particle in a "retarded position". We obtain the new operator Y.

(b) The polarization operator W by a pure Lorentz transformation. We then get the operator W'.

These two new operators $\mathbf{Y}$ and $\mathbf{W}'$ are the usual nonrelativistic operators in that they obey the usual commutation rules (20), (21), (22), (23), and (26), where P^0 plays the part of the Hamiltonian.

The "total angular momentum" is composed of an orbital part and a proper part as is shown in Eq. (30), together with a relativistic corrective term which can be interpreted as an effect of the Thomas precession. If we neglect this term, we are led to the well-known problem of coupling two angular momenta.

[16] It can be easily seen that these parts are commuting.

[17] See, in reference 4, the link between W' and the Thomas factor.

Motion of Test Bodies in General Relativity*

A. H. TAUB
Digital Computer Laboratory, University of Illinois, Urbana, Illinois and
Institute for Advanced Study, Princeton, New Jersey
(Received 18 April 1963)

Two distributions in the sense of Schwartz are introduced to describe a test body. Papapetrou's equations for a pole–dipole test body are derived in a coordinate-independent manner. The supplementary condition $S^{\sigma\mu}U_\mu = 0$ arises in a natural fashion in the course of the derivation. The method for treating test bodies with higher-order poles is indicated.

1. INTRODUCTION

THE Einstein theory of general relativity relates the metric tensor $g_{\mu\nu}$ of space–time to the matter contained in it by the field equations

$$R^{\mu\nu} - \tfrac{1}{2}g^{\mu\nu}R = -kc^2\theta^{\mu\nu}, \tag{1.1}$$

where $R^{\mu\nu}$ and R are the Ricci tensor and scalar curvature formed from the $g_{\mu\nu}$,

$$k = 8\pi G/c^2 = 1.864 \times 10^{-27} \text{ cm g}^{-1},$$

G is Newton's constant of gravitation, c is the velocity of light of the special theory of relativity, and $\theta^{\mu\nu}$ is the stress–energy tensor of the matter present. It is a consequence of Eqs. (1.1) and the Bianchi identities that

$$\theta^{\mu\nu}_{;\nu} = 0, \tag{1.2}$$

where the summation convention is used and the semicolon denotes the covariant derivative with respect to the metric $g_{\mu\nu}$.

One may consider a family of metric tensors, each member of the family being labeled by the value of a parameter e, that is, one may consider the $g_{\mu\nu}$ as functions of a parameter e as well as functions of coordinates; $g_{\mu\nu}(x; e)$. Associated with each of these metrics there will be a stress tensor $\theta^{\mu\nu}(x; e)$ computed from the right-hand side of Eq. (1.1). We shall denote

$$h_{\mu\nu}(x) = (d/de)g_{\mu\nu}(x; e)|_{e=0}, \tag{1.3}$$

and

$$T^{\mu\nu}(x) = (d\theta^{\mu\nu}/de)(x; e)|_{e=0}. \tag{1.4}$$

We shall further refer to $g_{\mu\nu}(x, 0)$ and $\theta^{\mu\nu}(x, 0)$ as the background metric and stress–energy tensors, respectively. $T^{\mu\nu}(x)$ and $h_{\mu\nu}(x)$ are considered as tensors in the space–time with the background metric, and are said to be the pertubring stress–energy and perturbed metric, respectively. The equations satisfied by these tensors will be derived below from Eqs. (1.1) and (1.2).

The tensor $T^{\mu\nu}$ will also be said to be the stress–energy tensor of a test system. It is part of the purpose of this paper to discuss those consequences of Eqs. (1.1) and (1.2) which involve the tensors $g_{\mu\nu}(x, 0)$ and $T^{\mu\nu}$ but do not involve the tensor $h_{\mu\nu}(x)$. The determination of $h_{\mu\nu}(x)$ and higher derivatives of $g_{\mu\nu}(x, e)$ and $\theta^{\mu\nu}(x, e)$ (evaluated at $e = 0$) will be treated in other papers.

We shall assume, as is the case for a fluid, that the tensor $T^{\mu\nu}$ has a unique timelike proper vector defined by the equations

$$T^{\mu\nu}u_\nu = \rho(1 + \epsilon/c^2)u^\mu, \tag{1.5}$$

$$g_{\mu\nu}u^\mu u^\nu = 1. \tag{1.6}$$

The vector field u^μ will be said to be the velocity vector of the test system. The scalar quantity ρ will be called the rest-mass density of the system and will be defined by the requirement

$$(\rho u^\mu)_{;\mu} = 0. \tag{1.7}$$

This equation is called the equation of conservation of mass. The scalar ϵ is said to be the specific internal energy of the test system.

The solutions of the equations

$$dx^\mu/ds = u^\mu(x) \tag{1.8}$$

define a three-parameter congruence of world lines in space–time,

$$x^\mu = x^\mu(\xi^1, \xi^2, \xi^3; s), \tag{1.9}$$

where

$$x^\mu = x^\mu(\xi^1, \xi^2, \xi^3; 0) \tag{1.10}$$

are the parametric equations of a fixed but arbitrary spacelike hypersurface in space–time. We call this

* This work was supported in part by the National Science Foundation.

hypersurface the initial hypersurface and require that it be intersected by each world line once and only once. The totality of world lines given by Eqs. (1.9) with the ξ^i which enter into those equations restricted to a compact domain in the spacelike hypersurface are said to form a world tube.

Test systems whose stress–energy tensors vanish outside of a particular world tube will be called limited test systems. The size of the system is measured by the size of the compact domain defining the world tube. A test body is a limited test system whose world tube is shrunk to a single world line by a suitable limiting process. This concept of a test body together with a particular definition of the limiting process involved was introduced by Robertson in 1937.[1] He proved, among other things, that a certain type of test body moves along a geodesic in the space–time with tensor $g_{\mu\nu}(x, 0)$.

In this paper we shall define a test body as a test system whose stress–energy tensor and whose rest densities are, respectively, tensor-valued[2] and scalar-valued distributions, in the sense of Schwartz, whose support is (that is, which are nonvanishing on) a single world-line of the congruence given by Eqs. (1.9). It follows from Eqs. (1.5) that, if $T^{\mu\nu}$ is a tensor-valued distribution and u_μ is a regular velocity field, then ρ or $(1 + \epsilon/c^2)$ is a scalar-valued distribution. We assume that ρ is the distribution since we shall wish to treat cases where $\epsilon \equiv 0$. Schwartz has shown[3] that a distribution with a single point as a support is a linear combination of the Dirac data function and a finite number of its derivatives. This theorem will be applied to the description of the tensor and scalar distributions in a family of spacelike hypersurfaces intersecting the world tube described by Eqs. (1.9).

We shall use the following notation: The world line of the family given by Eqs. (1.9) which is the support of the distributions will be assigned the values $\xi^i = \xi^i_0$ and written as

$$X^\mu(s) = x^\mu(\xi^1_0, \xi^2_0, \xi^3_0, s).$$

The unit tangent vector to this world line will be written as

$$U^\mu = U^\mu(s) = u^\mu(x(\xi_0; s)).$$

The invariant volume element in a spacelike hypersurface with unit normal vector n_ρ is

$$n_\rho\, dV = (-g)^{\frac{1}{2}}\epsilon_{\mu\nu\sigma\rho}\frac{\partial x^\mu}{\partial u}\frac{\partial x^\nu}{\partial v}\frac{\partial x^\sigma}{\partial w}\,du\,dv\,dw, \qquad (1.11)$$

where the parametric equations of the hypersurface are given by

$$x^\mu = x^\mu(u, v, w), \qquad (1.12)$$

g is the determinant of the metric tensor, and $\epsilon_{\mu\nu\sigma\tau}$ is the completely alternating tensor density.

In a spacelike hypersurface, the tensor distribution $T^{\mu\nu}$ and the scalar distributions ρ have supports which consist of single points. Hence Schwartz's theorem applies. We shall restrict ourselves to the case where only the Dirac delta function and first derivatives of this function enter. Such test bodies will be shown to be equivalent to the pole–dipole particles treated by Papapetrou.[4] The extension to systems where higher derivatives occur may be made by using the techniques given here.

2. INTEGRAL PROPERTIES OF THE DISTRIBUTIONS

Before giving a coordinate-invariant formulation of the properties of the distributions $T^{\mu\nu}$ and ρ, we discuss these properties in a particular coordinate system. We proceed as Robertson does and introduce a Gaussian coordinate system which has the hypersurface given by Eqs. (1.12), as the hypersurface surface $t = t_0$, and in which the line element has the form

$$ds^2 = dt^2 - h_{ij}\,dx^i\,dx^j \qquad (i = 1, 2, 3). \qquad (2.1)$$

The line element

$$dl^2 = h_{ij}(x^i, t)\,dx^i\,dx^j \qquad (2.2)$$

is positive definite. Moreover, if u, v, and w of Eqs. (1.12) are identified with x^1, x^2, and x^3, respectively, Eq. (1.11) becomes

$$n_\rho\, dV = h^{\frac{1}{2}}\,dx^1\,dx^2\,dx^3\,\delta^4_\rho = \delta^4_\rho\,h^{\frac{1}{2}}\,d^3x, \qquad (2.3)$$

where $h = \det ||h_{ij}||$.

We then require that $T^{\mu\nu}$ and ρ be such that if $f_{\mu\nu}$ is an arbitrary symmetric tensor, and f_μ are arbitrary vector, then in this Gaussian coordinate system,

$$U^4\int T^{\mu\nu}f_{\mu\nu}\,h^{\frac{1}{2}}\,d^3x = M'^{\mu\nu}(t_0)f_{\mu\nu}(X^i, t_0) - M'^{\mu\nu i}(t_0)f_{\mu\nu,i}(X^i, t_0), \qquad (2.4)$$

$$U^4\int \rho u^\mu f_\mu\,h^{\frac{1}{2}}\,d^3x = N'^{\mu}(t_0)f_{\mu}(X^i, t_0) - N'^{\mu i}(t_0)f_{\mu,i}(X^i, t_0). \qquad (2.5)$$

[1] H. P. Robertson, Proc. Edinburgh Math. Soc. 5, 63 (1937).

[2] A Lichnerowicz, *Propagateurs et Commutateurs en Relativité Générale* (Publications Mathematiques No. 10, Institut des Hautes Études Scientifiques, Paris, 1961).

[3] L. Schwartz, *Théorie des Distributions* (Hermann & Cie., Paris, 1950), p. 99, Théorème XXXV.

[4] A. Papapetrou, Proc. Roy. Soc. London **A209**, 248 (1951).

The integrals on the left of these equations are carried out over the three-dimensional hypersurface $t = t_0$, the hypersurface being given in parameteric form by Eqs. (1.12) and which may be any spacelike surface in space–time. The quantities $M'^{\mu\nu}(t_0) = M'^{\nu\mu}(t_0)$, $M'^{\mu\nu i}(t_0) = M'^{\nu\mu i}(t_0)$, $N'^{\mu}(t_0)$, and $N'^{\mu\nu}(t_0)$ depend on this hypersurface. This dependence is indicated by the variable t_0. The functions $f_{\mu\nu}$,

$$f_{\mu\nu,i} = \partial f_{\mu\nu}/\partial x^i, \qquad f_{\mu,i} = \partial f_{\mu}/\partial x^i,$$

and f_μ, occurring on the right-hand side of the above equation are to be evaluated at the point of intersection of the world line X^μ, the support of the distributions and the hypersurface $t = t_0$. The $f_{\mu\nu,i}$ and $f_{\mu,i}$ are the partial derivatives of the functions $f_{\mu\nu}$ and f_μ, the components of an arbitrary tensor and vector, respectively, with respect to the coordinates in the hypersurface.

In case $f_\mu = fu_\mu$, Eq. (2.5) becomes

$$U^4 \int \rho f h^{\frac{1}{2}}\, d^3x = [N'^{\mu}(t_0)U_\mu - N'^{\mu i}U_{\mu,i}] \times f(X^i, t_0) - N'^{\mu i}U_\mu f_{,i}.$$

We shall require that

$$N'^{\mu i}U_\mu = 0, \tag{2.6}$$

and write

$$N'^{\mu}(t_0)U_\mu - N'^{\mu i}U_{\mu,i} = \mu(t_0), \tag{2.7}$$

where $U_{\mu,i}$ is $u_{\mu,i}$ evaluated at the point of intersection of the hypersurface $t = t_0$ with the world line X^μ. Hence we have

$$U^4 \int \rho f h^{\frac{1}{2}}\, d^3x = \mu(t_0) f(X^i, t_0). \tag{2.8}$$

This equation differs in character from Eqs. (2.4) and (2.5) in that the right-hand side does not contain a term involving the derivative of f in the hypersurface. Equations (2.6) which ensure this are imposed in order to take into account the fact that mass dipoles are not to be allowed, as follows from the discussion below [see paragraph before Eq. (2.19)]. We shall see in the sequel that Eqs. (2.6) will provide the three conditions that complete Papapetrou's equations of motion for a pole–dipole test body.

Equations (2.4), (2.5), and (2.8) may be written in a form which holds in any coordinate system. Consider the tensor

$$P^\sigma_\alpha = \delta^\sigma_\alpha - (u^\sigma n_\alpha/u^\rho n_\rho), \tag{2.9}$$

where u^σ is defined by Eqs. (1.5) and (1.6), and n_ρ is defined by Eqs. (1.11). It may be readily verified that

$$P^\sigma_\alpha u^\alpha = 0, \tag{2.10}$$

and that

$$P^\sigma_\alpha P^\alpha_\beta = P^\sigma_\beta. \tag{2.11}$$

Further, in the Gaussian coordinate system used above where $n_\rho = \delta^4_\rho$,

$$\begin{aligned} P^4_\alpha &= \delta^4_\alpha - (u^4\delta^4_\alpha/u^4) = 0, \\ P^i_j &= \delta^i_j, \qquad P^i_4 = -(u^i/u^4). \end{aligned} \tag{2.12}$$

Hence in this coordinate system, the quantities

$$M'^{\mu\nu\sigma} = P^\sigma_\alpha M^{\mu\nu\alpha},$$

where the $M^{\mu\nu\alpha}$ are arbitrary, have the property that

$$M'^{\mu\nu 4} = 0.$$

Equations (2.4) may then be written as

$$\begin{aligned} U^\rho \int T^{\mu\nu} f_{\mu\nu} n_\rho\, dV &= M'^{\mu\nu}(t_0) f_{\mu\nu}(X^i, t_0) - P^\sigma_\alpha M^{\mu\nu\alpha} f_{\mu\nu,\sigma}(X^i, t_0) \\ &= M^{\mu\nu} f_{\mu\nu}(X) - P^\sigma_\alpha M^{\mu\nu\alpha} f_{\mu\nu;\sigma}(X), \end{aligned} \tag{2.13}$$

where

$$M^{\mu\nu} = M'^{\mu\nu} - P^\sigma_\alpha M^{\mu\rho\alpha}\Gamma^\nu_{\rho\sigma} - P^\sigma_\alpha M^{\rho\nu\alpha}\Gamma^\mu_{\rho\sigma}, \tag{2.14}$$

and the Γ's are the Christoffel symbols evaluated at the intersection of the world line X^μ and the hypersurface. Similarly, Eq. (2.5) may be written as

$$U^\rho \int \rho u^\mu f_\mu n_\rho\, dV = N^\mu f_\mu(X) - P^\sigma_\alpha N^{\mu\alpha} f_{\mu;\sigma}(X), \tag{2.15}$$

where

$$N^\mu = N'^{\mu} - P^\sigma_\alpha N^{\rho\alpha}\Gamma^\mu_{\rho\sigma}. \tag{2.16}$$

Equation (2.8) becomes

$$U^\rho \int \rho f n_\rho\, dV = \mu f(X), \tag{2.17}$$

where

$$\mu = N^\mu U_\mu - P^\sigma_\alpha N^{\mu\alpha} U_{\mu;\sigma}. \tag{2.18}$$

Equations (2.13), (2.15), and (2.17) may be evaluated in any coordinate system, and they reduce to Eqs. (2.4), (2.5), and (2.8) in the Gaussian coordinate system. The quantities $M^{\mu\nu}$, $M^{\mu\nu\sigma}$, N^μ, $N^{\mu\sigma}$, and μ in the general coordinate system are functions of the point of intersection of the hypersurface over which the integration is carried out, and the world line which is the support of the distributions. The values of these quantities in a

general coordinate system are obtained from their values in the Gaussian one by the tensor transformation laws indicated by their indices, since the left-hand sides of Eqs. (2.13), (2.15), and (2.17) are scalars, and the $f_{\mu\nu}$, $f_{\mu\nu;\sigma}$, f_μ, $f_{\mu;\sigma}$, and f, occurring on the right-hand sides, are arbitrary tensors.

The quantities $M'^{\mu\nu}$ and $M'^{\mu\nu i}$ entering into Eqs. (2.4) are identical with the quantities labeled as $M^{\mu\nu}$ and $M^{i\mu\nu}$ introduced by Papapetrou, as may be seen as follows: Let us write

$$f_{\mu\nu} = a_{\mu\nu} + b_{\mu\nu i}(x^i - X^i),$$

where $a_{\mu\nu}$ and $b_{\mu\nu i}$ are constants. If this expression is substituted into Eq. (2.4) we obtain

$$U^4\left[a_{\mu\nu}\int T^{\mu\nu}(h^{\frac{1}{2}})\,dV + b_{\mu\nu i}\int T^{\mu\nu}(x^i - X^i)(h^{\frac{1}{2}})\,dV\right] = M'^{\mu\nu}(t_0)a_{\mu\nu} - b_{\mu\nu i}M'^{\mu\nu i}(t_0).$$

On equating the coefficients of $a_{\mu\nu}$ and $b_{\mu\nu i}$ we have

$$M'^{\mu\nu}(t_0) = U^4\int T^{\mu\nu}(h^{\frac{1}{2}})\,dV,$$

$$M'^{\mu\nu i}(t_0) = -U^4\int T^{\mu\nu}(x^i - X^i)(h^{\frac{1}{2}})\,dV,$$

which are Papapetrou's equations (2.7) and (3.5), except for differences in notation. Similarly, if we set

$$f_\mu = u_\mu(a + b_i(x^i - X^i))$$

in Eq. (2.5) and equate coefficients of a and b_i, we find

$$N'^\mu U_\mu - N'^{\mu i}U_{\mu;i} = U^4\int \rho(h^{\frac{1}{2}})\,d^3x,$$

$$N'^{\mu i}U_\mu = U^4\int \rho(x^i - X^i)(h^{\frac{1}{2}})\,d^3x.$$

It is evident from the last equation that the condition expressed by Eqs. (2.6) is the invariant formulation of the requirement that no mass dipoles exist.

Equation (1.5) when applied to the distributions $T^{\mu\nu}$ and ρ may be written in the Gaussian coordinate system as

$$U^4\int T^{\mu\nu}u_\nu f_\mu(h^{\frac{1}{2}})\,d^3x = U^4\int \rho(1 + \epsilon/c^2)u^\mu f_\mu(h^{\frac{1}{2}})\,d^3x,$$

where f_μ is an arbitrary vector field. It then follows from Eqs. (2.4), (2.5), (2.14), and (2.16) that

$$M^{\mu\nu}U_\nu - P^\sigma_\alpha M^{\mu\nu\alpha}U_{\nu;\sigma} = N^\mu(1 + \epsilon/c^2) - P^\sigma_\alpha N^{\mu\alpha}(1 + \epsilon/c^2)_{,\sigma}, \tag{2.19}$$

and that

$$P^\sigma_\alpha M^{\mu\nu\alpha}U_\nu = (1 + \epsilon/c^2)P^\sigma_\alpha N^{\mu\alpha}. \tag{2.20}$$

Equation (2.6) then becomes

$$P^\sigma_\alpha M^{\mu\nu\alpha}U_\nu U_\mu = 0. \tag{2.21}$$

On multiplying Eqs. (2.19) by U_μ, summing and using Eqs. (2.18), we obtain

$$M^{\mu\nu}U_\mu U_\nu - 2P^\sigma_\alpha M^{\mu\nu\alpha}U_{\nu;\sigma}U_\mu = \mu(1 + \epsilon/c^2). \tag{2.22}$$

3. EQUATIONS OF CONSERVATION OF ENERGY AND MOMENTUM

In this section we shall derive differential equations that must be satisfied by the stress–energy tensor $T^{\mu\nu}$ for a general test system. The equations will then be written in integral form and used in the case where the tensor $T^{\mu\nu}$ is a distribution satisfying Eqs. (2.4) or (2.7).

Equations (1.2) which hold for $\theta^{\mu\nu}(x, e)$ may be written as

$$(\partial\theta^{\mu\nu}/\partial x^\nu) + \theta^{\mu\rho}\Gamma^\nu_{\rho\nu} + \theta^{\rho\nu}\Gamma^\mu_{\rho\nu} = 0,$$

where the Γ's are the Christoffel symbols computed from the $g_{\mu\nu}(x, e)$. If these equations are differentiated with respect to e and evaluated at $e = 0$, we obtain

$$(\partial T^{\mu\nu}/\partial x^\nu) + T^{\mu\rho}\mathring{\Gamma}^\nu_{\rho\nu} + T^{\rho\nu}\mathring{\Gamma}^\mu_{\rho\nu} + \theta^{\mu\rho}(x, 0)A^\nu_{\rho\nu} + \theta^{\rho\nu}(x, 0)A^\mu_{\rho\nu} = 0, \tag{3.1}$$

where the $\mathring{\Gamma}$'s are the Christoffel symbols computed from the base metric $g_{\mu\nu}(x, 0)$, and

$$A^\mu_{\rho\nu} = \tfrac{1}{2}g^{\mu\lambda}(h_{\rho\lambda;\nu} + h_{\nu\lambda;\rho} - h_{\rho\nu;\lambda}). \tag{3.2}$$

In this equation the covariant differentiation is with respect to the base metric. The $g^{\mu\lambda}$ entering the equation is the tensor inverse to the base metric $g_{\mu\nu}(x, 0)$.

Equation (3.1) may be written as

$$T^{\mu\nu}_{;\nu} + \theta^{\mu\rho}(x, 0)A^\nu_{\rho\nu} + \theta^{\rho\nu}(x. 0)A^\mu_{\rho\nu} = 0. \tag{3.3}$$

It should be remarked that the last two terms of this equation represent the action of the gravitational field due to the test system on the matter which is responsible for the presence of the base gravitational field represented by the tensor $g_{\mu\nu}(x, 0)$. We shall assume that the tensor $\theta^{\mu\rho}(x, 0)$ vanishes in the world tube of the test system. In that case, Eqs. (3.3) reduce to

$$T^{\mu\nu}_{;\nu} = 0. \tag{3.4}$$

This equation may be written as

$$(T^{\mu\nu}\lambda_\mu)_{;\nu} = T^{\mu\nu}\lambda_{\mu;\nu}, \tag{3.5}$$

where λ_μ is an arbitrary vector field.

We integrate both sides of Eqs. (3.5) over a region of space–time bounded by two spacelike hypersurfaces whose equations in the Gaussian coordinate system are given by $t = t_0$ and $t = t_0 + dt_0$. It then is a consequence of Gauss' theroem that in this coordinate system

$$\frac{d}{dt}\int \lambda_\mu T^{\mu 4}(h^{\frac{1}{2}})\, d^3x = \int \lambda_{\mu;\nu} T^{\mu\nu}(h^{\frac{1}{2}})\, d^3x. \tag{3.6}$$

We shall apply this form of Eqs. (3.4) to the case where $T^{\mu\nu}$ is a distribution satisfying Eqs. (2.4).

It follows from Eqs. (3.6) and (2.4) that

$$\frac{d}{dt}\left(\frac{M'^{\rho 4}}{U^4}\right) + \Gamma^\rho_{\mu\nu}\frac{M'^{\mu\nu}}{U^4} - \frac{M'^{\mu\nu i}}{U^4}\Gamma^\rho_{\mu\nu,i} = 0,$$

$$\frac{d}{dt}\left(\frac{M'^{\rho 4 i}}{U^4}\right) - \frac{M'^{\rho 4}}{U^4}\frac{dX^i}{dt} + \frac{M'^{\rho i}}{U^4} + \frac{M'^{\mu\nu i}}{U^4}\Gamma^\rho_{\mu\nu} = 0,$$

$$M'^{\mu 4 i}\frac{dX^i}{dt} + M^{\mu 4 j}\frac{dX^i}{dt} = M^{\mu i j} + M^{\mu j i}.$$

In deriving these equations we have made use of the fact that

$$\frac{d}{dt} f(X^\nu, t) = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x^i}\frac{dX^i}{dt}.$$

The three equations given above are those given by Papapetrou and may be written as

$$\frac{d}{ds}\left(\frac{M'^{\rho 4}}{U^4}\right) + \Gamma^\rho_{\mu\nu}M'^{\mu\nu} - M'^{\mu\nu\sigma}\Gamma^\rho_{\mu\nu,\sigma} = 0, \tag{3.7}$$

$$\frac{d}{ds}\left(\frac{M'^{\rho 4\sigma}}{U^4}\right) - \frac{M'^{\rho 4}}{U^4}U^\sigma + M'^{\rho\sigma} + M'^{\mu\nu\sigma}\Gamma^\rho_{\mu\nu} = 0, \tag{3.8}$$

$$M'^{\mu 4 i}U^i + M'^{\mu 4 j}U^i = U^4(M'^{\mu i j} + M'^{\mu j i}), \tag{3.9}$$

if it is remembered that in this coordinate system $M'^{\mu\nu 4} = 0$. The twenty-four equations (3.9) for the thirty nonvanishing $M'^{\mu\nu i}$ may be written as

$$P^\alpha_\nu P^\beta_\sigma (M'^{\mu\nu\sigma} + M'^{\mu\sigma\nu}) = 0.$$

It may be readily verified that these equations have the solution given by Papapetrou:

$$M'^{\mu\nu\sigma} = -\tfrac{1}{2}P^\sigma_\alpha(S^{\alpha\mu}U^\nu + S^{\alpha\nu}U^\mu)$$

$$= -\tfrac{1}{2}(S^{\sigma\mu}U^\nu + S^{\sigma\nu}U^\mu + U^\sigma U^\nu v^\mu + U^\sigma U^\mu v^\nu), \tag{3.10}$$

where

$$S^{\alpha\beta} = -S^{\beta\alpha}, \tag{3.11}$$

and

$$v^\mu = S^{\mu\nu}n_\nu/U^\nu n_\nu = S^{\mu 4}/U^4, \tag{3.12}$$

the last equation holding in the Gaussian coordinate system used above.

It is a consequence of Eqs. (3.10) that in this Gaussian coordinate system,

$$M'^{4\rho\sigma} = -\tfrac{1}{2}U^4(S^{\sigma\rho} + U^\sigma v^\rho + v^\sigma U^\rho). \tag{3.13}$$

Thus by taking the antisymmetric part of Eqs. (3.8) we obtain

$$(DS^{\rho\sigma}/Ds) + U^\rho P^\sigma - U^\sigma P^\rho = 0, \tag{3.14}$$

where

$$DS^{\rho\sigma}/Ds = (dS^{\rho\sigma}/ds) + S^{\mu\sigma}\Gamma^\rho_{\mu\nu}U^\nu + S^{\rho\mu}\Gamma^\sigma_{\mu\nu}U^\nu, \tag{3.15}$$

$$P^\sigma = [(M'^{\sigma 4}/U^4) + v^\mu U^\nu \Gamma^\sigma_{\mu\nu}]. \tag{3.16}$$

Multiplying Eqs. (3.13) by $P^\alpha_\rho P^\beta_\sigma$ and summing, we obtain the three equations

$$P^\alpha_\rho P^\beta_\sigma (DS^{\rho\sigma}/Ds) = 0. \tag{3.17}$$

On multiplying Eqs. (3.14) by U_σ, we obtain

$$P^\rho = mU^\rho + U_\sigma(DS^{\rho\sigma}/Ds), \tag{3.18}$$

where

$$m = U_\sigma P^\sigma = [P^4 - U_\sigma(DS^{4\sigma}/Ds)]/U^4. \tag{3.19}$$

Equations (3.18) and (3.19) contain three independent equations, as do Eqs. (3.17). These six equations are equivalent to the six Eqs. (3.14). The remaining independent equations in Eqs. (3.8) may be used to express $M'^{\rho\sigma}$ or $M^{\rho\sigma}$ in terms of P^σ, U^σ, and $S^{\sigma\rho}$.

It follows from Eqs. (3.13) and the symmetrized form of Eqs. (3.8) that

$$2M'^{\rho\sigma} = (d/ds)(U^\sigma v^\rho + v^\rho U^\sigma) + P^\rho U^\sigma + P^\sigma U^\rho + S^{\sigma\mu}U^\nu\Gamma^\rho_{\mu\nu} + S^{\rho\mu}U^\nu\Gamma^\sigma_{\mu\nu}. \tag{3.20}$$

It then follows from Eqs. (2.14) and (3.10) that

$$2M^{\rho\sigma} = \frac{D}{Ds}(U^\sigma v^\rho + U^\rho v^\sigma) + P^\rho U^\sigma + P^\sigma U^\rho$$

$$= U^\sigma\left(P^\rho + \frac{Dv^\rho}{Ds}\right) + U^\rho\left(P^\sigma + \frac{dv^\sigma}{Ds}\right) + v^\rho\frac{DU^\sigma}{Ds} + v^\sigma\frac{DU^\rho}{Ds}, \tag{3.21}$$

where

$$\frac{D}{Ds}U^\sigma = \frac{dU^\sigma}{ds} + U^\mu\Gamma^\sigma_{\mu\nu}U^\nu,$$

and

$$\frac{Dv^\sigma}{Ds} = \frac{dv^\sigma}{ds} + v^\rho\Gamma^\sigma_{\rho\mu}U^\mu.$$

It is evident from Eqs. (3.21) and the tensor char-

acter of the various quantities other than P^ρ entering this equation, that the P^ρ are the components of a vector.

It follows from Eqs. (3.16) and (3.20), that Eqs. (3.7) may be written as

$$(D/Ds)P^\rho + \tfrac{1}{2}S^{\mu\nu}U^\sigma R^\rho_{\mu\sigma\nu} = 0, \qquad (3.22)$$

where $R^\rho_{\mu\sigma\nu}$ is the Riemann Christoffel tensor formed from the $g_{\mu\nu}$, and is evaluated on the world line X^σ as are all the quantities entering into this equation.

We see therefore that the test body which has a stress–energy tensor satisfying (2.4) [or (2.13)] is characterized by the ten independent quantities m, U^ρ, $S^{\rho\sigma}$, where U^ρ satisfies Eq. (1.6). In terms of these quantities, $M^{\mu\nu}$ is defined by Eqs. (3.21) and (3.18), and $M'^{\mu\nu\sigma}$ is given by Eqs. (3.10). The vector P^σ is defined by Eqs. (3.18) and must satisfy Eqs. (3.22). The six quantities $S^{\rho\sigma}$ must satisfy the three Eqs. (3.17), but are otherwise undetermined by the arguments given in this section.

These three equations may be written as

$$\frac{DS^{\alpha\beta}}{Ds} + U^\alpha U_\gamma \frac{DS^{\beta\gamma}}{Ds} - U^\beta U_\gamma \frac{DS^{\alpha\gamma}}{Ds} = 0. \qquad (3.23)$$

The remaining conditions for the determination of $S^{\alpha\beta}$ are provided by Eqs. (2.21). It is a consequence of these equations and Eqs. (3.10) that we must have

$$P^\sigma_\alpha(S^{\alpha\mu}U^\nu + S^{\alpha\nu}U^\mu)U_\mu U_\nu = 0,$$

that is,

$$P^\sigma_\alpha S^{\alpha\mu}U_\mu = 0.$$

In view of Eqs. (2.12), these imply that

$$S^{\alpha\mu}U_\mu = AU^\alpha.$$

However, since $S^{\alpha\mu}$ is antisymmetric, we must have

$$S^{\alpha\mu}U_\mu = 0. \qquad (3.24)$$

These equations have been proposed by Pirani[5] as the equation to supplement Eqs. (3.23). Corinaldesi and Papapetrou[6] have argued that the equations

$$v^\rho = 0,$$

or an equation similar to this, namely $S^{i4} = 0$ where the S^{i4} are evaluated in a particular coordinate system, should be used. The arguments given above indicate that Pirani's conditions are the natural ones from the point of distributions.

[5] F. A. E. Pirani, Acta Phys. Polon. **15**, 389 (1956).
[6] E. Corinaldesi and A. Papapetrou, Proc. Roy. Soc. London **A209**, 259 (1951).

4. CONSERVATION OF MASS

Just as the quantities $M^{\mu\nu}$ and $M^{\mu\nu\sigma}$ characterize the distribution $T^{\mu\nu}$, so the quantities N^μ and $N^{\mu\sigma}$ characterize the distribution ρ. Their properties may be determined by using the equation of conservation of mass, Eq. (1.8), in a manner analogous to the manner in which Eq. (3.4) was used. Because of the existence of Eq. (1.5), we were able to establish relations between the M's and the N's. In this section we shall express the latter quantities in terms of the variables U^μ, $S^{\mu\nu}$, m, and ϵ, and derive some consequences of Eq. (1.8).

It follows from Eqs. (2.22), (3.10), (3.21), and (3.25) that

$$\mu(1 + \epsilon/c^2) = m + (Dv^\rho/Ds)U_\rho + P^\sigma_\alpha S^{\alpha\nu}U_{\nu;\sigma},$$

or

$$\mu(1 + \epsilon/c^2) = m + S^{\sigma\nu}U_{\nu;\sigma}, \qquad (4.1)$$

since

$$v^\rho U_\rho = 0.$$

It may be verified that Eq. (2.20) may be written as

$$(1 + \epsilon/c^2)P^\sigma_\alpha N^{\mu\alpha} = -\tfrac{1}{2}P^\sigma_\alpha S^{\alpha\mu} = -\tfrac{1}{2}(S^{\sigma\mu} + U^\sigma v^\mu), \qquad (4.3)$$

and that Eq. (2.19) becomes

$$(1 + \epsilon/c^2)N^\mu = \frac{1}{2}\left[(m + S^{\sigma\nu}U_{\nu;\sigma})U^\mu + P^\mu + \frac{Dv^\mu}{Ds} - \frac{P^\sigma_\alpha S^{\alpha\mu}}{1 + \epsilon/c^2}\frac{\epsilon_{,\sigma}}{c^2}\right],$$

or

$$(1 + \epsilon/c^2)N^\mu = \mu(1 + \epsilon/c^2)U^\mu + \frac{1}{2}\frac{D}{Ds}\left(\frac{v^\mu}{1 + \epsilon/c^2}\right)(1 + \epsilon/c^2) + \frac{1}{2}\left[U_\sigma\frac{DS^{\mu\sigma}}{Ds} - S^{\sigma\nu}U_{\nu;\sigma}U^\mu - \frac{S^{\sigma\mu}}{1 + \epsilon/c^2}\frac{\epsilon_{,\sigma}}{c^2}\right]. \qquad (4.4)$$

The last equation may be studied further by using Eq. (1.7), the conservation-of-mass equation, which may be written as

$$(\rho u^\mu f)_{;\mu} = \rho u^\mu f_{,\mu}, \qquad (4.5)$$

where f is an arbitrary scalar function. It then follows from Gauss' theorem that in the Gaussian coordinate system we have

$$\frac{d}{dt}\int \rho u^4 f(h^{\frac{1}{2}})\,d^3x = \int (\rho u^\mu f_{,\mu})(h^{\frac{1}{2}})\,d^3x. \qquad (4.6)$$

We shall use this form of the conservation-of-mass equation even when ρ is a distribution.

We evaluate Eqs. (4.6) by means of Eqs. (2.5). We obtain

$$\frac{d}{dt}\left[\left(\frac{N'^4 f}{U^4}\right) - \frac{N'^{4i} f_{,i}}{U^4}\right] = \frac{1}{U^4}\left[N'^{\mu} f_{,\mu} - N'^{\mu i} f_{,\mu i}\right]$$

for arbitrary functions f. Thus we must have

$$d/dt(N'^4/U^4) = 0,$$

$$-\frac{d}{dt}\left(\frac{N'^{4i}}{U^4}\right) + \frac{N'^4}{U^4}\frac{dX^i}{dt} - N'^i = 0, \tag{4.7}$$

and

$$N'^{4i}\frac{dX^i}{dt} + N'^{4j}\frac{dX^i}{dt} = N'^{ij} + N'^{ji}.$$

These equations may be written as

$$dw/dt = 0,$$

$$N'^{\mu} = wU^{\mu} - d/ds(N'^{4\mu}/U^4), \tag{4.8}$$

$$P^{\alpha}_{\sigma}N'^{\sigma\beta} + P^{\beta}_{\sigma}N'^{\sigma\alpha} = 0, \tag{4.9}$$

respectively, where

$$w = N'^4/U^4 = \text{constant},$$

and we have set $N'^{\mu 4} = 0$. Equations (4.9) are satisfied as a consequence of Eqs. (4.3), from which it follows that

$$N'^{4\sigma}/U^4 = -\tfrac{1}{2}[v^{\sigma}/(1 + \epsilon/c^2)].$$

Hence Eqs. (4.8) become

$$N'^{\mu} = wU^{\mu} + \tfrac{1}{2}(d/ds)[v^{\mu}/(1 + \epsilon/c^2)].$$

In view of Eq. (2.16), this may be written as

$$N^{\mu} = wU^{\mu} + \tfrac{1}{2}(D/Ds)[v^{\mu}/(1 + \epsilon/c^2)]. \tag{4.10}$$

It is a consequence of Eqs. (2.18), (4.3), and (4.10) that

$$\mu(1 + \epsilon/c^2) = w(1 + \epsilon/c^2) + \tfrac{1}{2}S^{\sigma\mu}U_{\mu;\sigma}, \tag{4.11}$$

or, as follows from Eq. (4.1),

$$w(1 + \epsilon/c^2) = m + \tfrac{1}{2}S^{\sigma\mu}U_{\mu;\sigma}. \tag{4.12}$$

Equation (4.11) is a consequence of the algebraic structure of Eqs. (4.8) and (4.9), and holds irrespective of the conservation-of-energy equations or Eqs. (1.5). The factor $1 + \epsilon/c^2$ enters because of the latter equations, as does the appearance of the antisymmetric tensor $S^{\sigma\mu}$. In all cases, however, an equation similar to (4.11) may be derived from Eqs. (4.8) and (4.9). This equation will involve the antisymmetric part of $U_{\mu;\sigma}$, which is related to the angular velocity of the test body. It is clear from Eq. (4.11) (or the corresponding equation in the general case) that, although w is a constant, μ will be a constant only under very special circumstances.

Subtracting $(1 + \epsilon/c^2)$ times Eq. (4.10) from Eq. (4.4), and using Eq. (4.11), we obtain

$$\frac{S^{\sigma\mu}}{c^2 + \epsilon^2}\epsilon_{,\sigma} = U_{\sigma}\frac{DS^{\mu\sigma}}{Ds} = S^{\sigma\mu}\frac{DU_{\mu}}{Ds}. \tag{4.13}$$

This equation together with the equation for $d\epsilon/ds$ which will be derived below, may be obtained from the application of Eqs. (2.4) and (2.5) to the integral form of the following consequence of equations (1.5), (1.7), and (3.4):

$$T^{\mu\nu}u_{\mu;\nu} = \rho(\epsilon_{,\mu}/c^2)u^{\mu}.$$

Pirani has pointed out that it is a consequence of Eqs. (3.24) that Eqs. (3.23) may be written as

$$DS^{\alpha\beta}/Ds = (U^{\alpha}S^{\beta\gamma} - U^{\beta}S^{\alpha\gamma})(DU_{\gamma}/Ds). \tag{4.14}$$

Hence

$$(DS^{\alpha\beta}/Ds)(DU_{\beta}/Ds) = 0, \tag{4.15}$$

and

$$\frac{DS^{\alpha\beta}}{Ds}U_{\alpha;\beta} = -S^{\alpha\gamma}\frac{DU_{\gamma}}{Ds}\frac{DU_{\alpha}}{Ds} = 0. \tag{4.16}$$

It follows from Eqs. (3.22) and the symmetry properties of the Riemann tensor that

$$\frac{DP^{\rho}}{Ds}U_{\rho} = \frac{D}{Ds}\left(mU^{\rho} + U_{\sigma}\frac{DS^{\rho\sigma}}{Ds}\right)U_{\rho} = 0.$$

Hence we obtain

$$dm/ds = 0. \tag{4.17}$$

Equation (4.1) then implies that

$$\frac{d}{ds}[\mu(1 + \epsilon/c^2)] = S^{\sigma\nu}(DU_{\nu;\sigma}/Ds). \tag{4.18}$$

If we follow Pirani and introduce the angular momentum four-vector

$$H^{\mu} = \tfrac{1}{2}(-g)^{-\frac{1}{2}}\epsilon^{\mu\nu\rho\sigma}U_{\nu}S_{\rho\sigma},$$

where g is the determinant of the metric tensor and $\epsilon^{\mu\nu\rho\sigma}$ is the alternating tensor density, then Eqs. (4.13) may be written as

$$(DU_{\mu}/Ds) - (\epsilon_{,\mu}/c^2 + \epsilon^2) = AU_{\mu} + BH_{\mu}, \tag{4.19}$$

for any vector V_{μ} satisfying

$$S^{\sigma\mu}V_{\mu} = 0$$

is in the plane of the vectors U_{μ} and H_{μ}. Equation (4.20) may also be written as

$$\left(\delta^{\rho}_{\mu} - \frac{H^{\rho}H_{\mu}}{H^{\sigma}H_{\sigma}}\right)\left\{\frac{D}{Ds}\left[(1 + \epsilon/c^2)U_{\rho}\right] - \frac{\epsilon_{,\rho}}{c^2}\right\} = 0. \quad (4.20)$$

This equation relates the derivatives of ϵ evaluated along the world line to the acceleration vector and the angular-momentum vector.

5. CONCLUSION

The discussion given above shows that the theory of a pole–dipole test body given by Papapetrou may be given a concise and coordinate-independent formulation by the use of Schwartz's theory of distributions. The theory of a single-pole particle given by Robertson is contained as a special case of the above, the case for which $S^{\sigma\mu} = 0$. By using two related distributions, one tensor-valued to describe the energy–momentum properties of the test body, and one scalar-valued to describe the material properties, and by requiring that the latter not show any dipole properties, that is by imposing Eq. (2.6) on the mass distribution, we were able to derive the condition

$$S^{\sigma\mu}U_{\mu} = 0.$$

These equations together with Eqs. (3.14), (3.18), and (3.22), previously given by Papapetrou, are a complete set of equations describing the motion of the pole–dipole test body. These equations have been discussed in some detail by Schiff[7] for the case of the Schwarschild field.

The relations between the vector and tensor characterizing the mass distribution and the tensors characterizing the stress–energy distribution have been given in Sec. 4.

Test bodies with higher-order poles may be treated in a manner analogous to that used above for pole–dipole test bodies. The only modification needed is to introduce additional terms into the right-hand sides of Eqs. (2.4) and (2.5). These terms should involve higher derivatives of $f_{\mu\nu}$ and f_{μ} evaluated on the world line of the test body.

ACKNOWLEDGMENT

It is a pleasure to acknowledge my indebtedness to A. Lichnerowicz who suggested that the Schwartz theorem on distributions used in this paper is connected with the description of test bodies, and with whom I had many fruitful discussions of the results reported in this paper.

[7] L. I. Schiff, Proc. Natl. Acad. Sci. U. S. 46, 871 (1960).

Poles in Coupled Scattering Amplitudes*

A. J. DRAGT
Department of Physics, University of California, Berkeley, California
AND
R. KARPLUS†
Department of Physics and Astronomy, University of Maryland, College Park, Maryland
(Received 12 June 1963)

Elastic scattering amplitudes which have the analytic structure of the Mandelstam representation and which satisfy the unitarity condition and substitution law are severely restricted. Because of these restrictions, poles cannot occur independently and arbitrarily in amplitudes coupled by the unitarity condition. Instead, the locations and residues of poles in coupled amplitudes must satisfy the same relations as do poles in perturbation theory amplitudes.

I. INTRODUCTION

IT has been shown that an elastic scattering amplitude which has the analytic structure given by the Mandelstam representation and satisfies the unitary condition and substitution law[1] is severely restricted.[2] In this note we shall explore a particular example that illustrates these restrictions.

We shall consider the elastic scattering amplitude $f_{AA}(s, t, u)$ for the process

$$A + A \to A + A \tag{1}$$

illustrated in Fig. 1, where the energy–momentum vectors p_i of the spinless particles A (or antiparticles $\bar{A}$) are introduced. The three arguments s, t, and u are defined by the usual relations

$$\begin{aligned} s &= (p_1 + p_2)^2, \qquad t = (p_1 + p_3)^2, \\ u &= (p_2 + p_3)^2, \qquad s + t + u = 4m_A^2, \end{aligned} \tag{2}$$

and m_A is the mass of the particles ($\hbar = c = 1$).

Now we shall assume that f_{AA} has poles with residues Γ at $t = m_B^2$ and at $u = m_B^2 (m_A^2 < m_B^2 < 2m_A^2)$, corresponding to a bound state in the channel $A + \bar{A}$, but has no other poles on the physical sheet. It then follows that the $A + \bar{A}$ states are coupled to $B + B$ states, and that the amplitude $f_{AB}(t, u, s)$ for the process (Fig. 2),

$$A + \bar{A} \to B + B, \tag{3}$$

where t, u, and s are again defined as in Eq. (2), has poles at $u = m_A^2$ and at $s = m_A^2$ on the physical sheet.[3] The residues of f_{AB} at these poles are equal to Γ. The result is trivial in a perturbation theory based on an interaction Lagragian of the form $g\psi_A^\dagger \psi_A \phi_B$. We wish to show that the result also follows if we assume that the amplitudes f_{AA} and f_{AB} are unitary, are analytic (Mandelstam representation), and satisfy the substitution law. The result supplements the work of Polkinghorne, who showed that the general requirements produce scattering amplitudes whose minimum set of singularities is the same as that of a perturbation-theoretic amplitude provided one assumes the existence of all poles known to occur in perturbation theory.[4] This last assumption is no longer necessary, because the occurrence of poles in coupled amplitudes is not independent. We should stress that our conclusions are purely kinematic (independent of the details of the interaction), and give no insight into the reason why f_{AA} has poles in the first place.

In Sec. II, the representations of the amplitudes f_{AA} and f_{AB} are given. In Sec. III, the unitarity condition is applied to f_{AA} in the u and t channels to show that f_{AB} has the expected poles with residue $\pm\Gamma$. The ambiguity of sign is removed in Sec. IV by applying the unitarity condition to f_{AB} in the u and t channels. Secs. III and IV also contain a discussion of the Landau curves nearest the physical regions.

II. REPRESENTATIONS OF AMPLITUDES

We begin with a statement of the regions of analyticity for f_{AA} and f_{AB} that we shall assume. Our assumptions can be most easily expressed by means of a Mandelstam representation. The Man-

* This study was supported in part by the Air Force Office of Scientific Research.
† On leave of absence from the University of California, Berkeley.
[1] J. M. Jauch and F. Rohrlich, *The Theory of Photons and Electrons* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1955), p. 161. The consequences of the substitution law are often referred to as crossing relations.
[2] M. Froissart, Phys. Rev. **123**, 1053 (1961). A. Martin, Phys. Rev. Letters **9**, 410 (1962).
[3] For notational convenience, the variables s, t, u have been ordered in such a way that f_{AA} and f_{AB} are symmetric functions under exchange of their second and third arguments.
[4] J. C. Polkinghorne, Nuovo Cimento **23**, 360 (1962); **25**, 901 (1962). H. P. Stapp, Phys. Rev. **125**, 2139 (1962).

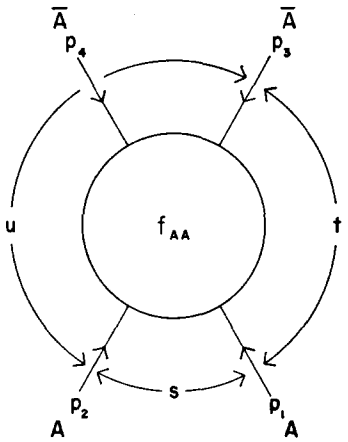


FIG. 1. Lorentz-in-variant variables for A–A scattering.

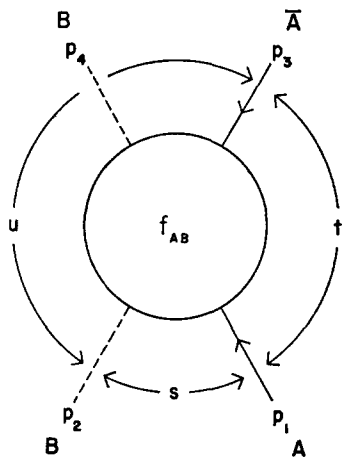


FIG. 2. Lorentz-invariant variables for the reaction $A + \bar{A} \to 2B$.

delstam representation as usually employed is both a statement about the region of analyticity and the behavior at infinity. However, we shall use it only as a statement about the region of analyticity, and all integrals which occur may be interpreted as integrals over a finite interval closed by a large circular contour. The elastic amplitude f_{AA} then has the form

$$\begin{aligned} f_{AA}(s,t,u) &= \Gamma\left(\frac{1}{m_B^2 - t} + \frac{1}{m_B^2 - u}\right) \\ &+ \frac{1}{\pi^2}\iint_{4m_A^2} \frac{\rho(s',t')}{(s'-s)(t'-t)}\,ds'\,dt' \\ &+ \frac{1}{\pi^2}\iint_{4m_A^2} \frac{\rho(s',u')}{(s'-s)(u'-u)}\,ds'\,du' \\ &+ \frac{1}{\pi^2}\iint_{4m_A^2} \frac{\bar{\rho}(t',u')}{(t'-t)(u'-u)}\,dt'\,du' \\ &+ \frac{1}{\pi}\int_{4m_A^2} \frac{\rho_1(s')}{(s'-s)}\,ds' \\ &+ \frac{1}{\pi}\int_{4m_A^2} \frac{\bar{\rho}_1(u')}{(u'-u)}\,du' \\ &+ \frac{1}{\pi}\int_{4m_A^2} \frac{\bar{\rho}_1(t')}{(t'-t)}\,dt', \end{aligned} \tag{4}$$

where the spectral function $\bar{\rho}(t, u)$ is symmetrical,

$$\bar{\rho}(t,u) = \bar{\rho}(u,t). \tag{5}$$

This symmetry property is crucial to our argument. Similarly, the reaction amplitude f_{AB} has the form

$$\begin{aligned} f_{AB}(t,u,s) &= \frac{1}{\pi^2}\int_\tau \frac{dt'}{t'-t}\int_{(m_A+m_B)^2} \frac{ds'}{s'-s}\,r(t',s') \\ &+ \frac{1}{\pi^2}\int_\tau \frac{dt'}{t'-t}\int_{(m_A+m_B)^2} \frac{du'}{u'-u}\,r(t',u') \\ &+ \frac{1}{\pi^2}\iint_{(m_A+m_B)^2} \frac{ds'\,du'}{(s'-s)(u'-u)}\,\bar{r}(s',u') \\ &+ \frac{1}{\pi}\int_\tau \frac{dt'}{t'-t}\,r_1(t') \\ &+ \frac{1}{\pi}\int_\sigma \frac{ds'}{s'-s}\,\bar{r}_1(s') + \frac{1}{\pi}\int_\sigma \frac{du'}{u'-u}\,\bar{r}\,(u'), \end{aligned} \tag{6}$$

where $\bar{r}(s, u)$ is symmetrical;

$$\bar{r}(s,u) = \bar{r}(u,s). \tag{7}$$

The parameters σ, τ are, for the time being, unspecified. It is our intention to place some restrictions on them and on the function $\bar{r}_1$.

III. UNITARITY OF f_{AA}

The physical unitarity condition on f_{AA} for real u in the elastic strip $4m_A^2 < u < 4m_B^2$ is

$$\begin{aligned} &\text{Im } f_{AA}(s_0,t_0,u) \\ &= \left(\frac{u-4m_A^2}{16u}\right)^{\frac{1}{2}}\int_{-1}^{1}\int_{-1}^{1} dx_1\,dx_2 K(x_0,x_1,x_2) \\ &\quad\times f_{AA}(s_1,t_1,u) f_{AA}^*(s_2,t_2,u), \end{aligned} \tag{8}$$

where

$$\begin{aligned} x_i &= 1 + [2t_i/(u-4m_A^2)] \\ &= -1 - [2s_i/(u-4m_A^2)], \end{aligned} \tag{9}$$

and

$$K(x_0,x_1,x_2) = \frac{\theta(1-x_0^2-x_1^2-x_2^2+2x_0x_1x_2)}{(1-x_0^2-x_1^2-x_2^2+2x_0x_1x_2)^{\frac{1}{2}}}. \tag{10}$$

Now we must substitute the representation of f_{AA} [Eq. (4)] into Eq. (8). This operation is carried out more simply in two steps, if we first introduce a one-dimensional representation for the amplitude. Thus we write

$$\begin{aligned} f_{AA}(s,t,u) &= \frac{\Gamma}{t-m_B^2} + \frac{1}{\pi}\int_{4m_A^2} \frac{ds'}{s'-s}\,A_s(s',t',u) \\ &+ \frac{1}{\pi}\int_{4m_A^2} \frac{dt'}{t'-t}\,A_t(s',t',u), \end{aligned} \tag{11}$$

with functions A that are obtainable by comparison with Eq. (4). By the usual procedure we then

construct the formula for the double spectral function,[5]

$$\bar{\rho}(t_0, u) = \left[\frac{(u - 4m_A^2)}{4u}\right]^{\frac{1}{2}} \iint dx_1\, dx_2 K(x_0, x_1, x_2)$$
$$\times \{A_s(s_1, t_1, u)A_s^*(s_2, t_2, u)$$
$$+ [A_t(s_1, t_1, u) + \pi\Gamma\delta(t_1 - m_B^2)]$$
$$\times [A_t^*(s_2, t_2, u) + \pi\Gamma\delta(t_2 - m_B^2)]\}, \quad (12)$$

where

$$K(x_0, x_1, x_2) = \frac{\theta(x_0 - x_1x_2 - (x_1^2 - 1)^{\frac{1}{2}}(x_2^2 - 1)^{\frac{1}{2}})}{(x_0^2 + x_1^2 + x_2^2 - 2x_0x_1x_2 - 1)^{\frac{1}{2}}}, \quad (13)$$

and the x_i are related to the invariant arguments as in Eq. (9). The structure of this equation is such that the singularities of $\bar{\rho}$ in the real t_0, u plane occur on lines (Landau curves) all of which are asymptotic to the straight line $u = 4m_A^2$. By evaluating the integral we find quite easily that

$$\bar{\rho}(t, u) = 0, \quad t < 4m_B^2 + 4m_B^4/(u - 4m_A^2), \quad (14)$$

and that

$$\bar{\rho}(t, u) = \pi^2\Gamma^2(ut)^{\frac{1}{2}} \times [(u - 4m_A^2)(u - 4m_B^2) - 4m_B^4]^{\frac{1}{2}}, \quad (15)$$

$$4m_B^2 + \frac{4m_B^4}{u - 4m_A^2} < t < m_B^2 + 4m_A^2 + \frac{8m_A^2m_B^2 + 4m_Am_B[u^2 + u(m_B^2 - 4m_A^2)]^{\frac{1}{2}}}{u - 4m_A^2}.$$

Beyond this region, the double spectral function no longer depends on the poles only. The two leading Landau curves we have found are

$$t = 4m_B^2 + 4m_B^4/(u - 4m_A^2), \quad (16)$$

and

$$t = m_B^2 + 4m_A^2 + \frac{8m_A^2m_B^2 + 4m_Am_B[u^2 + u(m_B^2 - 4m_A^2)]^{\frac{1}{2}}}{u - 4m_A^2}. \quad (17)$$

These two Landau curves, of course, extend beyond the value $u = 4m_B^2$, because the elastic contribution to the absorptive part remains. The total double spectral function, however, depends there on the amplitude for transitions to coupled channels as well, so that Eqs. (14) and (15) no longer necessarily hold.

We now examine the unitarity condition in the t channel for $t > 4m_A^2$. The contributions from two-body channels are written explicitly, but the contribution from n-body channels are indicated only schematically as the last term. (These are discussed more extensively in the Appendix.)

$$\text{Im } f_{AA}(s_0, t, u_0) = \left(\frac{t - 4m_A^2}{16t}\right)^{\frac{1}{2}} \iint K(x_0, x_1, x_2)$$
$$\times f_{AA}(s_1, t, u_1)f_{AA}^*(s_2, t, u_2)\, dx_1\, dx_2$$
$$+ \theta(t - 4m_B^2)\left(\frac{t - 4m_B^2}{64t}\right)^{\frac{1}{2}}$$
$$\times \iint K(x_0, x_1', x_2')f_{AB}(t, u_1', s_1')f_{AB}^*(t, u_2', s_2')\, dx_1'\, dx_2'$$
$$+ \theta(t - (2m_A + m_B)^2) \sum_n \int f_{An}f_{An}^*. \quad (18)$$

Here the angular and invariant variables are related by the equations

$$x_i = 1 + \frac{2u_i}{t - 4m_A^2} = -1 - \frac{2s_i}{t - 4m_A^2}, \quad (19a)$$

$$x_i' = \frac{t + 2s_i' - 2m_A^2 - 2m_B^2}{[(t - 4m_A^2)(t - 4m_B^2)]^{\frac{1}{2}}} = -\frac{t + 2u_i' - 2m_A^2 - 2m_B^2}{[(t - 4m_A^2)(t - 4m_B^2)]^{\frac{1}{2}}}. \quad (19b)$$

We are especially interested in the consequences of Eq. (18) in the strip $4m_A^2 < u_0 < 4m_B^2$. There we have considerable information about $\bar{\rho}(t, u)$ and therefore about the imaginary part

$$\text{Im } f_{AA}(s_0, t, u_0) = \frac{1}{\pi}\int_{4m_A^2} \frac{\rho(s', t)}{s' - s_0}\, ds' + \frac{1}{\pi}\int_{4m_A^2} \frac{\bar{\rho}(t, u')}{u' - u_0}\, du' + \bar{\rho}_1(t). \quad (20)$$

In particular, we wish to identify the term in Eq. (18) that gives rise to the singularity of Im f_{AA} on the Landau curve, Eq. (16). The contribution of elastic scattering to Eq. (18) can be obtained through an equation like Eq. (12), in which the variables u and t have been interchanged. It then follows from Eq. (16) and (17) that this contribution does not give rise to any singularities of Im $f_{AA}(s_0, t, u_0)$ in the strip we are considering.

The contributions of the n-particle production amplitudes are important in the strip, but they give rise only to Landau curves that are asymptotic to the n-body thresholds (see Appendix). The curve in Eq. (17) is an example determined by a three-body production amplitude; it is asymptotic to $t = (2m_A + m_B)^2$.

We are, therefore, left with the term

[5] S. Mandelstam, Phys. Rev. 115, 1741 (1959).

$$\text{Im } f_{AA}^{(2B)}(s_0, t, u_0) = \theta(t - 4m_B^2)[(t - 4m_B^2)/64t]^{\frac{1}{2}} \times \iint K(x_0, x_1', x_2') f_{AB}(t, u_1', s_1') \times f_{AB}^*(t, u_2', s_2')\, dx_1'\, dx_2', \quad (21)$$

which must not be zero. It, of course, gives rise to Landau curves asymptotic to $t = 4m_B^2$. It is again advantageous to introduce a one-dimensional representation, this time for f_{AB},

$$f_{AB}(t, u, s) = \frac{1}{\pi}\int_\sigma \frac{ds'}{s' - s} B(t, u', s') + \frac{1}{\pi}\int_\sigma \frac{du'}{u' - u} B(t, s', u'), \quad (22)$$

and to extract the spectral function $\bar{\rho}^{(2B)}(t, u)$ of Im $f_{AA}^{(2B)}$. In view of the preceding discussion, $\bar{\rho}^{(2B)}(t, u) = \bar{\rho}(t, u)$ below the Landau curve, Eq. (17). Hence[6]

$$\bar{\rho}^{(2B)}(t, u_0) = \left(\frac{t - 4m_B^2}{4t}\right)^{\frac{1}{2}} \times \iint_1^\infty dx_1'\, dx_2' K(x_0, x_1', x_2') \times B(t, u_1', s_1') B^*(t, u_2', s_2'), \quad (23)$$

$$u_0 < 4m_A^2 + \frac{16m_A^2 m_B^2 t}{[t - (2m_A + m_B)^2][t - (2m_A - m_B)^2]}.$$

The leading Landau curves of this function are

$$u_0 = 4\sigma + 4(\sigma + m_B^2 - m_A^2)^2/(t - 4m_B^2), \quad (24)$$

and others determined by the singularities of $B(t, u, s)$. We see immediately that we must have

$$\sigma = m_A^2, \quad (25)$$

so that Eqs. (24) and (16) are equivalent. To learn more, we use Eq. (25) and write the expression for $B(t, u, s)$ that follows from Eq. (6),

$$B(t, u, s) = \frac{1}{\pi}\theta[s - (m_A + m_B)^2] \times \left[\int_\tau \frac{dt'}{t' - t} r(t', s) + \frac{1}{2}\int_{(m_A+m_B)^2} \frac{du'}{u' - u}\bar{r}(s, u')\right] + \theta(s - m_A^2)\bar{r}_1(s). \quad (26)$$

[6] The region defined by the inequality in Eq. (23) is bounded by the Landau curve of Eq. (17). A more symmetric form for the Landau curve and its attached Landau surface is given by the equation

$$q(m_A, m_A, u)q(2m_A, m_B, t) = m_A m_B,$$

where $q(m_1, m_2, u)$ is the relative momentum in the center-of-mass system of two particles having masses m_1, m_2 and a total energy $u^{\frac{1}{2}}$. This way of expressing the Landau surface displays that it is accessible only from corresponding half-planes of the u and t variables.

We now use the additional information about the explicit form of $\bar{\rho}(t, u)$ we have from Eq. (15). Because of the θ function in the kernel of Eq. (23), only the last term of $B(t, u, s)$ [Eq. (26)] contributes to the value of $\bar{\rho}^{(2B)}(t, u)$ in the regions of validity of Eqs. (15) and (23). We conclude that

$$\bar{\rho}^{(2B)}(t, u_0) = \left(\frac{t - 4m_B^2}{4t}\right)^{\frac{1}{2}} \iint_1^\infty dx_1'\, dx_2' K(x_0, x_1', x_2') \times \theta(s_1' - m_A^2)\bar{r}_1(s_1')\theta(s_2' - m_A^2)\bar{r}_1(s_2') = \frac{\pi^2\Gamma^2\theta[(u_0 - 4m_A^2)(t - 4m_B^2) - 4m_B^4]}{\{u_0 t[(u_0 - 4m_A^2)(t - 4m_B^2) - 4m_B^4]\}^{\frac{1}{2}}}, \quad (27)$$

$$u_0 < 4m_A^2 + \frac{16m_A^2 m_B^2 t}{[t - (2m_A + m_B)^2][t - (2m_A - m_B)^2]}.$$

Suppose that $\bar{r}_1(s)$ is singular at $s = M^2$, $M > m_A$. Then combining this singularity with that produced by the θ function at $s = m_A^2$, we conclude that $\bar{\rho}^{(2B)}(t, u_0)$ must be singular on the Landau curve given by

$$u_0^2(t - 4m_B^2) - 2u_0[2(m_B^2 - m_A^2)^2 + 2M^2 m_A^2 + (t - 2m_A^2 - 2m_B^2)(M^2 + m_A^2)] + (t - 4m_A^2)(M^2 - m_A^2)^2 = 0. \quad (28')$$

This Landau curve has as asymptotes $t = 4m_B^2$ and $u_0 = (M + m_A)^2$. However, there are no Landau curves that approach values of u between $4m_A^2$ and $(2m_A + m_B)^2$ as t becomes large, and we conclude that $\bar{r}_1(s)$ must be analytic in the interval $m_A^2 < s < (m_A + m_B)^2$. Our conclusion is valid provided we assume the curves given by Eq. (28) do not intersect any of the Landau curves associated with the n-body channels of Eq. (18) in the region $t > (2m_A + m_B)^2$, $u_0 > (2m_A + m_B)^2$. If a curve of Eq. (28) does intersect a Landau curve associated with one of the n-body channels, the possibility exists that the part of the curve which extends beyond the intersection and is asymptotic to $u = (M + m_A)^2$ may no longer be singular on the physical sheet. That this can indeed occur has been shown explicitly by Gribov and Dyatlov for the case of a ladder diagram of three rungs in which one of the end boxes occurring in the diagram has an anomalous threshold.[7] Our assumption of no intersection will enable us to prove that f_{AB} has a normal threshold in the s channel, and our results are therefore consistent with theirs. The example of Gribov and Dyatlov clearly illustrates the need

[7] V. N. Gribov and I. T. Dyatlov, Zh. Eksperim. i Teor. Fiz. 42, 1268 (1962) [English transl.: Soviet Phys.—JETP 15, 879 (1962)].

for an understanding of inelastic processes in any meaningful formulation of an S-matrix program.[8]

We now make the ansatz

$$\bar{r}_1(s) = \theta(s - m_A^2)[\pi\Gamma\delta(s - m_A^2) + R(s)], \quad s < (m_A + m_B)^2 \tag{29}$$

to insert into Eq. (27). The first term is chosen to exactly reproduce the known double spectral function. The remainder therefore has to produce a double spectral function which vanishes identically:

$$0 = \iint_1^\infty dx_1\, dx_2 K(x_0, x_1, x_2) R(s_1)\theta(s_1 - m_A^2) \times [R(s_2) + 2\pi\Gamma\delta(s_2 - m_A^2)]\theta(s_2 - m_A^2). \tag{30}$$

But now the left-hand side of the equation is analytic in the entire region

$$u_0 < 4m_A^2 + \frac{16m_A^2 m_B^2 t}{[t - (2m_A + m_B)^2][t - (2m_A - m_B)^2]}. \tag{31}$$

It follows that the integral in Eq. (30), as a function of u_0 for fixed t, must be analytic also. But the integral is not analytic if both factors $R(s_1)\theta(s_1 - m_A^2)$ and $[R(s_2) + 2\pi\Gamma\delta(s_2 - m_A^2)] \times \theta(s_2 - m_A^2)$ have a singularity in the interval $0 < s < (m_A + m_B)^2$. Hence one of the two factors must be analytic; because of the occurrence of the step function, this factor has to be identically zero. If we put this result back into the ansatz Eq. (29), we obtain the final conclusion

$$\bar{r}_1(s) = \pm\pi\Gamma\delta(s - m_A^2). \tag{32}$$

In other words, $f_{AB}(t, u, s)$ has simple poles at $u = m_A^2$ and at $s = m_A^2$; the single spectral integrals in these two channels may be taken with the threshold $(m_A + m_B)^2$ if these poles are included explicitly. In view of footnote 8, our result that f_{AB} has simple poles at $u = m_A^2$ and $s = m_A^2$ is independent of our assumption concerning the nonintersection of two-body and n-body Landau curves. However, our assumption is necessary to raise the threshold of the single spectral integrals beyond $1.8\ m_A^2$ to $(m_A + m_B)^2$.

It is also clear that the condition $m_A < m_B$ assumed in the introduction is essential for our argument. This inequality assures that there is an energy interval in which two A particles scatter elastically. Only in this energy interval does the unitarity condition not couple f_{AA} to other amplitudes. Hence one obtains complete information about the double spectral function from f_{AA} alone.

[8] To study the behavior of $\bar{r}_1(s)$ near $s = m_A^2$, it is sufficient to consider only those Landau curves of Eq. (28) which lie entirely outside the n-body inelastic region $t, u > (2m_A + m_B)^2$. For these curves, there is no possibility of intersection with an n-body Landau curve. We then find, independently of our assumption, that $\bar{r}_1(s)$ is analytic in the interval $m_A^2 < s < 1.8\ m_A^2$.

IV. UNITARITY OF f_{AB}

The physical unitarity condition on $f_{AB}(t, u, s)$ for real u in the elastic strip $(m_A + m_B)^2 < u < (m_A + 2m_B)^2$ is

$$\mathrm{Im}\, f_{AB}(t_0, u, s_0) = \frac{q_{AB}}{2(u)^{\frac{1}{2}}} \iint dx_1\, dx_2 K(x_0, x_1, x_2) \times f_{AB}(t_1, u, s_1) f_{AB}^*(t_2, u, s_2), \tag{33}$$

where

$$q_{AB} = \{[u - (m_A + m_B)^2] \times [u - (m_A - m_B)^2](4u)^{-1}\}^{\frac{1}{2}}, \tag{34}$$

and

$$x_i = 1 + t_i/2q_{AB}^2 = -1 - \{s_i - [(m_A^2 - m_B^2)^2/u](2q_{AB}^2)^{-1}\}. \tag{35}$$

Now we must substitute the representation of f_{AB} [Eq. (6)] into Eq. (33). As before, we first introduce a one-dimensional representation for the amplitude. Thus we write

$$f_{AB}(t, u, s) = \pm[\Gamma/(m_A^2 - s)] + \beta(u) + \frac{1}{\pi}\int_{(m_A+m_B)^2} \frac{ds'}{s' - s} B_s(t', u, s') + \frac{1}{\pi}\int_\tau \frac{dt'}{t' - t} B_t(t', u, s'), \tag{36}$$

with functions β and B that can be obtained by comparison with Eqs. (6) and (32). Again, by the usual procedure we obtain the double spectral function $r(t, u)$,

$$r(t_0, u) = \frac{q_{AB}}{u^{\frac{1}{2}}} \iint dx_1\, dx_2 K(x_0, x_1, x_2) \times \{\theta(t_1 - \tau)B_t(t_1, u, s_1)\theta(t_2 - \tau)B_t^*(t_2, u, s_2) + [\pm\pi\Gamma\delta(s_1 - m_A^2) + B_s(t_1, u, s_1) \times \theta(s_1 - (m_A + m_B)^2)][\pm\pi\Gamma\delta(s_2 - m_A^2) + B_s(t_2, u, s_2)\theta(s_2 - (m_A + m_B)^2)]\}, \tag{37}$$

where the x_i are as defined in Eq. (35). The singularities of r in the real t–u plane occur on Landau curves that are asymptotic to $u = (m_A + m_B)^2$. The leading Landau curve of the first term in the

integrand of Eq. (37) is

$$t_0 = 4\tau + \tau^2/q_{AB}^2. \tag{38}$$

The leading Landau curve of the second term is

$$t_0 = 4m_A^2 + m_B^2\{2m_A^2 - m_B^2 + 2[(m_B^2 - m_A^2)^2/u]\}(q_{AB}^2)^{-1}. \tag{39}$$

We find that the curve is monotonically asymptotic to $t_0 = 4m_A^2$ when $m_B^2 < 2m_A^2$, but that the curve has a minimum,

$$t_0 = 4m_A^2 - (2m_A^2 - m_B^2)^2/m_A^2, \tag{40}$$

when $m_B^2 > 2m_A^2$. Such behavior is characteristic of an anomalous threshold on the physical sheet in the t channel. We see that the development of the anomalous threshold follows unambiguously from our assumptions and does not require the use of analytic continuation with respect to a mass variable. In order to avoid any difficulties associated with an anomalous threshold in t, we shall continue to maintain the inequality $m_B^2 < 2m_A^2$, thus enabling us to set $\tau = 4m_A^2$. With this proviso, the leading Landau curve of Eq. (37) comes from the two pole terms in the integrand. These terms contribute an amount

$$r(t_0, u) = \pi^2\Gamma^2\theta(D)/2(t_0 u D)^{\frac{1}{2}}, \tag{41}$$

where

$$D = q_{AB}^2 t_0 + 4q_{AB}^2(u - m_A^2 - 2m_B^2) - (u - m_A^2 - 2m_B^2)^2. \tag{42}$$

Note that only Γ^2 appears and that there is therefore no ambiguity of sign.

As a result of crossing symmetry, it is also possible to calculate $r(t_0, u)$ by applying unitarity in the t channel for $4m_A^2 < t < 4m_B^2$. We find

$$\text{Im } f_{AB}(t, u_0, s_0) = \left(\frac{t - 4m_A^2}{16t}\right)^{\frac{1}{2}} \iint dx_1\, dx_2 K(x_0, x_1, x_2) \times f_{AA}(s_1, t, u_1) f_{AB}^*(t, u_2, s_2), \tag{43}$$

where

$$x_1 = 1 + 2u_1/(t - 4m_A^2) = -1 - 2s_1/(t - 4m_A^2),$$

$$x_2 = \frac{2s_2 + t - 2m_A^2 - 2m_B^2}{[(t - 4m_A^2)(t - 4m_B^2)]^{\frac{1}{2}}} = -\frac{2u_2 + t - 2m_A^2 - 2m_B^2}{[(t - 4m_A^2)(t - 4m_B^2)]^{\frac{1}{2}}}. \tag{44}$$

Taking into account the poles of f_{AB} at $s_2 = m_A^2$ and $u_2 = m_A^2$ and the pole of f_{AA} at $\mu = m_B^2$, we obtain for the double spectral function

$$r(t, u_0) = \pm\pi^2\Gamma^2\theta(D)/2(tu_0 D)^{\frac{1}{2}}. \tag{45}$$

In order to achieve agreement with Eq. (41), we must choose the plus sign, and the sign of Γ is thus determined.

V. CONCLUSION

In this note we have given an example of the restrictions imposed upon elastic scattering amplitudes by the unitarity condition, analyticity, and the substitution law. We have found that, due to these restrictions, poles appearing in amplitudes coupled by the unitarity condition are not independent and must satisfy the same relations as do poles in perturbation theory amplitudes. In addition, we have attempted to determine the thresholds for single spectral integrals once the pole terms have been explicitly removed. In doing so, we were forced to make some restrictive assumptions about the behavior of n-body inelastic Landau curves. This circumstance clearly illustrates the necessity for an understanding of inelastic processes in any meaningful formulation of an S-matrix program.

APPENDIX

In this appendix, we shall discuss the behavior of certain Landau surfaces of the singularities of the absorptive part of an elastic scattering amplitude f_{AA} for physical energies $\omega > 2m_A$. Let us consider a particular reaction amplitude f_{An} for the transition between a two-particle state $A + \bar{A}$ and an n-particle state $n \geq 3$, whose particles have masses m_i. If we denote the contribution of f_{An} to the imaginary part of f_{AA} by $\text{Im } f_{AA}^{(n)}$, we can write schematically

$$\text{Im } f_{AA}^{(n)} = \int f_{An} f_{An}^*. \tag{A1}$$

We now note that the reaction will be dominated by the low-angular-momentum states at threshold if the interaction is of short range. Hence we expect that the partial-wave series for $\text{Im } f_{AA}^{(n)}$ will converge in an ever expanding ellipse as the total energy ω approaches the production threshold. In other words, the Landau surface of the singularities of $\text{Im } f_{AA}^{(n)}$ is asymptotic to the threshold surface $\omega = \sum_1^n m_i = M_n$.

To be more precise, we introduce coordinates to describe the n-particle state in the energy range slightly above threshold, where all particles move nonrelativistically. We use momenta $\mathbf{k}_i$ conjugate to the $n - 1$ Jacobian coordinates to define the

internal behavior of the n-particles, and Eulerian angles ϕ, θ, ψ to define the orientation of the internal axes with respect to the direction of the incident A particles in the overall center-of-mass system. The amplitude is then independent of the angle ϕ. We can write

$$f_{An} = \sum_{l,\lambda} (2l + 1) D^{l}_{0\lambda}(\phi, \theta, \psi) f^{l}_{\lambda}(k_i), \qquad \text{(A2)}$$

where the D functions are the usual elements of the rotation matrices. The total kinetic energy is

$$\sum_{1}^{n-1} \frac{k_i^2}{2m_i} = W - M_n, \qquad \text{(A3)}$$

which depends on the appropriate reduced masses. Alternatively, we could use the internal angular momenta l_i associated with the Jacobian coordinates,

$$f^{l}_{\lambda}(\mathbf{k}_i) = \sum f^{l,l_1,l_2,\cdots}_{\lambda,\lambda_1,\lambda_2,\cdots}(k_i) \prod_{i=1}^{n-1} Y^{\lambda_i}_{l_i}(\theta_i, \phi_i). \qquad \text{(A4)}$$

We can now rewrite Eq. (A1) more explicitly,

$$\begin{aligned} \text{Im } f^{(n)}_{AA}(\omega, \cos\theta) &= \sum_{ll'\lambda\lambda'} (2l + 1)(2l' + 1) \\ &\times \int \sin\theta_1 \, d\theta_1 \, d\phi_1 \, d\psi_1 \\ &\times D^{l}_{0\lambda}(\phi_1, \theta_1, \psi_1) \, D^{l'*}_{0\lambda'}(\phi_2, \theta_2, \psi_2) \\ &\times \int d^3k_1 \cdots d^3k_{n-1} f^{l}_{\lambda}(\mathbf{k}_i) f^{l'}_{\lambda'}(\mathbf{k}_i) \, \delta(W - \omega). \end{aligned} \qquad \text{(A5)}$$

Here ϕ_1, θ_1, ψ_1 and ϕ_2, θ_2, ψ_2 are the Eulerian angles relative to the incident and outgoing two-particle momenta, respectively. To carry out the angular integration, we use the addition theorem for the D functions,

$$\begin{aligned} &D^{l'*}_{0\lambda'}(\phi_2, \theta_2, \psi_2) \\ &\quad = \sum_{u=-l'}^{l'} D^{l'}_{u0}(0, \theta, 0) \, D^{l'*}_{u\lambda'}(\phi_1, \theta_1, \psi_1). \end{aligned} \qquad \text{(A6)}$$

The orthogonality properties of the D's make the integration trivial, with the result

$$\begin{aligned} &\text{Im } f^{(n)}_{AA}(\omega, \cos\theta) \\ &= \sum_{l} (2l + 1) P_l(\cos\theta) \sum_{\lambda} \int k_1^2 \, dk_1 \cdots k_{n-1}^2 \, dk_{n-1} \\ &\quad \times \delta(W - \omega) \sum_{\substack{l_1 \cdots l_{n-1} \\ \lambda_1 \cdots \lambda_{n-1}}} |f^{l,l_1,l_2,\cdots}_{\lambda,\lambda_1,\lambda_2,\cdots}(k_i)|^2. \end{aligned} \qquad \text{(A7)}$$

We shall now use the effect of the centrifugal barrier, which reduces the amplitudes for large angular momenta when the wavelength is long compared to the range of interaction. For the wave-number k_i we use the estimate

$$k_i^2 < (\omega^2 - M_n^2), \qquad i = 1, 2, \cdots n - 1. \qquad \text{(A8)}$$

Since all particles are produced within the interaction volume, the individual terms in Eq. (A7) depend on energy as

$$f^{l,l_1,l_2,\cdots}_{\lambda,\lambda_1,\lambda_2,\cdots} \sim k_1^{l_1} k_2^{l_2} \cdots k_{n-1}^{l_{n-1}}, \qquad \text{(A9)}$$

and we may drop all terms except the one with the smallest total exponent,

$$l_1 + l_2 + \cdots + l_{n-1} = l, \qquad \text{(A10)}$$

permitted by the law of addition for angular momenta. We then obtain

$$\begin{aligned} &\text{Im } f^{(n)}_{AA}(\omega, \cos\theta) \\ &\simeq \rho(\omega) \sum_{l} (2l + 1) C_l P_l(\cos\theta) (\omega^2 - M_n^2)^l, \end{aligned} \qquad \text{(A11)}$$

where ρ is an angular-momentum-independent function that describes the phase space for s waves and C_l is an energy-independent coefficient that contains information about the zero-energy properties of the interaction and other factors. As expected, the effects are dominated by the s wave. If we set

$$\lim_{l\to\infty} (C_l)^{1/l} = C^{-1}, \qquad \text{(A12)}$$

then the semimajor axis x_0 of the ellipse of convergence of the partial-wave series for Im $f^{(n)}_{AA}$ is

$$x_0 \geq \tfrac{1}{2}[C(\omega^2 - M_n^2)]^{-1}, \qquad \text{(A13)}$$

which becomes infinite at threshold. The invariant momentum transfer, which is usually used to describe the Landau surfaces, is

$$\begin{aligned} t_0 &= -\tfrac{1}{2}(\omega^2 - 4m_A^2)(1 - x_0) \\ &\gtrsim (\omega^2 - 4m_A^2)[4C(\omega^2 - M_n^2)]^{-1}, \end{aligned} \qquad \text{(A14)}$$

which also approaches infinity at the n-particle threshold.

If the particles have spin, we nevertheless expect the same conclusion to apply, because it depends on the behavior of amplitudes for very large angular momentum. The relative energy dependence of the high-angular-momentum terms is unaffected by finite spin of the particles.

For the contribution of two-body reaction amplitudes, the same result can be established by a simplified procedure because the final states have no internal degrees of freedom.

Approach to a Phase Transition in a One-Dimensional System

EUGENE HELFAND
Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey
(Received 30 July 1963)

The manner in which a one-dimensional system approaches a phase transition as the range of its attractive pairwise interactions becomes large is analyzed, as a continuation of studies on such systems by Kac, Uhlenbeck, Hemmer, and the author. Explicitly, consideration is given to an Ising ferromagnetic model and the corresponding lattice gas with interaction $-\alpha\gamma \exp(-\gamma|i-j|)$ as $\gamma \to 0$. One finds that the fluid isotherms are actually analytic, but that in a certain region they approach the zero slope characteristic of a phase transition in the essentially singular fashion $\exp(-\text{const}/\gamma)$.

I. INTRODUCTION

MOST approximate theories of the statistical thermodynamics of continuum or lattice fluids have had only limited success with regard to questions concerning the phase transition and especially the critical point. Our most detailed knowledge comes from the study of systems amenable to exact treatment, such as two-dimensional lattice models, and from the analysis of expansions when rather long series are available. Recently, exact investigations have been carried out by Kac, Uhlenbeck, Hemmer, and Helfand[1-6] on certain one-dimensional systems which approach phase transitions. The purpose of this paper is to further illuminate the nature of this transition.

The model studied by Kac, Uhlenbeck, and Hemmer[2,3,4] is a one-dimensional fluid of hard rods of length δ which also attract one another with a potential $-\alpha\gamma \exp(-\gamma|x_i - x_j|)$. Kac and Helfand[5] have studied the lattice system wherein each of the N sites in a linear array is occupied by a particle with spin corresponding to $\mu_i = \pm 1$. The interaction between the particles on sites i and j is $-J\mu_i\mu_j\gamma \exp(-\gamma|i-j|)$. (There is a corresponding lattice gas.)

A means of determining the partition function even for finite γ has been presented by Kac,[1] but most of the explicit consideration has been directed at the long-range-force problem, i.e., the limit as $\gamma \to 0$. In this limit one observes a first-order phase transition, with a critical point. The equation of state in the single-phase region is van der Waals-like, having the form

$$p = p_S - \alpha\rho^2, \tag{1}$$

where p_S is the pressure of a fluid with only the short-range forces. The density discontinuity at the transition (rather than van der Waals loops) is explicitly predicted by the theory. The phase transition remains in the theory, apparently, even as one makes corrections in powers of γ. Van Hove[7] tells us, however, that a one-dimensional system with finite range interaction, i.e., finite γ, cannot exhibit a phase transition. This quandary is resolved in the present paper wherein a fuller consideration of the small γ case reveals that terms which go like $\exp(-\text{const}/\gamma)$ are material. In particular, in the "transition region", the slope of the pressure–density isotherm vanishes in this essentially singular fashion as $\gamma \to 0$.

This result is also interesting since one finds that in the neighborhood of the critical point there is a critical region of range $O(\gamma^{\frac{1}{2}})$ in temperature and $O(\gamma^{\frac{1}{2}})$ in density.[4,5] The dilemma encountered in an effort to interpret this region is that unless γ vanishes there is no phase transition, but if γ vanishes there is no region. If the appearance of the phase transition is characterized by the function $\exp(-\text{const}/\gamma)$, one might argue that, by comparison, a region of order γ to a power is large. Actually, the only reasonable lines along which this anomaly can be resolved is through consideration of two-dimensional systems,[5] where a true phase transition can occur for finite γ. This problem is currently under investigation.

We begin in Sec. II with a review of the formulation applicable to the statistical calculation of the properties of the one-dimensional lattice system with an exponential interaction. The partition function is shown to depend on the lowest eigenvalue

[1] M. Kac, Phys. Fluids 2, 8 (1959).
[2] M. Kac, G. E. Uhlenbeck, and P. C. Hemmer, J. Math. Phys. 4, 216 (1963).
[3] G. E. Uhlenbeck, P. C. Hemmer, and M. Kac, J. Math. Phys. 4, 229 (1963).
[4] P. C. Hemmer, M. Kac, and G. E. Uhlenbeck, J. Math. Phys. 5, 60 (1964).
[5] M. Kac and E. Helfand, J. Math. Phys. 4, 1078 (1963).
[6] E. Helfand, Ann. Rev. Phys. Chem. 14, (1963).
[7] L. Van Hove, Physica 16, 137 (1950).

of a differential operator of the type

$$-(\partial^2/\partial x^2) + U(x, \xi),$$

where the parameter ξ is linearly related to the chemical potential for the lattice gas or to the external field for the ferromagnetic problem. For certain values of temperature and chemical potential, U has two minima separated by a barrier of width $O(\gamma^{-\frac{1}{2}})$ and height $O(\gamma^{-1})$. For $\zeta = 0$, the wells are of equal depth. The transition occurs as ζ passes through zero, since then the roles of the higher- and lower-well interchange, and for $\gamma \to 0$ the lowest eigenfunction is entirely in the lower well. In Sec. III we observe that for γ nonvanishing, the transition from one well to the other takes place more gradually, viz., within a range of change of chemical potential or pressure which goes predominantly as $\exp(-\text{const}/\gamma)$. Explicit calculations are presented of some of the salient properties effected by this modification.

II. REVIEW AND FORMULATION

Consider a one-dimensional lattice of N spins which can be up or down, corresponding to $\mu = \pm 1$. Let the pair interaction be of the form (ferromagnetic)

$$-J\mu_i\mu_j\gamma \exp(-\gamma |i - j|),$$

and the interaction of each spin with an external field H be $-H\mu_i$. Thermodynamic properties are derivable from the partition function

$$Q_N = \{ \sum_{\{\mu\}=\pm 1} \exp [\tfrac{1}{2}\nu\gamma \sum_{i,j} \mu_i\mu_j \times \exp(-\gamma |i - j|) + \zeta \sum_i \mu_i]\} e^{-\frac{1}{2}\nu\gamma N}, \quad (2)$$

where $\nu = J/kT$ is the reduced inverse temperature, and $\zeta = H/kT$ is the reduced field. The self-interaction, i.e., the $i = j$ term, has been included in the summation and then removed with the final factor.

There is a corresponding lattice gas.[8] Consider N sites, each of which may be either occupied by one particle or vacant. There is an attraction, $-\alpha\gamma \exp(-\gamma |i - j|)$ between particles on sites i and j. If we assign $\mu = +1$ to an occupied site and $\mu = -1$ to a vacant one, the interaction energy which we can now associate with the pair of sites i and j is

$$-\alpha \cdot \tfrac{1}{2}(\mu_i + 1) \cdot \tfrac{1}{2}(\mu_j + 1)\gamma \exp(-\gamma |i - j|).$$

Multiplying this out, we get $\mu_i\mu_j$ terms, like the interactions in the ferromagnet problem; μ_i terms, giving effects like those caused by the external field; and constants. In addition, to calculate the grand partition function we must include factors of the fugacity z, raised to the power of the number of occupied sites, which may be written z to the power $\frac{1}{2}\sum_i (\mu_i + 1)$. As a result, the grand partition function is

$$G_N = \exp\{N\zeta - \nu\gamma N[(e^\gamma - 1)^{-1} + \tfrac{1}{2}]\} \times \sum_{\{\mu\}} \exp\left[\frac{\nu}{2} \sum_{ij} \mu_i\mu_j \times \exp(-\gamma |i - j|) + \zeta \sum_i \mu_i\right], \quad (3)$$

where $\nu = \alpha/4kT$ and ζ is now related to the fugacity by

$$\zeta = \tfrac{1}{2} \ln z + 2\nu\gamma(e^\gamma - 1)^{-1}. \quad (4)$$

The computational problem of evaluating the magnetic Q_N and the gas G_N are identical.

The essential innovation in calculation of the partition function, due to Kac,[1] is employment of the mathematical identity

$$\exp[\tfrac{1}{2}\nu\gamma \sum_{ij} \mu_i\mu_j \exp(-\gamma |i - j|)] = E\{\exp[(\nu\gamma)^{\frac{1}{2}} \sum_i X(i)\mu_i]\}, \quad (5)$$

from which one easily shows that

$$Q_N = (2e^{-\frac{1}{2}\nu\gamma})^N E\{\prod_k \cosh[(\nu\gamma)^{\frac{1}{2}} X(k) + \zeta]\}. \quad (6)$$

$E\{\cdots\}$ indicates that one is to take the expectation value over a Gaussian random process, $X(i)$, which is characterized by mean zero,

$$E\{X(i)\} = 0, \quad (7)$$

and covariance

$$E\{X(i)X(j)\} = \exp(-\gamma |i - j|). \quad (8)$$

This stochastic process is the Ornstein–Uhlenbeck process, useful in the study of Brownian motion. It is a Markoff process which is the feature rendering Eq. (6) tractable.

The functional integral involved in taking the expectation value may be written as the ordinary integral

$$Q_N = (2e^{-\frac{1}{2}\nu\gamma})^N \int dx_1 \cdots dx_N \prod_k \cosh[(\nu\gamma)^{\frac{1}{2}} x_k + \zeta] \times W(x_1)P(x_1 | x_2, \gamma)P(x_2 | x_3, \gamma) \cdots P(x_{N-1} | x_N, \gamma), \quad (9)$$

where

$$W(x) = (2\pi)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}x^2), \quad (10)$$

[8] T. D. Lee and C. N. Yang, Phys. Rev. 87, 410 (1952).

and

$$P(x \mid x', \gamma) = [2\pi(1 - e^{-2\gamma})]^{-\frac{1}{2}} \times \exp[-\tfrac{1}{2}(x' - xe^{-\gamma})^2/(1 - e^{-2\gamma})]. \quad (11)$$

The integrand is a chain of two variable functions and as such Eq. (9) may be regarded as a repeated operation of an integral operator, thus expressible in terms of the eigenvalues. The characteristic equation corresponding to this operator is conveniently expressed in the form[5]

$$\exp\{\tfrac{1}{2}\log\cosh[(\nu\gamma)^{\frac{1}{2}}x + \zeta]\} \times \exp\{\gamma[(d^2/dx^2) - \tfrac{1}{4}x^2 + \tfrac{1}{2}]\} \times \exp\{\tfrac{1}{2}\log\cosh[(\nu\gamma)^{\frac{1}{2}}x + \zeta]\}\varphi(x) = \lambda\varphi(x). \quad (12)$$

For N large, the contribution of the highest eigenvalue, λ_0, dominates, and one obtains

$$\lim_{N\to\infty} N^{-1}\log Q_N = \ln 2 - \tfrac{1}{2}\nu\gamma + \ln\lambda_0. \quad (13)$$

All that has been said so far applies independently of the value of γ. The range of the interaction potential is $O(1/\gamma)$ so that the long-range-force problem corresponds to $\gamma \to 0$. In this limit, the operator exponents in Eq. (12) may be combined merely by adding. Corrections, involving commutators of the operators, are of higher order in γ.

Therefore, the eigenvalue problem which must be considered for the long-range-force problem is

$$[-(d^2/dx^2) + U(x, \zeta)]\varphi(x) = \kappa\varphi(x), \quad (14)$$

$$U(x, \zeta) = \tfrac{1}{4}x^2 - \gamma^{-1}\log\cosh[(\nu\gamma)^{\frac{1}{2}}x + \zeta]. \quad (15)$$

λ_0 is related to the lowest eigenvalue κ_0 by

$$\lambda_0 = \exp[\gamma(\tfrac{1}{2} - \kappa_0)]. \quad (16)$$

For $\zeta = 0$ and high enough temperatures, one may seek solutions to this eigenvalue problem by expanding the log cosh function. The equation becomes

$$[-(d^2/dx^2) + \tfrac{1}{4}(1 - 2\nu)x^2 + \tfrac{1}{12}\nu^2\gamma x^4 + \cdots]\varphi = \kappa\varphi, \quad (17)$$

which to lowest order in γ is the Weber equation (Schrödinger's equation for the harmonic oscillator).

For $\nu > \frac{1}{2}$, U has negative curvature at the origin. Actually U now has a pair of minima, as one might suspect from the fact that for large $|x|$, U goes as $\frac{1}{4}x^2$ (cf. Fig. 1). The minima are located at the points $\pm|\eta_0|(2/\gamma)^{\frac{1}{2}}$ where η_0 is the solution of

$$\eta_0(2\nu)^{-\frac{1}{2}} = \tanh[(2\nu)^{\frac{1}{2}}\eta_0] \quad (18)$$

(which has a positive and negative solution of the same magnitude). For finite temperatures below the critical ($\nu_c = J/kT_c = \frac{1}{2}$), the wells which are of equal depth are separated by $O(\gamma^{-\frac{1}{2}})$ and are of depth $O(\gamma^{-1})$. This becomes an infinitely high, infinitely wide barrier as $\gamma \to 0$. The lowest eigenvalue corresponds to two eigenfunctions, one centered about the positive well, the other in the negative well. This degeneracy has been shown to lead to what may be interpreted as two phases.

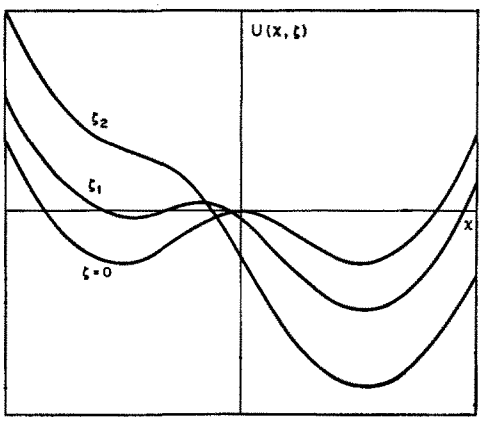


FIG. 1. $U(x, \zeta)$ vs x for three chemical potentials, such that $0 < \zeta_1 < \zeta_2$, and at a fixed subcritical temperature. The scale of x is proportional to $\gamma^{-\frac{1}{2}}$, while that of U is proportional to γ^{-1}.

At the critical point, the quadratic term in (17) vanishes so that the properties of the system are related to the solution of the quartic oscillator. Furthermore, within a temperature region of $O(\gamma^{\frac{1}{2}})$ of the critical temperature, the quartic and quadratic terms of $U(x, 0)$ make contributions of equal order in γ. The manner in which the correlation functions become long range, and the modifications of thermodynamic functions in this region are discussed in references 4 and 5.

In order to achieve an understanding of the lattice gas, the nonzero ζ case must be discussed; i.e., we must consider variations of the chemical potential which will vary the density from its critical value $\frac{1}{2}$. $U(x, \zeta)$ is no longer symmetric in x. As $\gamma \to 0$, the eigenfunction corresponding to the lowest eigenvalue is for the most part localized about the absolute minimum, at the point $\eta_\zeta(2/\gamma)^{\frac{1}{2}}$, where η_ζ is one of the solutions of

$$\eta_\zeta(2\nu)^{-\frac{1}{2}} = \tanh[(2\nu)^{\frac{1}{2}}\eta_\zeta + \zeta]. \quad (19)$$

With the temperature fixed at some subcritical value, consider varying the chemical potential (cf. Fig. 1). For a range of ζ, Eq. (19) will lead to two minima, but one is lower than the other, except for $\zeta = 0$ when the above-mentioned degeneracy exists. For $\gamma \to 0$ as ζ passes through zero, the eigenfunction

will shift from one minimum to the other. This results in a change in the rate of variation of κ with ζ, or in physical terms, a finite change of density at the transition. The major observation of the next section is that for γ small, but still finite, this transformation occurs rapidly but not discontinuously.

Finally, let us examine the application to the lattice gas of the method proposed by Kac, Uhlenbeck, and Hemmer[2] for handling the system in the two-phase region. The central problem is that of which linear combination of the degenerate eigenfunctions to employ.

From the earlier equations one easily finds, as $\gamma \to 0$,

$$\beta p = \ln 2 + \zeta - \nu + \tfrac{1}{2}\gamma(1 - \nu) - \gamma\kappa_0, \tag{20}$$

$$\rho = \tfrac{1}{2}\partial\beta p/\partial\zeta = \tfrac{1}{2} - \tfrac{1}{2}\gamma\partial\kappa_0/\partial\zeta. \tag{21}$$

To determine $\partial\kappa_0/\partial\zeta$, write the eigenfunction as a linear combination of the eigenfunctions centered about the two minima:

$$\begin{aligned} \alpha_1\varphi_0^{(1)} + \alpha_2\varphi_0^{(2)} &= \varphi_0, \\ \alpha_1^2 + \alpha_2^2 &= 1. \end{aligned} \tag{22}$$

Differentiate eigenvalue Eq. (14) with respect to ζ, set $\zeta = 0$, multiply by φ_0, and integrate over x. Making use of the isolation of $\varphi_0^{(1)}$ and $\varphi_0^{(2)}$, and Eq. (18), one finds

$$-\gamma(\partial\kappa_0/\partial\zeta)|_{\zeta=0} = \alpha_1^2\eta^{(1)}(2\nu)^{-\frac{1}{2}} + \alpha_2\eta^{(2)}(2\nu)^{-\frac{1}{2}}. \tag{23}$$

The densities in the pure phases are given by[5]

$$\rho_i = \tfrac{1}{2}[1 + \eta^{(i)}(2\nu)^{-\frac{1}{2}}], \tag{24}$$

from which it follows that

$$\rho = \alpha_1^2\rho_1 + \alpha_2^2\rho_2. \tag{25}$$

Thus the quantities α_i^2 may be interpreted as the volume fraction of phase i. Having established this result it may then be shown[3] that the distribution functions in the two-phase region are linear combinations, weighted with volume fraction, of the distribution functions appropriate to the pure phases.

III. APPROACH TO THE PHASE TRANSITION

Let us proceed to a consideration of the system where the range of the potential, $1/\gamma$, is long but finite. When one contrasts the eigenvalue of a two-well problem with the solutions for two isolated wells, one finds a shift of the eigenvalue and a splitting. In the Appendix, the lowest eigenvalue of Eq. (14), with $\zeta = 0$ and subcritical temperature, is calculated assuming only that the barrier between the wells is high and wide, but not infinite. The solution is achieved by an extension of the method employed by Dennison and Uhlenbeck[9] for the ammonia molecule. $U(x, 0)$ is taken as quadratic in the vicinity of the minima. Locally the solutions are then Weber functions. These are joined by a WKB approximation for the region beyond applicability of the quadratics. The lowest eigenvalue is given by (see Appendix)

$$\kappa_0 = \kappa_0^{(0)} + ks - k\Delta_0.$$

The quantity $\kappa_0^{(0)}$ is the eigenvalue appropriate to an isolated quadratic well. There is a small shift, ks, due to the deviations from this quadratic shape. Part of this shift is calculated in the Appendix. The remainder could be determined, but since this shift is of little physical significance it is not corrected. Finally there is a term $k\Delta_0$, where k^2 is twice the curvature at the bottom of the well,

$$\begin{aligned} k^2 = 2U''(\eta_0(2\nu)^{-\frac{1}{2}}, 0) &= 1 - 2\nu\,\mathrm{sech}^2\,(2\nu)^{\frac{1}{2}}\eta_0, \\ &= 1 - 2\nu + \eta_0^2; \end{aligned} \tag{26}$$

and Δ_0 is given by

$$\Delta_0 = (\pi ek)^{\frac{1}{2}} \exp\left\{-\int_{-t}^{t} [U(x, 0) - U(\eta_0(2\nu)^{\frac{1}{2}}, 0) - \tfrac{1}{2}k]^{\frac{1}{2}}\,dx\right\}, \tag{27}$$

with t the classical turning point. Thus Δ_0 is proportional to the exponential of a quantity of $O(-\gamma^{-1})$. This term arises as a result of the presence of the two wells of equal depth. It is analogous to resonance-splitting terms frequently encountered. There is another eigenvalue $\kappa_0^{(0)} + ks + k\Delta_0$.

Consider slight variation of ζ from zero. The important change that will occur in $U(x, \zeta)$ is that one of the wells will move down slightly by an amount kc,

$$kc = \tfrac{1}{2}\eta_0\gamma^{-1}(2\nu)^{-\frac{1}{2}}\zeta + O(\zeta^2), \tag{28}$$

while the other goes up by the same amount. Slight changes in the locations of the minima and in the well curvature lead to higher-order effects and will be neglected. The solution of the eigenvalue problem with small difference of well depth, detailed in the Appendix, is

$$\kappa = \kappa_0^{(0)} + ks + k\Delta_\zeta, \tag{29}$$

$$\Delta_\zeta = (\Delta_0^2 + c^2)^{\frac{1}{2}} - c. \tag{30}$$

[9] D. M. Dennison and G. E. Uhlenbeck, Phys. Rev. 41, 313 (1932).

The lowest eigenfunction, which in the equal-well case is symmetric, becomes asymmetric by an amount characterized by $R = \Delta_1/\Delta_0$.

The scale of important variations of c is Δ_0, the measure of the equal-well splitting. For $0 < c \ll \Delta_0$, the eigenfunction is nearly symmetric ($R \approx 1$), and almost the entire splitting occurs:

$$\Delta_1/\Delta_0 = 1 - (c/\Delta_0) + \tfrac{1}{2}(c/\Delta_0)^2 + \cdots . \tag{31}$$

However, for $c \gg \Delta_0$, the splitting is almost entirely lost,

$$\Delta_1/\Delta_0 = (\Delta_0/c) - \tfrac{1}{8}(\Delta_0/c)^3 + \cdots , \tag{32}$$

and the wavefunction is almost entirely localized in the lower well.

Thus, the shift in the eigenfunction from one well to the other, with the resulting change in the rate of variation of κ with ζ, is not sudden, but is, rather, continuous, taking place in a range of values of ζ and κ (i.e., chemical potential and pressure) proportional to Δ_0, which has dominant behavior proportional to $e^{-\theta(\nu)/\gamma}$. This is an extremely small interval for small γ, so that the slope in the region, which in the limit becomes the coexistence region, goes as Δ_0. These facts may be expressed quantitatively starting from the pressure–chemical-potential relation in the neighborhood of $\zeta = 0$,

$$\begin{aligned}\beta p &= \ln 2 + \zeta - \nu + \log\cosh\,[(2\nu)^{\frac{1}{2}}\eta_1 + \zeta] \\ &\quad + \gamma(1 - 2\nu + \eta_0^2)^{\frac{1}{2}}\Delta_1 + O(\gamma) + O(\zeta^2),\end{aligned} \tag{33}$$

which yields the density

$$\begin{aligned}\rho &= \tfrac{1}{2}\partial\beta p/\partial\zeta \\ &= \tfrac{1}{2}[1 - \eta_1(2\nu)^{-\frac{1}{2}}] + \gamma(1 - 2\nu + \eta_0^2)^{\frac{1}{2}}\partial\Delta_1/\partial\zeta \\ &\quad + \text{unimportant terms.}\end{aligned} \tag{34}$$

The slope of the p–ρ curve is

$$\begin{aligned}\partial\beta p/\partial\rho &= 2\rho[\tfrac{1}{4}\nu^{-1}\partial\eta_1^2/\partial\zeta \\ &\quad + \gamma(1 - 2\nu + \eta_0^2)\partial^2\Delta_1/\partial\zeta^2]^{-1} + \cdots .\end{aligned} \tag{35}$$

The major contribution to the denominator near the "transition" is

$$\begin{aligned}\partial^2\Delta_1/\partial\zeta^2 &= [(\Delta_0^2 + c^2)^{-\frac{1}{2}} \\ &\quad + c^2(\Delta_0^2 + c^2)^{-\frac{3}{2}}](\partial c/\partial\zeta)^2,\end{aligned} \tag{36}$$

which, for instance, at $\zeta = 0$ is proportional to Δ_0^{-1}, yielding an $O(\Delta_0)$ slope.

It is possible, using the above techniques, to determine the apportionment of the eigenfunction between the two wells. It was shown in the previous section that, in the $\gamma \to 0$ limit, this is related to the volume fraction of the two phases. A similar result can be demonstrated directly in the large, but finite range case; viz., R, which is essentially the ratio of the eigenfunction in the negative x well to that in the positive x well, is related to the numbers of sites we can assign to the two phase by

$$R^2 = N_g/N_l = \text{ratio of vol. fractions.} \tag{37}$$

The apportionment N_g and N_l is made by the conditions

$$N = N_g + N_l, \tag{38}$$

$$N\rho = N_g\rho_g + N_l\rho_l, \tag{39}$$

where ρ_g and ρ_l are assigned the value applicable in the $\gamma \to 0$ limit [cf. Eq. (24) wherein $\eta^{(g)} = -|\eta_0|$ and $\eta^{(l)} = +|\eta_0|$]. Equation (37) is confirmed by substituting the value of R from the Appendix.

APPENDIX

In this Appendix we discuss, in more detail, features of the solution of the basic eigenvalue equation [(14) and (15)].

As a preliminary, consider the $\zeta = 0$ problem. The function $U(x, 0)$ is symmetric about $x = 0$, and below the critical point it has two minima at $\pm x_0$ where $x_0 = |\eta_0|(2/\gamma)^{\frac{1}{2}}$. In the vicinity of the minima, U may be represented by

$$\begin{aligned}U(x, 0) &= U(x_0, 0) + \tfrac{1}{4}k^2(x - x_0)^2, \quad x \approx x_0 \\ &= U(x_0, 0) + \tfrac{1}{4}k^2(-x - x_0)^2, \quad x \approx -x_0,\end{aligned} \tag{A1}$$

where k is given by Eq. (26). We may consider three regions of x values. Region I is from ∞ to some point w, where $w < x_0$ is chosen so that U may be approximated by the quadratic (A1) in Region I. Obviously this approximation also breaks down beyond some $w' > x_0$ but this has only minor consequences, which could be studied by obvious extensions, but will be neglected. In Region II from w to $-w$, the condition $|(\partial U/\partial x)/(U - \kappa)^{\frac{1}{2}}| \ll 1$, necessary for application of the WKB approximation, is valid. A w chosen such that $x_0 - w$ goes as some fractional power of $1/\gamma$ will do. Region III extends from $-w$ to $-\infty$, and again U is approximately quadratic.

The eigenfunction in Region I is

$$\varphi_I(x) = A_I D_n[k^{\frac{1}{2}}(x - x_0)], \tag{A2}$$

where D_n is a Weber function[10] and n is related

[10] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge University Press, Cambridge, England, 1952), p. 347.

to κ by

$$\kappa = U(x_0, 0) + k(n + \tfrac{1}{2}). \tag{A3}$$

Likewise,

$$\varphi_{III} = A_{III} D_n[k^{\frac{1}{2}}(-x - x_0)], \tag{A4}$$

(for the equal-well-depth problem, the lowest state is symmetric). In Region II the WKB eigenfunction is

$$\varphi_{II} = A_{II} p^{-\frac{1}{2}} \cosh\left(\int_0^x p\, dx\right), \tag{A5}$$

$$p = [U(x, 0) - U(x_0, 0) - k(n + \tfrac{1}{2})]^{\frac{1}{2}}. \tag{A6}$$

The condition on n arises by matching the logarithmic derivative of φ_I and φ_{II} at w. It is convenient to employ the asymptotic form for the Weber function,

$$D_n(z) \sim \exp(-\tfrac{1}{4}z^2) z^n - (2\pi)^{\frac{1}{2}}[\Gamma(-n)]^{-1} e^{n\pi i} \times \exp(\tfrac{1}{4}z^2) z^{-n-1} + \cdots .$$

Retaining the largest terms, one finds

$$n = s - \Delta_0. \tag{A7}$$

The shift term s arises from the deviations of U from a single parabolic well. According to the above calculation, it is given by

$$\begin{aligned} s = (2\pi)^{-\frac{1}{2}} &\exp[-\tfrac{1}{2}k(x_0 - w)^2] \\ &\times \{[U(w, 0) - U(x_0, 0) - \tfrac{1}{2}k]^{\frac{1}{2}} \\ &- \tfrac{1}{4}[U(w, 0) - U(x_0, 0) - \tfrac{1}{2}k]^{-1} \\ &\times \partial U(x, 0)/\partial x - \tfrac{1}{2}k^{\frac{1}{2}}(x_0 - w)\}. \end{aligned} \tag{A8}$$

There is another contribution to s from the deviations of U from Eq. (A1) at large positive and negative x, but since we will not find the term s interesting, this correction need not be recorded here. The quantity Δ_0 is related to the splitting which occurs in a two-well system. There is another level corresponding to $n = s + \Delta_0$. Δ_0 is given by

$$\begin{aligned} \Delta_0 = (2/\pi)^{\frac{1}{2}} &[U(w, 0) - U(x_0, 0) - \tfrac{1}{2}k]^{\frac{1}{2}} \\ &\times \exp\left\{\tfrac{1}{2}k(x_0 - w)^2 - \int_{-w}^{w} [U(x, 0) \right. \\ &\left. - U(x_0, 0) - \tfrac{1}{2}k]^{\frac{1}{2}}\, dx\right\}. \end{aligned} \tag{A9}$$

It would be desirable to eliminate w which was incompletely specified. This can be achieved by explicitly employing Eq. (A1) for U in the range $w > |x| > t$, where t is the classical turning point given by

$$U(t) - U(x_0) - \tfrac{1}{2}k = 0. \tag{A10}$$

One then obtains

$$\begin{aligned} \Delta_0 = (\pi e k)^{-\frac{1}{2}} &\exp\left\{-\int_{-t}^{t} [U(x, 0) \right. \\ &\left. - U(x_0, 0) - \tfrac{1}{2}k]^{\frac{1}{2}}\, dx\right\}. \end{aligned} \tag{A11}$$

From the fact that t is $O(\gamma^{-\frac{1}{2}})$ and U is $O(\gamma^{-1})$, the dominant behavior of Δ_0 may be estimated to be $\exp[-\theta(\nu)/\gamma]$, where $\theta(\nu)$ monotonically increases with ν.

To study the phase transition we need consider the ζ (chemical-potential) dependence of the eigenvalue (pressure). Several changes occur in $U(x, \zeta)$ as ζ varies from zero. The locations of the minima change slightly and the curvature at the bottoms of the wells are altered a bit. These are unimportant effects to leading order in γ and ζ, and will be omitted here for simplicity. More pertinently, one well will move up and the other down by an amount

$$\tfrac{1}{2}\eta_0 \gamma^{-1}(2\nu)^{-\frac{1}{2}}\zeta \equiv kc. \tag{A12}$$

Consider the same three regions of x. In Region I, $U(x, \zeta)$ is given approximately by

$$U(x, \zeta) = U(x_0, 0) - kc + \tfrac{1}{4}k^2(x - x_0)^2. \tag{A13}$$

The eigenfunction is

$$\varphi_I = A_I D_n[k^{\frac{1}{2}}(x - x_0)], \tag{A14}$$

where now n and κ are related by

$$\kappa = U(x_0, 0) - kc + \tfrac{1}{2}k(n + \tfrac{1}{2}). \tag{A15}$$

In Region III the local quadratic is

$$U(x, \zeta) = U(x_0, 0) + kc + \tfrac{1}{4}k^2(-x - x_0)^2, \tag{A16}$$

so that

$$\varphi_{III} = A_{III} D_{n-2c}[k^{\frac{1}{2}}(-x - x_0)]. \tag{A17}$$

In the intermediate region, the WKB approximation is

$$\varphi_{II} = \tfrac{1}{2}A_{II} p^{-\frac{1}{2}}\left[\exp\left(\int_0^x p\, dx\right) + R \exp\left(-\int_0^x p\, dx\right)\right]. \tag{A18}$$

Matching logarithmic derivatives at $\pm w$ yields, as values for n and R,

$$n = s - \Delta_\zeta, \tag{A19}$$

$$\Delta_\zeta = R\Delta_0, \tag{A20}$$

$$R\Delta_0 = (\Delta_0^2 + c^2)^{\frac{1}{2}} - c. \tag{A21}$$

Furthermore, to lowest order one finds

$$A_{III}/A_I = R. \tag{A22}$$

Tensor Artificial Viscosity for Numerical Hydrodynamics*

WILLIAM D. SCHULZ
University of California, Lawrence Radiation Laboratory, Livermore, California
(Received 27 February 1963)

Arguments are presented to the effect that a tensor artificial viscosity is a more suitable quantity to use than the usual scalar viscosity. A specific tensor viscosity is constructed for the case of a two-dimensional Lagrangian system.

I. INTRODUCTION

NUMERICAL hydrodynamics consists of constructing a model of a fluid suitable for placing on a digitial computer. In order to integrate hydrodynamic equations by replacing them with finite-difference equations, a mechanism must be introduced to smooth out the discontinuities which occur when shocks are present. Von Neumann and Richtmeyer[1] first solved this problem in one space dimension by introducing an artificial viscosity which spread shock discontinuities out over a specified number of zones. Essentially, they constructed an artificial medium whose final behavior, after experiencing smoothed-out shock waves, was sufficiently like that of the true medium so that it could be used in physical equations in place of the true medium. This was done for a Cartesian-space one-dimensional gas. When people first generalized this system to more complicated gases, the artificial viscosity was taken to be a scalar and treated analogously to the pressure. Both a tensor and a scalar artificial viscosity reduce to the same expression for the Cartesian one-dimensional case, *but not for non-Cartesian one-dimensional cases.* The simple assumption of a tensor artificial viscosity leads to different equations for the cylindrical and spherical one-dimensional cases.

This paper will present the case for a tensor artificial viscosity. The artificial viscosity induced later in this paper has been used successfully by the author for many production problems. In addition to its tensor property, it reduces to a somewhat different, and better, one-dimensional Cartesian-space expression than the basic Richtmeyer–von Neumann–Rosenbluth viscosity.

II. GENERAL COMPARISON

The tensor viscosity to be presented is for a two-space-dimension Lagrangian system. Before going into the detailed arguments for the viscosity, we can present some general arguments for a tensor viscosity.

x^i is the Cartesian coordinate, u^i the velocity comp, P the pressure, q a scalar viscosity (term), Q^{ij} the (stress tensor for) tensor viscosity, ρ the density, τ the specific volume $= \rho^{-1}$, and ϵ the internal energy,

$$\frac{\partial}{\partial t} = \frac{\partial}{\partial t}\bigg| + u^i \frac{\partial}{\partial x^i}$$

is the Lagrangian time derivative. The scalar viscosity equations are

$$\rho(\partial u^i/\partial t) = -(\partial/\partial x^i)(P + q), \tag{2.1}$$

Momentum equation,

$$\partial\epsilon/\partial t = -(P + q)(\partial\tau/\partial t), \tag{2.2}$$

(Int) energy equation.

These are the standard equations and thus require no justification, at least not from the viewpoint of this paper. One may not like artifical viscosities, but that is another matter. The tensor viscosity equations are

$$\rho \frac{\partial u^i}{\partial t} = -\frac{\partial P}{\partial x^i} - \frac{\partial Q^{ji}}{\partial x^j}, \tag{2.3}$$

Momentum equation,

$$\frac{\partial \epsilon}{\partial t} = -P \frac{\partial \tau}{\partial t} - \frac{Q^{ji}}{\rho} \frac{\partial u^i}{\partial x^j}, \tag{2.4}$$

(Int) energy equation.

These equations are new, and thus a few comments are in order. The way Q^{ji} enters the momentum equation is essentially a matter of definition. Total energy conservation then suffices to determine the form of term containing Q^{ji} in the energy equation. That is, let us deduce the equation of total energy conservation. We need

$$\partial\rho/\partial t \,| + \nabla\cdot\rho\mathbf{u} = 0, \tag{2.5}$$

the usual expression for the conservation of mass.

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

[1] J. von Neumann and R. D. Richtmeyer, J. Appl. Phys. **21**, 232 (1950), or see the last chapter of R. D. Richtmeyer's, *Difference Methods for Initial Value Problems* (Interscience Publishers, Inc., New York, 1957).

Expanding the divergence we obtain

$$\partial\rho/\partial t \mid + \rho\nabla\cdot\mathbf{u} = 0, \tag{2.6}$$

substitute $\rho = \tau^{-1}$ and obtain

$$\partial\tau/\partial t \mid - \tau\nabla\cdot\mathbf{u} = 0. \tag{2.7}$$

Let f be an arbitrary function; then consider

$$\rho\frac{\partial f}{\partial t}\bigg| = \frac{\partial(\rho f)}{\partial t}\bigg| - f\frac{\partial\rho}{\partial t}\bigg|;$$

substitute from (2.6) for $\partial\rho/\partial t$ and obtain

$$\rho(\partial f/\partial t) \mid = \partial(\rho f)/\partial t \mid + \nabla\cdot(\rho f\mathbf{u}). \tag{2.8}$$

Define

$$\mathbf{Q}^i = (Q^{1i}, Q^{2i}, Q^{3i}).$$

Multiply the momentum equation (2.3) by u^i, multiply the energy equation (2.4) by ρ and add:

$$\rho\frac{\partial}{\partial t}\left(\frac{\mathbf{u}^2}{2}\right)\bigg| + \mathbf{u}\cdot\nabla P + u^i\nabla\cdot\mathbf{Q}^i + \rho\frac{\partial\epsilon}{\partial t}\bigg| + P\rho\frac{\partial\tau}{\partial t}\bigg| + \mathbf{Q}^i\cdot\nabla u^i = 0.$$

Using (2.7) and (2.8) and rearranging, we obtain

$$\frac{\partial}{\partial t}\left(\rho\frac{\mathbf{u}^2}{2} + \rho\epsilon\right)\bigg| = -\nabla\cdot\left[\left(\rho\frac{\mathbf{u}^2}{2} + \rho\epsilon\right)\mathbf{u} + P\mathbf{u} + u^i\mathbf{Q}^i\right]. \tag{2.9}$$

If we integrate this over a volume, we can use the divergence theorem on the right side. Thus Q^{ii} will only appear in a surface integral and if Q^{ii} vanishes on the boundaries of the system, it will not contribute to the energy conservation expression. The term containing Q^{ii} in the internal energy equation is so chosen that Q^{ii} appears inside a divergence in the total energy conservation equation (2.9).

Let us reduce the previous equations to one-dimensional ones. The one-dimensional scalar viscosity equations are

$$\rho(\partial u^R/\partial t) = -(\partial/\partial R)(P + q), \tag{2.10}$$

$$\partial\xi/\partial t = -(P + q)(\partial\tau/\partial t). \tag{2.11}$$

These equations hold whether R is a Cartesian, cylindrical, or spherical coordinate. The specific volume τ is calculated differently in each of these cases of course.

The reduction of the divergence of a tensor to cylindrical or spherical coordinates is a less familiar process than that required for the above scalar equations. Using the machinery of tensor analysis, let us write the tensor viscosity equations in covariant form. Let $t^{ab\cdots}$ be an arbitrary contravariant tensor, and ξ^l, $l = 1$ to 3, be an arbitrary coordinate system. The covariant derivative of $t^{ab\cdots}$ is

$$t^{ab\cdots};\ l = \frac{\partial t^{ab\cdots}}{\partial\xi^l} + \begin{Bmatrix} a \\ pl \end{Bmatrix} t^{pb\cdots} + \begin{Bmatrix} b \\ pl \end{Bmatrix} t^{ap\cdots} + \cdots,$$

where

$$\begin{Bmatrix} l \\ ik \end{Bmatrix} = \tfrac{1}{2}g^{ls}(g_{is,k} + g_{ks,i} - g_{ik,s}),$$

$$ds^2 = g_{ij}d\xi^i d\xi^l = \text{invariant length}.$$

Also let

$$(\partial t^{ab\cdots}/\partial t) = (\partial t^{ab\cdots}/\partial t)\,|_{\xi^i\text{fixed}} + u^i t^{ab\cdots};j.$$

The covariant tensor viscosity equations are

$$\rho(\partial u^i/\partial t) = -g^{ij}P_{;j} - Q^{ji}_{;j}, \tag{2.12}$$

$$\frac{\partial\xi}{\partial t} = -P\frac{\partial\tau}{\partial t} - \frac{Q^{ji}}{\rho}g_{ji}u^l_{;j}. \tag{2.13}$$

For cylindrical coordinates,

$$\xi^1 = R, \quad \xi^2 = Z, \quad \xi^3 = \theta,$$

$$g_{11} = g_{22} = 1, \quad g_{33} = R^2, \quad g^{ii} = g_{ii}^{-1},$$

$$g_{ij} = g^{ij} = 0, \quad \text{if } i \neq j.$$

For spherical coordinates,

$$\xi^1 = R, \quad \xi^2 = \phi, \quad \xi^3 = \theta,$$

$$g_{11} = 1, \quad g_{22} = R^2\sin^2\theta, \quad g_{33} = R^2, \quad g^{ii} = g_{ii}^{-1},$$

$$g_{ij} = g^{ij} = 0, \quad \text{if } i \neq j.$$

By turning the crank, we can now obtain the desired one-dimensional equations.

Define

$$\delta = \begin{cases} 0 & \text{Cartesian,} \\ 1 & \text{cylindrical,} \\ 2 & \text{spherical.} \end{cases}$$

Then the resulting one-dimensional tensor viscosity equations are

$$\rho\frac{\partial u^R}{\partial t} = -\frac{\partial P}{\partial R} - \frac{1}{R^\delta}\frac{\partial}{\partial R}(R^\delta Q^{RR}), \tag{2.14}$$

$$\frac{\partial\epsilon}{\partial t} = -P\frac{\partial\tau}{\partial t} - \frac{Q^{RR}}{\rho}\frac{\partial u^R}{\partial R}. \tag{2.15}$$

Q^{RR} is all that remains of the tensor Q^{ij}. Its value in the given coordinate system is to be taken the same as that of the scalar q in the previous set (2.10) and (2.11). One could use the standard Richtmeyer–von Neumann expression for it. Its transformation properties, of course, are different.

By integrating the momentum equation (2.14), we can see the first advantage of a tensor viscosity:

$$\int_{R_1}^{R_2} \rho \frac{\partial u^R}{\partial t} R^\delta \, dR = -\int_{R_1}^{R_2} \frac{\partial P}{\partial R} R^\delta \, dR - R^\delta Q^{RR} \Big|_{R_1}^{R_2}.$$

We see that if Q^{RR} vanishes at the boundary points R_1 and R_2, then the artificial viscosity vanishes completely from the equation. In short, we have the proper expression for the sum of the "radial momentum" in the cylindrical and spherical case. The scalar equation (2.10) will only produce the proper expression for the Cartesian case.

The second advantage appears in the energy equation. The entropy of the system is changed by multiplying Q^{RR} by $\partial u^R/\partial R$ instead of $\partial \tau/\partial t$. $\partial u^R/\partial R$ is a function only of the compression of the system in the radial direction as is the case for a true shock. $\partial \tau/\partial t$ will, in general, contain contributions from the general radial convergence or divergence of the system as it moves inward or outward.

III. DEFINITIONS, NOTATION, AND TRANSFORMATION RELATIONS

For convenience, a few relations connecting Eulerian coordinates with Lagrangian coordinates will be discussed first. The artificial viscosity will be defined for a Lagrangian system; however, Eulerian coordinates are the ones which are most familiar in physics. In addition, Eulerian expressions serve as shorthand for longer Lagrangian equivalents.

$R(k, l, q)$ is the Eulerian coordinate; may be Cartesian or cylindrical;
$Z(k, l, q)$ is the Eulerian coordinate; always Cartesian;
$\mathbf{R}$ is the vector (R, Z);
k is the Lagrangian coordinate;
l is the Lagrangian coordinate; and
j is the Jacobian of transformation.

If $R_k = \partial R/\partial k$, $R_l = \partial R/\partial l$, etc., then $j = R_k Z_l - R_l Z_k$ = area Jacobian.

Let

$$\hat{R} = \begin{cases} R \text{ for cylindrical coordinates,} \\ 1 \text{ for Cartesian coordinates.} \end{cases}$$

Then a mass constant may be defined as

$$M = \rho j \hat{R}.$$

Consider the conversion of a Eulerian derivative into Lagrangian derivative:

$$\frac{\partial}{\partial R} = \left(\frac{\partial k}{\partial R}\right)\frac{\partial}{\partial k} + \left(\frac{\partial l}{\partial R}\right)\frac{\partial}{\partial l},$$

$$\frac{\partial}{\partial Z} = \left(\frac{\partial k}{\partial Z}\right)\frac{\partial}{\partial k} + \left(\frac{\partial l}{\partial Z}\right)\frac{\partial}{\partial l}.$$

Expressions for $\partial k/\partial R, \cdots, \partial l/\partial Z$, in terms of $R_k, \cdots, Z_l$ can be found as follows: for arbitrary g,

$$\frac{\partial g}{\partial k} = R_k \frac{\partial g}{\partial R} + Z_k \frac{\partial g}{\partial Z}, \qquad \frac{\partial g}{\partial l} = R_l \frac{\partial g}{\partial R} + Z_l \frac{\partial g}{\partial Z}.$$

Let $g = k$; then one can solve for $\partial k/\partial R$ and $\partial k/\partial Z$. Similarly, $g = l$ gives $\partial l/\partial R$ and $\partial l/\partial Z$. The result of this is

$$\partial k/\partial R = Z_l/j, \qquad \partial l/\partial R = -Z_k/j,$$
$$\partial k/\partial Z = -R_l/j, \qquad \partial l/\partial Z = R_k/j$$

which gives

$$\frac{\partial}{\partial R} = \frac{Z_l}{j}\frac{\partial}{\partial k} - \frac{Z_k}{j}\frac{\partial}{\partial l}, \qquad \frac{\partial}{\partial Z} = -\frac{R_l}{j}\frac{\partial}{\partial k} + \frac{R_k}{j}\frac{\partial}{\partial l}.$$

Define a vector $\bar{\mathbf{R}}$ lagging $\mathbf{R}$ by 90° as the "normal vector" to $\mathbf{R}$:

$$\bar{\mathbf{R}} = (Z, -R) = \text{normal vector to } \mathbf{R}.$$

A useful vector operator can be defined as

$$\mathbf{D} = \frac{1}{j}\left[\bar{\mathbf{R}}_l \frac{\partial}{\partial k} - \bar{\mathbf{R}}_k \frac{\partial}{\partial l}\right]$$
$$= \frac{1}{j}\left[\frac{\partial}{\partial k}(\bar{\mathbf{R}}_l \cdots) - \frac{\partial}{\partial l}(\bar{\mathbf{R}}_k \cdots)\right].$$

Then

$$\nabla f = \mathbf{D}f$$

in both cylindrical and Cartesian coordinates. However,

$$\nabla \cdot \mathbf{f} = \mathbf{D} \cdot \mathbf{f}, \quad \text{Cartesian},$$

$$\nabla \cdot \mathbf{f} = \frac{f_R}{R} + \mathbf{D} \cdot \mathbf{f} = \frac{1}{R}\mathbf{D} \cdot (R\mathbf{f}), \quad \text{cylindrical},$$

where f_R is the Rth component of $\mathbf{f}$.

Langrangian time derivatives, that is, partial derivatives with k and l held fixed, will be written as follows:

$$u(k, l, t) = \partial R/\partial t = R, \quad \text{velocity},$$
$$v(k, l, t) = \partial Z/\partial t = Z, \quad \text{velocity},$$
$$\mathbf{u} \text{ is the vector } (u, v).$$

These relations between u and v, and $\partial R/\partial t$ and $\partial Z/\partial t$ are what defines the k, l coordinates as "Lagrangian" ones. The previously defined Lagrangian derivative can now be deduced. Let g be an arbitrary function,

$$\frac{\partial g}{\partial t} = \frac{\partial g}{\partial t}\Big|_{\mathbf{R}} + \frac{\partial g}{\partial R}\frac{\partial R}{\partial t} + \frac{\partial g}{\partial Z}\frac{\partial Z}{\partial t};$$

thus

$$\partial/\partial t = \partial/\partial t\,|_R + \mathbf{u}\cdot\nabla,$$

as had been specified.

The combined Cartesian–cylindrical equations to which these relations will be applied can be deduced from the covariant equations (2.12) and (2.13). They are

$$\rho\frac{\partial u}{\partial t} = -\frac{\partial P}{\partial R} - \frac{1}{\hat{R}}\frac{\partial}{\partial R}(\hat{R}Q^{RR}) - \frac{\partial Q^{ZR}}{\partial Z},$$
$$\rho\frac{\partial v}{\partial t} = -\frac{\partial P}{\partial Z} - \frac{1}{\hat{R}}\frac{\partial}{\partial R}(\hat{R}Q^{RZ}) - \frac{\partial Q^{ZZ}}{\partial Z}, \tag{3.1}$$

$$\frac{\partial \epsilon}{\partial t} = -P\frac{\partial \tau}{\partial t} - \frac{1}{\rho}\left(Q^{RR}\frac{\partial u}{\partial R} + Q^{ZR}\frac{\partial u}{\partial Z} + Q^{RZ}\frac{\partial v}{\partial R} + Q^{ZZ}\frac{\partial v}{\partial Z}\right). \tag{3.2}$$

Using a little hindsight, let us apply the following transformation to the components of the tensor.

$$Q_{RR} = \frac{1}{j}\left[\frac{\partial R}{\partial k}\frac{\partial Z}{\partial l}q_A - \frac{\partial R}{\partial l}\frac{\partial Z}{\partial k}q_B\right],$$
$$Q_{ZR} = \frac{1}{j}\frac{\partial Z}{\partial k}\frac{\partial Z}{\partial l}(q_A - q_B),$$
$$Q_{RZ} = \frac{1}{j}\frac{\partial R}{\partial k}\frac{\partial R}{\partial l}(q_B - q_A), \tag{3.3}$$
$$Q_{ZZ} = \frac{1}{j}\left[\frac{\partial R}{\partial k}\frac{\partial Z}{\partial l}q_B - \frac{\partial R}{\partial l}\frac{\partial Z}{\partial k}q_A\right].$$

Substituting in the preceeding equations and performing a fair amount of manipulation, there results

$$\frac{\partial \mathbf{u}}{\partial t} = -\frac{\nabla P}{\rho} - \frac{1}{M}\left[\frac{\partial}{\partial k}(\hat{R}\bar{\mathbf{R}}_l q_A) - \frac{\partial}{\partial l}(\hat{R}\bar{\mathbf{R}}_k q_B)\right], \tag{3.4}$$

$$\frac{\partial \epsilon}{\partial t} = -P\frac{\partial \tau}{\partial t} - \frac{q_A(\bar{\mathbf{R}}_l\cdot\mathbf{u}_k)}{\rho j} - \frac{q_B(-\bar{\mathbf{R}}_k\cdot\mathbf{u}_l)}{\rho j}. \tag{3.5}$$

The remainder of this paper will be devoted to a consideration of these relatively simple equations. In point of fact, Eqs. (3.4) and (3.5) were the first deduced from physical arguments as sketched in the next section. Their relation to a formal tensor viscosity, that is, Eqs. (3.3), was deduced afterward.

IV. ARTIFICIAL VISCOSITY

First let us consider the Richtmeyer–von Neumann viscosity. Consider motion in the R direction only, no motion in the Z direction. Under these circumstances, the Lagrangian coordinates can be chosen orthogonal with k and l increasing in the R and Z directions, respectively, i.e.,

$$R_l = Z_k = 0.$$

Assume also a γ-law gas equation of state,

$$P = \xi(\gamma - 1)/\tau.$$

Then, in both the momentum and energy equation, p is replaced by $(p + q_N)$,

$$q_N = \begin{cases} c_0^2\rho(\partial u/\partial k)^2 & \text{if } (\partial u/\partial k) < 0, \\ 0 & \text{otherwise,} \end{cases}$$

$$c_0 = (N/\pi)[\tfrac{1}{2}(\gamma + 1)]^{\frac{1}{2}};$$

N is the distance the shock is spread over in k space ($\Delta k = N$ from beginning of shock to the end) in a Cartesian system. This is all that will be said about the Richtmeyer–von Neumann paper.

The following is a list of some physical conditions one would like an artificial viscosity to satisfy.

(a) A uniform expansion or contraction over the entire medium is considered to correspond to a reversible process (infinitesimal volumes are collapsed with infinitesimal slowness). The artificial viscosity should be independent of such motion.[2]

(b) The velocity component (of the medium) parallel to the shock front should be continuous.

(c) Angular momentum should be conserved.

(d) There should be no artificial viscosity under a velocity field corresponding to a rigid rotation.

The Richtmeyer–von Neumann one-dimensional viscosity does not satisfy condition (a). This is the case where

$$\frac{\partial u}{\partial R} < 0, \qquad \frac{\partial^2 u}{\partial R^2} = 0, \qquad \frac{\partial \tau}{\partial R} = 0;$$

then q_N is unequal to zero.

In place of this basic viscosity, let us use the following. Define

$$(\partial u/\partial k)_- = \min\{(\partial u/\partial k), 0\},$$

and then

$$q = -c_0^2\rho\left(\frac{\partial u}{\partial k}\right)_-\cdot\left|\frac{\partial}{\partial k}\left(\frac{\partial u}{\partial k}\right)_-\right|.$$

For the case of uniform zones, this q will vanish as

[2] The question as to whether the second coefficient of viscosity vanishes or not is apparently still open. That is, perhaps in a uniformly contracting system there is an entropy change. This however is a separate question. The artificial viscosity attempts to take care of shock heating only.

desired for the situation described by condition (a).[3]

In practice, one chooses $c_0^2 = 2$. For q_N, this spreads a simple shock out over three to four zones. That is, $(\partial u/\partial k)_-$ is nonvanishing only over three to four zones. While $\partial/\partial k(\partial u/\partial k_-)$ may be somewhat smaller in magnitude than $(\partial u/\partial k)_-$, it should be of the same order of magnitude. Thus we may hope that q, with the same c_0^2, will also spread a simple shock out over three to four zones. In actual practice, this turns out to be the case. The shape of the shock front, of course, departs from the simple sine wave of the Richtmeyer–von Neumann q_N. The tensor viscosity developed in this paper is thus designed around this new scalar viscosity.

Consider now a region in a general k, l space. $\mathbf{R}_k$ and $\mathbf{R}_l$ are vectors pointing along lines of constant l and k respectively. More significant directions, however, are those given by the normal vectors,

$$\bar{\mathbf{R}}_k = (Z_k, -R_k) = \text{normal vector to } \mathbf{R}_k,$$
$$\bar{\mathbf{R}}_l = (Z_l, -R_l) = \text{normal vector to } \mathbf{R}_l.$$

These normal vectors are the ones that appear in the **D** operator. A scalar viscosity appears in the momentum equation acted on by this operator,

$$\rho \frac{\partial \mathbf{u}}{\partial t} + \mathbf{D}p = \frac{-1}{j} \bar{\mathbf{R}}_l \frac{\partial q}{\partial k} + \frac{1}{j} \bar{\mathbf{R}}_k \frac{\partial q}{\partial l}.$$

Condition (b), however, requires directional properties. As a first attempt at a tensor equation we can write

$$\rho \frac{\partial \mathbf{u}}{\partial t} + \mathbf{D}p = \frac{-1}{j} \bar{\mathbf{R}}_l \frac{\partial q_A}{\partial k} + \frac{1}{j} \bar{\mathbf{R}}_k \frac{\partial q_B}{\partial l},$$

where q_A is a "one-dimensional" viscosity associated with shocks traveling in the $\bar{\mathbf{R}}_l$ direction, and analogously for q_B. Deriving an internal energy equation through the requirement of total energy conservation, one obtains

$$\frac{\partial \epsilon}{\partial t} + P \frac{\partial \tau}{\partial t} = -\frac{q_A}{M} \frac{\partial}{\partial k} (\hat{R}\bar{\mathbf{R}}_l \cdot \mathbf{u}) + \frac{q_B}{M} \frac{\partial}{\partial l} (\hat{R}\bar{\mathbf{R}}_k \cdot \mathbf{u}).$$

This is not a satisfactory energy equation since it is not Galilean invariant. To make it independent of a uniform velocity transformation, we must take $\hat{R}\bar{\mathbf{R}}_l$ and $\hat{R}\bar{\mathbf{R}}_k$ outside of the derivatives over k and l, respectively. Then

$$\frac{\partial \epsilon}{\partial t} + P \frac{\partial \tau}{\partial t} = -\frac{q_A}{\rho j} (\bar{\mathbf{R}}_l \cdot \mathbf{u}_k) - \frac{q_B}{\rho j} (-\bar{\mathbf{R}}_k \cdot \mathbf{u}_l).$$

Going back through the total energy conservation equation once more, there results

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{\nabla P}{\rho} = -\frac{1}{M} \left[\frac{\partial}{\partial k} (\hat{R}\bar{\mathbf{R}}_l q_A) - \frac{\partial}{\partial l} (\hat{R}\bar{\mathbf{R}}_k q_B) \right],$$

and we have Eqs. (3.4) and (3.5).

There remains only the problem of finding suitable expressions for q_A and q_B. For a shock driven in the $\bar{\mathbf{R}}_l$ direction, the previous one-dimensional viscosity can be generalized as follows: Define

$$(\partial u^A/\partial k) = \bar{\mathbf{R}}_l \cdot \mathbf{u}_k / |\bar{\mathbf{R}}_l|,$$

and also

$$(\partial u^A/\partial k)_- = \min \{(\partial u^A/\partial k), 0\};$$

then let

$$q_A = -2\rho \left(\frac{\partial u^A}{\partial k}\right)_- \cdot \left| \frac{\partial}{\partial k} \left(\frac{\partial u^A}{\partial k}\right)_- \right|. \tag{4.1}$$

Analogously,

$$(\partial u^B/\partial l) = (-\bar{\mathbf{R}}_k \cdot \mathbf{u}_l)/|\bar{\mathbf{R}}_k|,$$
$$(\partial u^B/\partial l)_- = \min \{(\partial u^B/\partial l), 0\},$$
$$q_B = -2\rho \left(\frac{\partial u^B}{\partial l}\right)_- \cdot \left| \frac{\partial}{\partial l} \left(\frac{\partial u^B}{\partial l}\right)_- \right|. \tag{4.2}$$

The tensor viscosity defined by Eqs. (3.4), (3.5), (4.1), and (4.2) has been very satisfactory for the class of problems it was designed for. That is, there is a general rule in computing which states that one can always find a situation which your code will not handle properly. Consider the physical conditions (c) and (d) listed earlier. They were chosen because they are conditions that the derived tensor viscosity does not satisfy. Angular momentum conservation requires a symmetric tensor. An examination of Eqs. (3.3) shows immediately that the tensor is not symmetric in general. The difficulty seems to lie in the fact that most problems of interest contain a great many long thin zones. The artificial viscosity used must have the property of spreading shocks over a small distance in the direction of fine zoning, and over a large distance in the direction of gross zoning, i.e., the viscosity is a zone-dependent quantity. This seems to preclude a symmetric tensor. For an Eulerian code with

[3] For nonuniform zones, this q will not vanish as desired. In order to take this into account, the following q was tried originally:

$$(\partial u/\partial R)_- = \begin{cases} \partial u/\partial R \text{ if } \partial u/\partial R < 0, \\ 0 \text{ otherwise,} \end{cases}$$
$$q_{\text{tentative}} = -c_0^2 \rho \left(\frac{\partial u}{\partial R}\right)_- \left| \frac{\partial}{\partial k} \left(\frac{\partial u}{\partial R}\right)_- \right| \left(\frac{\partial R}{\partial k}\right)^2.$$

This q was unsatisfactory. Shocks were propagated through a large zone surrounded by small zones in an unphysical manner.

It should be also noted that, in differencing, the average

$$\left| \frac{\partial}{\partial k} \left(\frac{\partial u}{\partial k}\right)_- \right|_{k-1/2} = \frac{1}{2} \left[\left| \frac{\partial}{\partial k} \left(\frac{\partial u}{\partial k}\right)_- \right|_k + \left| \frac{\partial}{\partial k} \left(\frac{\partial u}{\partial k}\right)_- \right|_{k-1} \right]$$

was taken, and not the absolute value of the average.

uniform zoning in the R and Z directions, there would be no trouble in constructing a symmetric tensor, in fact the previously constructed tensor is symmetric under these circumstances.

Condition (d) refers to a velocity field of the type

$$\mathbf{u} = \bar{\mathbf{R}}.$$

$(\partial u^A/\partial k)$ and $(\partial u^B/\partial l)$ do not vanish in general under this type of field. This could probably be remedied, but since this problem seems related to the problem of angular momentum conservation, it did not seem worth while to fix this one, and not be able to do anything about the other.

One final comment: From both analysis and experience, the shock stability criterion of the difference equations resulting from using the described tensor viscosity is less severe than that deduced from a scalar viscosity. Thus, a tensor viscosity leads to no additional difficulties as far as stability of the difference equations is concerned.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 1 JANUARY 1964

Motion of a Relativistic Damped Oscillator*

R. A. STRUBLE AND T. C. HARRIS
North Carolina State University, Raleigh, North Carolina
(Received 28 June 1963; revised September 1963)

Quantitative effects of damping in a nonlinear system are illustrated for the case of a relativistic oscillator. The elementary techniques employed furnish details of the decay phenomena which are more extensive than those normally furnished by stability theorems.

1. INTRODUCTION

A SO-CALLED relativistic oscillator describes the oscillatory motion of a particle where the rest mass has been replaced by the relativistic mass. Though the resulting model may be somewhat artificial from a physical point of view, it has been the subject of considerable interest and from a mathematical viewpoint, it affords a good example in which one might compare the quantitative effects of damping in a linear and a nonlinear system. Pignedoli[1] has studied cases in which the force acting upon the particle is a function of position only and in which the force is a function of velocity only. The former is notable in that it includes the analog of the classical harmonic oscillator which has been treated by Penfield and Zatskis[2] and MacColl.[3] Other studies have been made by Gheorghita,[4] Eftimiu and Klarsfeld,[5] and Dragos.[6]

In this note we present a brief analysis of the relativistic damped harmonic oscillator where the force acting on the point mass is linearly dependent upon both the displacement and the velocity. Though this case is a combination of those treated by Pignedoli, its solution cannot be obtained by superposition due to the nonlinearity inherent in the introduction of the relativistic mass. Recently, Mitchell and Pope[7] studied this problem using the PLK method and obtained the limiting form of the solution as the nonlinearity vanishes. However, in the damped case, their expansion contains secular terms and, hence, does not reveal the ultimate nature of the oscillation. The methods used here obviate this difficulty.

2. EQUATION OF MOTION

The equation of motion, in which the origin coincides with the equilibrium position of the particle, is

$$(d/dt)(m^*du/dt) = -[ku + r(du/dt)], \qquad (1)$$

where

$$m^* = m[1 - (du/dt)^2/c^2]^{-\frac{1}{2}}. \qquad (2)$$

* Work sponsored by the U. S. Army Research Office, Durham, North Carolina.

[1] A. Pignedoli, Atti. Mem. Accad. Sci. Mod. 9, 256 (1951).
[2] R. Penfield and H. Zatskis, J. Franklin Inst. 262, 121 (1956).
[3] L. A. MacColl, Am. J. Phys. 25, 535 (1957).
[4] St. Gheorghita, Rev. Univ. Bucuresti C. I. Parhon Politehn, Ser. Sti. Nat. 1, 19 (1953).
[5] C. Eftimiu and S. Klarsfeld, Ann. Univ. Bucuresti C. I. Parhon, Ser. Sti. Nat., 6, 53 (1957).
[6] L. Dragos, Ann. Univ. Bucuresti C. I. Parhon, Ser. Sti. Nat. 5, 41 (1956).
[7] T. P. Mitchell and D. L. Pope, J. Soc. Indt. Appl. Math. 10, 49 (1962).

uniform zoning in the R and Z directions, there would be no trouble in constructing a symmetric tensor, in fact the previously constructed tensor is symmetric under these circumstances.

Condition (d) refers to a velocity field of the type

$$\mathbf{u} = \bar{\mathbf{R}}.$$

$(\partial u^A/\partial k)$ and $(\partial u^B/\partial l)$ do not vanish in general under this type of field. This could probably be remedied, but since this problem seems related to the problem of angular momentum conservation, it did not seem worth while to fix this one, and not be able to do anything about the other.

One final comment: From both analysis and experience, the shock stability criterion of the difference equations resulting from using the described tensor viscosity is less severe than that deduced from a scalar viscosity. Thus, a tensor viscosity leads to no additional difficulties as far as stability of the difference equations is concerned.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 1 JANUARY 1964

Motion of a Relativistic Damped Oscillator*

R. A. STRUBLE AND T. C. HARRIS
North Carolina State University, Raleigh, North Carolina
(Received 28 June 1963; revised September 1963)

Quantitative effects of damping in a nonlinear system are illustrated for the case of a relativistic oscillator. The elementary techniques employed furnish details of the decay phenomena which are more extensive than those normally furnished by stability theorems.

1. INTRODUCTION

A SO-CALLED relativistic oscillator describes the oscillatory motion of a particle where the rest mass has been replaced by the relativistic mass. Though the resulting model may be somewhat artificial from a physical point of view, it has been the subject of considerable interest and from a mathematical viewpoint, it affords a good example in which one might compare the quantitative effects of damping in a linear and a nonlinear system. Pignedoli[1] has studied cases in which the force acting upon the particle is a function of position only and in which the force is a function of velocity only. The former is notable in that it includes the analog of the classical harmonic oscillator which has been treated by Penfield and Zatskis[2] and MacColl.[3] Other studies have been made by Gheorghita,[4] Eftimiu and Klarsfeld,[5] and Dragos.[6]

In this note we present a brief analysis of the relativistic damped harmonic oscillator where the force acting on the point mass is linearly dependent upon both the displacement and the velocity. Though this case is a combination of those treated by Pignedoli, its solution cannot be obtained by superposition due to the nonlinearity inherent in the introduction of the relativistic mass. Recently, Mitchell and Pope[7] studied this problem using the PLK method and obtained the limiting form of the solution as the nonlinearity vanishes. However, in the damped case, their expansion contains secular terms and, hence, does not reveal the ultimate nature of the oscillation. The methods used here obviate this difficulty.

2. EQUATION OF MOTION

The equation of motion, in which the origin coincides with the equilibrium position of the particle, is

$$(d/dt)(m^*du/dt) = -[ku + r(du/dt)], \qquad (1)$$

where

$$m^* = m[1 - (du/dt)^2/c^2]^{-\frac{1}{2}}. \qquad (2)$$

* Work sponsored by the U. S. Army Research Office, Durham, North Carolina.

[1] A. Pignedoli, Atti. Mem. Accad. Sci. Mod. 9, 256 (1951).
[2] R. Penfield and H. Zatskis, J. Franklin Inst. 262, 121 (1956).
[3] L. A. MacColl, Am. J. Phys. 25, 535 (1957).
[4] St. Gheorghita, Rev. Univ. Bucuresti C. I. Parhon Politehn, Ser. Sti. Nat. 1, 19 (1953).
[5] C. Eftimiu and S. Klarsfeld, Ann. Univ. Bucuresti C. I. Parhon, Ser. Sti. Nat., 6, 53 (1957).
[6] L. Dragos, Ann. Univ. Bucuresti C. I. Parhon, Ser. Sti. Nat. 5, 41 (1956).
[7] T. P. Mitchell and D. L. Pope, J. Soc. Indt. Appl. Math. 10, 49 (1962).

Here, m is the rest mass, m^* is the relativistic mass, c is the velocity of light in a vacuum, k the spring modulus, and r the damping constant. Introducing the nondimensional time variable $\tau = (k/m)^{\frac{1}{2}}t$, Eq. (1) becomes

$$\frac{d^2u}{d\tau^2} + \left(u + 2\alpha\frac{du}{d\tau}\right)\left[1 - \mu\left(\frac{du}{d\tau}\right)^2\right]^{\frac{1}{2}} = 0, \qquad (3)$$

where

$$\alpha = r(k/m)^{\frac{1}{2}}; \qquad \mu = k/mc^2.$$

General qualitative properties of the solutions of this equation are discussed in Sec. 3. Rigorous variational properties for μ small are obtained in Sec. 4 and a formal, asymptotic expansion in μ is considered in Sec. 5.

3. QUALITATIVE PROPERTIES

Multiplying Eq. (3) by $(du/d\tau)[1 - \mu(du/d\tau)^2]^{-\frac{1}{2}}$ and integrating, results in

$$(1/\mu)[1 - \mu(du/d\tau)^2]^{-\frac{1}{2}} + \tfrac{1}{2}u^2 = 1/\mu + W - 2\alpha f(\tau), \qquad (4)$$

where

$$f(\tau) \equiv \int_0^\tau (du/d\tau)^2\, d\tau.$$

The integration constant $1/\mu + W$ has been introduced so that W corresponds to the classical total energy of the motion.

Special solutions for the undamped relativistic oscillator ($\alpha = 0$) are illustrated in Fig. 1. These are the usual phase-plane trajectories, where $v = du/d\tau$. The trajectories are symmetric with respect to both the u axis and v axis and fill out the strip, $|v| \leq \mu^{-\frac{1}{2}}$. The origin corresponds to $W = 0$ and the trajectories close to the origin are nearly circular. Away from the origin, as W increases, the curves become elongated and approach the lines, $v = \pm\mu^{-\frac{1}{2}}$.

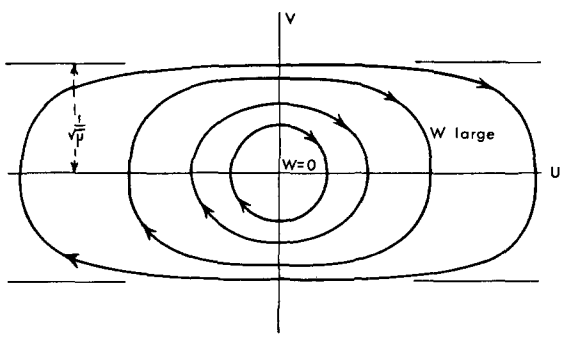


FIG. 1. Undamped oscillator.

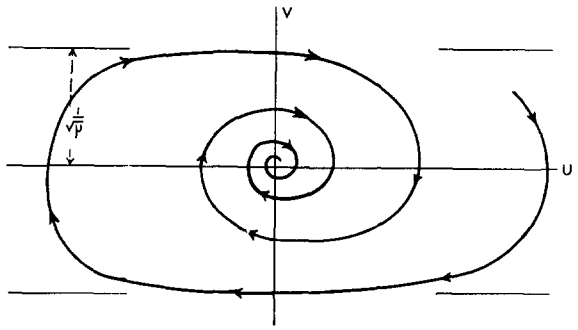


FIG. 2. Damped oscillator.

For this case, Eq. (4) may be expressed as

$$d\tau/\mu^{\frac{1}{2}} = u(1/\mu + W - u^2/2)\, du \times \{\pm[\mu^2(1/\mu + W - u^2/2)^2 - 1]^{\frac{1}{2}}\}^{-1},$$

and is readily reduced to standard elliptic form.[2,3] In particular, the least period T of the undamped relativistic oscillator may be written

$$T = 4\mu^{\frac{1}{2}}[w^{\frac{1}{2}}E - (2/\mu)w^{-\frac{1}{2}}K],$$

where

$$w = 2(2/\mu + W).$$

Here K and E are, respectively, complete elliptic integrals of the first and second kinds with modulus $(2W/w)^{\frac{1}{2}}$ and $T = 2\pi$ for $W = 0$, while $T \to \infty$ as $W \to \infty$.

For the general case with damping ($\alpha > 0$), the term $2\alpha f(\tau)$ is nondecreasing and, in effect, continuously reduces the energy along each trajectory. Thus, with damping, the trajectories cut across those in Fig. 1 and approach the origin. The resultant motion is either a damped oscillation, as in Fig. 2, or a nonoscillatory approach to equilibrium. Typical stability theorems[8] yield the same conclusion. However, it is our purpose here to compare the quantitative effects of damping in the relativistic oscillator with those in the linear system (3) when $\mu = 0$.

4. PROPERTIES OF THE EXACT VARIATIONAL SYSTEM

It is convenient to require that the damping coefficient satisfy the inequality $0 < \alpha < 1$. This restricts the analysis to oscillatory motion; however, we do not assume that the damping coefficient is necessarily small. For small μ, we introduce variational parameters A, θ by the equations

[8] S. Lefschetz, *Differential Equations: Geometric Theory* (Interscience Publishers, Inc., New York, 1957).

$$u = e^{-\alpha\tau}A\cos(\omega\tau - \theta), \tag{5}$$
$$du/d\tau = -\alpha e^{-\alpha\tau}A\cos(\omega\tau - \theta) - \omega e^{-\alpha\tau}A\sin(\omega\tau - \theta),$$

where $\omega = (1 - \alpha^2)^{\frac{1}{2}}$ is the corresponding linear frequency of Eq. (3) when $\mu = 0$. Substituting (5) in (3) yields the (exact) variational equations

$$dA/d\tau = -\mu e^{-2\alpha\tau}AF(A, \theta, \tau)\sin(\omega\tau - \theta), \tag{6}$$
$$d\theta/d\tau = \mu e^{-2\alpha\tau}F(A, \theta, \tau)\cos(\omega\tau - \theta),$$

where

$$F(A, \theta, \tau) \equiv \frac{A^2\Omega^2[\cos(\omega\tau - \theta) - 2\alpha\Omega][3 - 3\mu\Omega^2e^{-2\alpha\tau}A^2 + \mu^2\Omega^4e^{-4\alpha\tau}A^4]}{\omega[(1 - \mu\Omega^2e^{-2\alpha\tau}A^2)^{\frac{3}{2}} + 1]}, \tag{7}$$

and

$$\Omega = \Omega(\omega\tau - \theta) \equiv \alpha\cos(\omega\tau - \theta) + \omega\sin(\omega\tau - \theta).$$

We note that the function $F(A, \theta, \tau)$ assumes real values so long as the inequality

$$\mu\Omega^2e^{-2\alpha\tau}A^2 \leq 1$$

is maintained. This inequality, of course, corresponds to the requirement that real solutions lie in the strip $|v| = |du/d\tau| \leq \mu^{-\frac{1}{2}}$. Since $\Omega^2(\omega\tau - \theta) \leq 1$, $F(A, \theta, \tau)$ will be real for $\tau \geq 0$ and all θ, so long as $A^2 \leq \mu^{-1}$. Further, $F(A, \theta, \tau)$ is then majorized by a polynomial in A. Hence, if $\delta > 0$, there exists a $\Delta > 0$ such that

$$|F(A, \theta, \tau| \leq \Delta, \tag{8}$$

for $0 \leq A \leq \delta$, $\tau \geq 0$, and all θ, provided $\mu^{-\frac{1}{2}} \geq \delta$.

Now, if A_0 is an initial value for A satisfying $0 \leq A_0 < \delta$, the first equation in (6) implies that the inequality

$$|A(\tau) - A_0| \leq \mu\Delta\delta\int_0^\tau e^{-2\alpha\tau}\,d\tau < \mu\Delta\delta/2\alpha \tag{9}$$

holds on some interval $0 \leq \tau < b$. Consequently, if in addition, μ is chosen so small that $\mu < (2\alpha/\Delta\delta)(\delta - A_0)$, then both (8) and (9) remain valid for all $\tau \geq 0$. It also follows from the second equation in (6) that $|\theta(\tau) - \theta_0| < \mu\Delta/2\alpha$, where θ_0 is an arbitrary initial value for θ. Thus, if $\epsilon > 0$, there exists a $\mu_0 > 0$ with the property that whenever μ satisfies $0 \leq \mu \leq \mu_0$, the inequalities

$$|A(\tau) - A_0| < \epsilon, \qquad |\theta(\tau) - \theta_0| < \epsilon$$

hold for all $\tau \geq 0$. From this we see that for small μ, the effects of damping in the nonlinear oscillator are quantitatively very similar to those in the linear oscillator. In particular, the rates of decay and the frequencies are essentially the same. We will obtain somewhat more detailed information in the concluding section.

5. AN ASYMPTOTIC EXPANSION

Using a technique which has been introduced previously,[9,10] we assume an asymptotic series of the form

$$u = e^{-\alpha\tau}[A\cos(\omega\tau - \theta) + \mu u_1 + \mu^2u_2 + \cdots], \tag{10}$$

in the parameter μ. Here the amplitude and phase variables A and θ will be different from those introduced in Sec. 4 but will continue to reflect the basic characteristics of the motion. Except for transient effects, the additive corrections $u_1, u_2, \cdots$ depict the higher harmonics of the motion. We shall develop only the first-order solution (10).

Substituting the expansion (10) in (3) results in the equation

$$\left[\frac{d^2A}{d\tau^2} + 2\omega A\frac{d\theta}{d\tau} - A\left(\frac{d\theta}{d\tau}\right)^2\right]\cos(\omega\tau - \theta) + \left[A\frac{d^2\theta}{d\tau^2} + 2\frac{d\theta}{d\tau}\frac{dA}{d\tau} - 2\omega\frac{dA}{d\tau}\right]\sin(\omega\tau - \theta) + \mu\left(\frac{d^2u_1}{d\tau^2} + \omega^2u_1\right) = \tfrac{3}{8}\mu e^{-2\alpha\tau}A^3[-4\alpha\omega\sin(\omega\tau - \theta) + (1 - 4\alpha^2)\cos(\omega\tau - \theta) + 4\alpha\omega(1 - 2\alpha^2)\sin 3(\omega\tau - \theta) - (1 - 8\alpha^2 + 8\alpha^4)\cos 3(\omega\tau - \theta)] + O(\mu^2). \tag{11}$$

Upon balancing the fundamental harmonics in Eq. (11) and associating the remaining first-order terms (in μ) with u_1, we obtain the following:

$$\frac{d^2A}{d\tau^2} + 2\omega A\frac{d\theta}{d\tau} - A\left(\frac{d\theta}{d\tau}\right)^2 = \tfrac{3}{8}\mu(1 - 4\alpha^2)A^3e^{-2\alpha\tau},$$
$$A\frac{d^2\theta}{d\tau^2} - 2\omega\frac{dA}{d\tau} + 2\frac{d\theta}{d\tau}\frac{dA}{d\tau} = -\tfrac{3}{2}\mu\alpha\omega A^3e^{-2\alpha\tau}; \tag{12}$$

[9] R. A. Struble, *Nonlinear Differential Equations* (McGraw-Hill Book Company, Inc., New York, 1962), Chap. 8.

[10] R. A. Struble and J. E. Fletcher, J. Math. Phys. **2**, 880 (1961).

$$\frac{d^2u_1}{d\tau^2} + \omega^2 u_1 = \tfrac{3}{8}e^{-2\alpha\tau}A^3[4\alpha\omega(1 - 2\alpha^2)\sin 3(\omega\tau - \theta) - (1 - 8\alpha^2 + 8\alpha^4)\cos 3(\omega\tau - \theta)]. \quad (13)$$

It is readily verified that each solution of the reduced system

$$dA/d\tau = \tfrac{9}{16}\mu\alpha A^3 e^{-2\alpha\tau}, \qquad d\theta/d\tau = \tfrac{3}{16}\mu\omega A^2 e^{-2\alpha\tau} \quad (14)$$

satisfies (12) to first order in μ, and so it suffices to treat (14).[11] The first of these may be integrated to yield

$$A^2 = [A_0^{-2} + \tfrac{9}{16}\mu(e^{-2\alpha\tau} - 1)]^{-1} = A_0^2[1 - \tfrac{9}{16}\mu A_0^2(e^{-2\alpha\tau} - 1)] + O(\mu^2), \quad (15)$$

where A_0 is the initial amplitude for $\tau = 0$. If $A_0^2 < \mu^{-1}$, the right-hand member of Eq. (15) is bounded and we have

$$\lim_{\tau\to\infty}(A/A_0)^2 = 1 + \tfrac{9}{16}\mu A_0^2 + O(\mu^2).$$

Evidently, the effect of the nonlinear perturbation is to steadily increase the relative amplitude A/A_0 (See Fig. 3). The qualitative explanation for this is that the role of damping here is not only to remove energy but, because of the concomitant reduction of the relativistic mass, to produce a slightly larger amplitude of oscillation than that normally anticipated in a linear system.

From (14), we may obtain the relationship

$$A = K_0^{-1}e^{3\alpha\theta/\omega},$$

where K_0 is an integration constant. Then, the phase angle θ may be expressed in the form

$$\theta = \frac{\omega}{6\alpha}\ln\left[\frac{K_0^2}{A_0^{-2} + \tfrac{9}{16}\mu(e^{-2\alpha\tau} - 1)}\right] = \theta_0 - \tfrac{3}{32}(\omega/\alpha)\mu A_0^2(e^{-2\alpha\tau} - 1) + O(\mu^2), \quad (16)$$

where $\theta_0 = (\omega/\alpha)\ln(K_0A_0)$. Hence,

$$\lim_{\tau\to\infty}\theta = \theta_0 + \tfrac{3}{32}(\omega/\alpha)\mu A_0^2 + O(\mu^2).$$

According to this, the total change of phase angle is always finite but appears to become large as the damping coefficient α becomes small. In fact, Eqs. (14) imply that, without damping ($\alpha = 0$), A is constant and

$$\theta = \theta_0 + \tfrac{3}{16}\mu\omega A_0^2\tau.$$

This secular (timelike) variation in θ corresponds to a perturbation of the fundamental frequency which is to be expected in the nonlinear undamped system.

[11] The general form for these derivatives is suggested by the right-hand members of (12). It is only necessary to find appropriate constant factors.

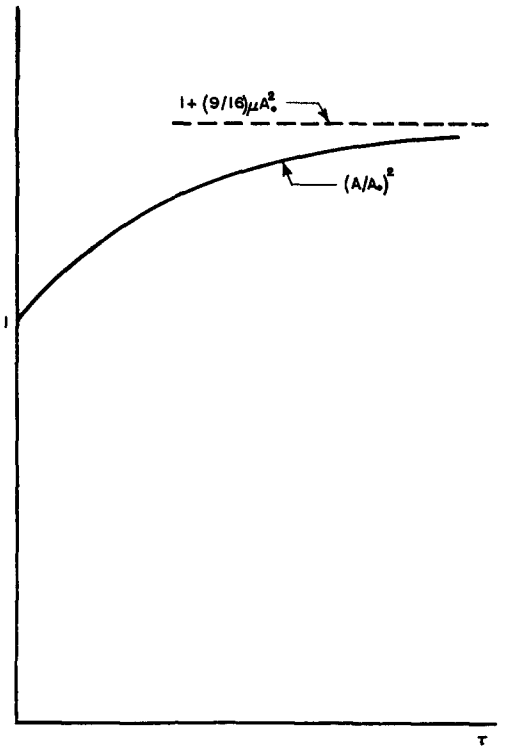


FIG. 3. Growth of relative amplitude.

The first-order solution is completed by the additive correction

$$u_1 = \tfrac{3}{32}A^3[e^{-2\alpha\tau}/(4 - 3\alpha^2)][(4\alpha^4 - 7\alpha^2 + 2)\times\cos 3(\omega\tau - \theta) + \alpha\omega(4\alpha^2 - 5)\sin 3(\omega\tau - \theta)],$$

which satisfies (13) to zero order in μ. Then u becomes

$$u = Ae^{-\alpha\tau}\cos(\omega\tau - \theta) + \tfrac{3}{32}[\mu/(4 - 3\alpha^2)]e^{-3\alpha\tau}A^3 \times [(4\alpha^4 - 7\alpha^2 + 2)\cos 3(\omega\tau - \theta) + \alpha\omega(4\alpha^2 - 5)\sin 3(\omega\tau - \theta)], \quad (17)$$

where A and θ are given by Eqs. (15) and (16).

The motion as described by Eq. (17) is a damped oscillation where the essential characteristics are reflected in the fundamental term. Due to the nonlinearity, higher harmonic terms are present in the expansion, but evidently these decay more rapidly then does the fundamental. Since the relative amplitude A/A_0 increases, the damping here is somewhat less effective than in the corresponding linear harmonic oscillator. On the other hand, the frequency is essentially that of the linear oscillator since the phase θ undergoes only a finite change. The variational properties obtained in Sec. 4 are borne out in the asymptotic expansion.

The writers wish to thank the Referees for some helpful comments.

Excitation Spectrum of an Impurity in the Hard-Sphere Bose Gas*

TOSHIO SODA†
The Enrico Fermi Institute for Nuclear Studies, The University of Chicago, Chicago, Illinois
(Received 27 August 1963)

The impurity excitation spectrum is calculated for the hard-sphere Bose gas with dilute impurity concentration, using the pseudopotential method of Lee, Huang, and Yang. The spectrum for the ideal case, in which the mass and interaction of impurity are the same with those of the bosons in the gas, is k^2 in the unit of $2m_a = \hbar = 1$ (m_a being the mass of boson, and k the momentum of impurity), with an effective mass correction of $-\frac{8}{45}[(\rho a^3)^{\frac{1}{2}}/\pi^{\frac{1}{2}}]k^2 + O[(\rho a^3)k^2, k^4]$, where ρ is the density of the boson gas, and a is the hard-sphere diameter of the boson. The general impurity case is also calculated and found to have an impurity spectrum rk^2, with a similar effective mass correction, where r is the mass ratio of the boson to the impurity.

I. INTRODUCTION

THERE has been some interest in experiments with impurity particles in the superfluid liquid helium. Theoretical investigations have been done by Landau and Pomeranchuk,[1] and by Girardeau.[2] The former authors assume the excitation spectrum and discuss the consequences. The latter author adapts the Lee–Low–Pines technique of polaron to the impurity problem, and examines the low-lying excitation spectrum for the case of a much heavier impurity particle for the hard-sphere boson gas. Recently Miller, Pines, and Nozieres[3] have also presented a formalism of treating the impurity particle in their work of elementary excitations in liquid helium, but they have not explicitly evaluated the actual spectrum with a specific potential.

There would be a great interest and illumination if the low lying excitation spectrum of an impurity is investigated in a mathematically rigorous way within the framework of hard-sphere Bose gas at low density, because the low-density hard-sphere Bose gas seems to have been investigated on a firmer foundation by Lee, Huang, and Yang[4] in treatments of the many-body boson problem. To understand the excitation spectrum in the above spirit is the aim of this paper. We consider the problem of impurity particles of dilute density in the hard-core boson gas. This problem also has implications for understanding what is the effect of statistics to the single-particle excitation.

* This work was supported in part by the United States Atomic Energy Commission. Part of this work was done by the author while at the Institute for Advanced Study, Princeton, New Jersey.

[1] L. D. Landau and I. Pomeranchuk, Dokl. Akad. Nauk SSSR **55**, 669 (1948).

[2] M. Girardeau, Phys. Fluids **4**, 249 (1961).

[3] A. Miller, D. Pines, and P. Nozieres, Phys. Rev. **127**, 1452 (1962).

[4] T. D. Lee, K. Huang, and C. N. Yang, Phys. Rev. **106**, 1135 (1957).

In the second section we treat the mathematically ideal case in which an impurity has the same mass as a boson and has the same form of interaction between boson and impurity as the boson–boson interaction. We also discuss the equivalence between a single-particle removal energy and self-energy of the ground state due to the presence of an impurity, which gives a check of our calculation to be correct.

In the third section we discuss the general impurity particle which has a different mass and interaction.

II. IMPURITY PARTICLE OF EQUAL MASS AND INTERACTION WITH BOSON

In this section we treat an ideal case, in which an impurity particle has the same mass and the interaction between the impurity and boson is the same as the boson–boson interaction. We assume that the density of impurity particles is so low that they interact only with bosons. Actually we consider only a single impurity particle in the boson gas in our calculation.

Let us consider the following two systems: system A and system B. System A contains $(N + 1)$ identical bosons in the volume Ω. We call these bosons "a" particles. System B contains N identical bosons plus one single impurity particle in the volume Ω. We call this impurity particle "b" particle.

Let us investigate the ground-state energy and low-lying excitation energy of system A. The Hamiltonian of system A is given as follows:

$$H_A = \sum_p p^2 + \tfrac{1}{2} \sum_{p,p',q} V_q a_{p+q}^+ a_{p'}^+ a_{p'+q} a_p, \qquad (1)$$

where a_p^+ and a_p are the creation and annihilation operators of a boson with momentum p. We choose the unit $2m_a = \hbar = 1$ in this paper. We use a pseudopotential to describe hard-sphere interaction

as given in the paper of Huang, Lee, and Yang,[4]

$$V_q = 8\pi a/\Omega, \tag{2}$$

where a is the hard-sphere diameter of the boson. If we assume the condensation to occur at the zero momentum state as Bogoliubov did, and keep the terms proportional to the total number of particles, we obtain the following Hamiltonian:

$$H_A = \frac{N(N+1)}{2\Omega} 8\pi a + \sum_p \left(p^2 + 8\pi a \frac{N+1}{\Omega}\right) a_p^+ a_p + \frac{(N+1)8\pi a}{\Omega} \sum_p (a_p^+ a_{-p}^+ + a_{-p} a_p). \tag{3}$$

In order to diagonalize H_A, we introduce the Bogoliubov transformation

$$a_k = u_k \alpha_k + v_k \alpha_{-k}^+, \tag{4}$$

where u_k and v_k are real coefficients and α_k is the annihilation operator of a so-called quasiparticle. The ground-state energy is given by[4]

$$E_g = \frac{N+1}{\Omega} 4\pi a \left[1 + \frac{128}{15\pi^{\frac{1}{2}}} \left(\frac{N+1}{\Omega} a^3\right)^{\frac{1}{2}}\right], \tag{5}$$

and the excitation energy is given by[4]

$$E = \sum_p p\left(p^2 + 16\pi a \frac{N+1}{\Omega}\right)^{\frac{1}{2}} \alpha_p^+ \alpha_p. \tag{6}$$

The above choice of the terms in the Hamiltonian is equivalent to picking up the terms which yield the correct result up to the order of $\rho a(\rho a^3)^{\frac{1}{2}}$, in the low-density perturbation expansion of (ρa^3). Here ρ is the density of the system.

Now let us turn to our problem of calculating the energy of the system B. The Hamiltonian is given as follows:

$$H_B = \sum_p p^2 a_p^+ a_p + \tfrac{1}{2} \sum_{p,p',q} V_q a_{p+q}^+ a_{p'}^+ a_{p'+q} a_p + \sum_p p^2 b_p^+ b_p + \sum_{p,p',q} V_q a_{p+q}^+ b_{p'}^+ b_{p'+q} b_p, \tag{7}$$

with the subsidiary condition

$$\sum_p b_p^+ b_p = 1, \tag{8}$$

which means that we have a single impurity particle. Here b_p is the annihilation operator of an impurity particle with momentum p.

If we use the same pseudopotential here and keep the corresponding terms to H_A in Eq. (3), which are proportional to the total number of particles when "b" is changed to "a", and yield the energy up to the same order of magnitude as the calculation for the system A, we obtain the following Hamiltonian:

$$H_B = \frac{N(N-1)}{2\Omega} 8\pi a + \sum_p \left(p^2 + 8\pi a \frac{N}{\Omega}\right) a_p^+ a_p + \frac{N}{\Omega} 4\pi a \sum_p (a_p^+ a_{-p}^+ + a_{-p} a_p) + \frac{N}{\Omega} 8\pi a + \sum_p p^2 b_p^+ b_p + \frac{N^{\frac{1}{2}}}{\Omega} 8\pi a [\sum_p (a_{-p}^+ b_p^+ b_0 + b_0^+ b_p a_{-p} + a_p^+ b_0^+ b_p + b_p^+ b_0 a_p) + \sum_{\substack{p,q \\ p\neq 0 \\ q\neq 0}} (a_{p+q}^+ b_{-p}^+ b_q + b_q^+ b_{-p} a_{p+q})]. \tag{9}$$

[*Note added in proof.* Exactly speaking, we retain all terms of the boson–impurity interaction, which contain at least one zero momentum state boson operator a_0^+ or a_0.]

Let us investigate the diagonalization of H_B in Eq. (9). We first diagonalize the "a" particle part using the Bogoliubov transformation,

$$H_B = \tfrac{1}{2} \frac{N(N-1)}{\Omega} 8\pi a + \sum_p p(p^2 + 16\pi a\rho)^{\frac{1}{2}} \alpha_p^+ \alpha_p + \tfrac{1}{2} \sum_p [p(p^2 + 16\pi a\rho)^{\frac{1}{2}} - (p^2 + 8\pi a\rho)] + \frac{N}{\Omega} 8\pi a + \sum_p p^2 b_p^+ b_p + \frac{N^{\frac{1}{2}}}{\Omega} 8\pi a \sum_p [(u_p \alpha_{-p}^+ - v_p \alpha_p) b_p^+ b_0 + b_0^+ b_p (u_p \alpha_{-p} - v_p \alpha_p^+) + (u_p \alpha_p^+ - v_p \alpha_{-p}) b_0^+ b_p + b_p^+ b_0 (u_p \alpha_p - v_p \alpha_{-p}^+)] + \frac{N^{\frac{1}{2}}}{\Omega} 8\pi a \sum_{\substack{p,q \\ p\neq 0\neq q}} [(u_{p+q} \alpha_{p+q}^+ - v_{p+q} \alpha_{-p-q}) b_{-p}^+ b_q + b_q^+ b_{-p} (u_{p+q} \alpha_{p+q} - v_{p+q} \alpha_{-p-q}^+)], \tag{10}$$

where ρ is given by

$$\rho = N/\Omega, \tag{11}$$

and u_p and v_p are given by

$$u_p^2 = \tfrac{1}{2}[1 + (p^2 + 8\pi a\rho)/p(p^2 + 16\pi a\rho)^{\frac{1}{2}}] \tag{12}$$

and

$$v_p^2 = \tfrac{1}{2}[-1 + (p^2 + 8\pi a\rho)/p(p^2 + 16\pi a\rho)^{\frac{1}{2}}]. \tag{13}$$

Let us diagonalize the "b" particle part of the Hamiltonian which contains the interaction of the quasiparticle with bosons. The diagonalization procedure allows us to use the perturbation calculation. Looking at Eq. (10) carefully, we find that the ground state, which is composed of N "a" particles and one "b" particle at rest ($p = 0$), has the self-energy due to the presence of a b_0 particle interacting with a quasiparticle α_p in the Hamiltonian. The diagram of this self-energy-type interaction is

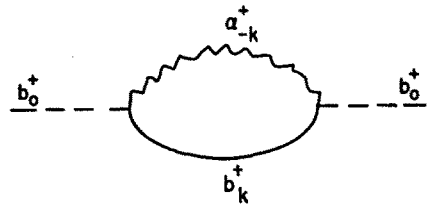


FIG. 1. Self-energy of the ground state in the second-order perturbation caused by the terms like $b_p{}^+ \alpha_{-p}{}^+ b_0$ and $b_0{}^+ \alpha_{-p}\, b_p$ in the Hamiltonian.

shown in Fig. 1. The actual calculation leads to

$$E_g^{(2)} = \frac{N(8\pi a)^2}{\Omega^2} \times \sum_p (u_p - v_p)^2 \frac{1}{-p(p^2 + 16\pi a\rho)^{\frac{1}{2}} - p^2} = -\frac{(8\pi a)^2}{8\pi^3} 4\pi\rho \int_0^\infty dp\, p^2 \times \frac{p^2}{p(p^2 + 16\pi a\rho)^{\frac{1}{2}}} \frac{1}{p(p^2 + 16\pi a\rho)^{\frac{1}{2}} + p^2}. \quad (14)$$

There is a linear divergence, but we can subtract it by using the correct pseudopotential as in reference 4, or by subtracting the free-scattering correction from K matrix due to Sawada.[5] We obtain the following result.

$$E_g^{(2)} = \frac{64}{3\pi^{\frac{1}{2}}} 4\pi\rho a(\rho a^3)^{\frac{1}{2}}. \quad (15)$$

We notice the following interesting relationship. This positive energy shift Eq. (15) of the ground-state energy together with the $8\pi a\rho$ term of the Hamiltonian in Eq. (10) is equivalent to a removal energy of a single "a" particle from $(N + 1)$ "a"-particle ground state, as it should be. Because the "b" particle is nothing different but its tag and the absolute (no symmetry requirement) ground state of the Boltzmann statistic is equivalent to the one of the Bose statistics, if there is no degeneracy in the ground state and the wavefunction is nodeless. The actual calculation from Eq. (5) shows

$$E_g((N+1)\text{"}a\text{"}) - E_g(N\text{"}a\text{"}) = \left[\frac{N(N+1)}{\Omega} - \frac{N(N-1)}{\Omega}\right] 4\pi a + \frac{128}{15\pi^{\frac{1}{2}}} \frac{4\pi a^{5/2}}{\Omega^2} [(N+1)^{\frac{1}{2}} N - N^{\frac{1}{2}}(N-1)] = 8\pi\rho a + \frac{64}{3\pi^{\frac{1}{2}}} 4\pi\rho a(\rho a^3)^{\frac{1}{2}}. \quad (16)$$

This result of Eq. (16) checks the correctness of our self-energy calculation of Eq. (15).

[*Note added in proof.* We obtain the result of Eq. (16) correctly by differentiation, $\partial E_g(N\text{"}a\text{"})/\partial N$, because the terms of $O(1)$ in Eq. (16) do not have much meaning. The author owes this remark to Professor G. Wentzel. Professor K. Sawada has pointed out to the author that the reverse check of the relationship between Eqs. (15) and (16) was done in Eqs. (15), (15′), (16), and the appendix of his and L. Bruch's paper, Phys. Rev. **131**, 1379 (1963).]

Let us now calculate the self-energy of "b" particle with momentum k by the second-order perturbation. The interaction term containing b_0 or b_0^+ in the Hamiltonian, has a contribution proportional to the inverse of the volume Ω and is vanishingly small in this case as $\Omega \to \infty$. However, the last term in the Hamiltonian of (10) contributes to the self-energy which is shown in Fig. 2. The energy of this diagram is given by

$$E_k^{(2)} = -\frac{N}{\Omega^2} (8\pi a)^2 \times \sum_p \left[(u_p - v_p)^2 \frac{1}{\omega_p + (k-p)^2 - k^2} - \frac{1}{p^2 + (k-p)^2 - k^2} \right], \quad (17)$$

where ω_p is the quasiparticle energy given by

$$\omega_p = p(p^2 + 16\pi\rho a)^{\frac{1}{2}}. \quad (18)$$

The second term of Eq. (17) is added to subtract the linear divergence at higher momentum in p. This is justified, as in the ground-state energy calculation, by using the correct pseudopotential or by subtracting free-scattering terms from the K matrix as in the Sawada method.[5] The actual calculation goes as follows:

$$E_k^{(2)} = -16\rho a^2 \int_0^\infty dp \int_{-1}^1 d\mu \left[\frac{p^2}{(p^2 + 16\pi\rho a)^{\frac{1}{2}}} \times \frac{1}{(p^2 + 16\pi\rho a)^{\frac{1}{2}} - p - 2k\mu} - \frac{p}{2p - 2k\mu} \right]. \quad (19)$$

After integrating over μ, cosine of the angle between $\mathbf{k}$ and $\mathbf{p}$, we obtain

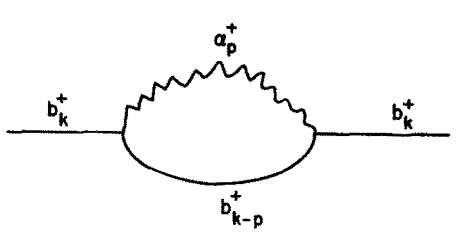


FIG. 2. Self-energy of "b" particle with finite momentum $\mathbf{k}$ in the second-order perturbation caused by the quasiparticle of the boson.

[5] K. Sawada, Phys. Rev. **116**, 1346 (1959).

$$E_k^{(2)} = \frac{8\rho a^2}{k}\int_0^\infty dp\Big[\frac{p^2}{(p^2+16\pi\rho a)^{\frac{1}{2}}} \times \ln\frac{(p^2+16\pi\rho a)^{\frac{1}{2}}+p-2k}{(p^2+16\pi\rho a)^{\frac{1}{2}}+p+2k} - p\ln\frac{p-k}{p+k}\Big]. \quad (20)$$

By changing the integration variable from p to t,

$$t = \tfrac{1}{2}[(p^2+16\pi\rho a)^{\frac{1}{2}}+p], \quad (21)$$

we have the following expression:

$$E_k^{(2)} = \frac{8\rho a^2}{k}\Big\{\int_e^\infty dt\Big[\frac{(t^2-4\pi\rho a)^2}{t^3}-t\Big] \times \ln\frac{t-k}{t+k} - \int_0^e dt\, t\ln\frac{t-k}{t+k}\Big\}, \quad (22)$$

where $e = (4\pi\rho a)^{\frac{1}{2}}$.

The integration of Eq. (22) and the expansion in small quantity k lead to the following result:

$$E_k^{(2)} = \frac{64}{3\pi^{\frac{1}{2}}}4\pi\rho a(\rho a^3)^{\frac{1}{2}} - \frac{64(\rho a^3)^{\frac{1}{2}}}{45\pi^{\frac{1}{2}}}k^2 + O(k^4). \quad (23)$$

In this calculation the k^2 term in Eq. (23) is not exact, because we have not taken the full Hamiltonian and it only gives the coefficient of the lowest order correctly in (ρa^3) expansion. The contribution of the neglected terms of the full Hamiltonian is of the order $(\rho a^3)k^2$ or higher in (ρa^3) expansion, containing terms of $O(k^4)$ as well. It is very interesting to notice that the constant term of the second-order self-energy of "b" particle with momentum k is equal to the positive energy shift of the ground state, due to the presence of a "b_0" particle in the interaction Hamiltonian, given in Eq. (15).

Now the excitation energy E_k of impurity "b" particle takes the following form within the second-order perturbation theory and up to the order of $(\rho a^3)^{\frac{1}{2}}k^2$:

$$\mathrm{E}_k = k^2 + E_k^{(2)} - E_g^{(2)} = k^2 - \frac{64}{45}\frac{(\rho a^3)^{\frac{1}{2}}}{\pi^{\frac{1}{2}}} + O[(\rho a^3)k^2, k^4]. \quad (24)$$

This is our final result. This impurity excitation is a single-particle excitation without statistics, because this impurity has a special tag to distinguish it from the surrounding bosons. If there is no tag on the "b" particle, then the effect of Bose statistics forces the excitation spectrum to take the following collective excitation spectrum:

$$E_k = k\{k^2+16\pi[(N+1)/\Omega]a\}^{\frac{1}{2}} \simeq (16\pi a\rho)^{\frac{1}{2}}k. \quad (25)$$

We can see the effect of statistics by comparing Eqs. (24) and (25); namely, the collective excitation has a linear spectrum for small k, whereas the impurity excitation has a quadratic spectrum in k.

III. GENERAL IMPURITY PARTICLE

In this section we treat the general case, in which the impurity particle has a different mass m_b to the boson mass m_a, and a different form of interaction than the boson–boson interaction. We introduce two parameters r and c. r is the mass ratio of boson to impurity.

$$r = m_a/m_b. \quad (26)$$

c is the hard-core diameter of interaction between boson and impurity, which is given by

$$c = \tfrac{1}{2}(a+a_i), \quad (27)$$

where a and a_i are the hard-core diameters of boson and impurity, respectively. Now the impurity and boson interaction is given in the following new form:

$$V_{ab} = \frac{4\pi c}{2[m_a m_b/(m_a+m_b)]} = 8\pi c\Big(\frac{1+r}{2}\Big). \quad (28)$$

The Hamiltonian is given by

$$H = \sum_p\Big(p^2+8\pi\frac{N}{\Omega}a\Big)a_p^+a_p + \frac{N}{\Omega}4\pi a\sum_p(a_p^+a_{-p}^+ + a_{-p}a_p) + \frac{N(N-1)}{2\Omega}8\pi a + \sum_p rp^2b_p^+b_p + \frac{N}{\Omega}8\pi c\Big(\frac{1+r}{2}\Big) + \frac{N^{\frac{1}{2}}}{\Omega}8\pi c\Big(\frac{1+r}{2}\Big) \times\{\sum_p(a_{-p}^+b_pb_0 + b_0b_pa_{-p} + a_p^+b_0b_p + b_p^+b_0a_p) + \sum_{\substack{p,q\\ p\neq0\\ q\neq0}}(a_{p+q}^+b_{-p}^+b_q + b_q^+b_{-p}a_{p+q})\}. \quad (29)$$

The procedures of calculation are the same as in the previous section. The second-order perturbation energy to the ground-state energy is

$$E_g^{(2)} = \frac{N}{\Omega^2}\Big(8\pi c\frac{1+r}{2}\Big)^2 \times\sum_p(u_p-v_p)^2\frac{1}{-p(p^2+16\pi\rho a)^{\frac{1}{2}}-rp^2} = -32\rho c^2\Big(\frac{1+r}{2}\Big)^2\int_0^\infty dp \times\frac{p^2}{(p^2+16\pi\rho a)^{\frac{1}{2}}}\frac{1}{(p^2+16\pi\rho a)^{\frac{1}{2}}+rp}. \quad (30)$$

After subtracting linear divergence, which is given by

$$\frac{N\{8\pi c[\frac{1}{2}(1+r)]\}^2}{\Omega^2}\sum_p\frac{1}{-(1+r)p},$$

we obtain

$$E_g^{(2)} = \frac{64\rho c^2}{2\pi^{\frac{1}{2}}r}\Big(\frac{1+r}{2}\Big)^2 4\pi(\rho a)^{\frac{1}{2}}$$

$$\times\left\{1-\frac{1}{1-r^2}-\frac{r}{(r^2-1)^{\frac{1}{2}}}\right.$$
$$\left.\times\ln\frac{1+[(r-1)/(r+1)]^{\frac{1}{2}}}{1-[(r-1)/(r+1)]^{\frac{1}{2}}}\right\}\quad\text{for } r>1, \tag{31}$$
$$=\frac{64\rho c^2}{3\pi^{\frac{1}{2}}}4\pi(\rho a)^{\frac{1}{2}}\quad\text{for } r=1, \tag{32}$$
$$=\frac{64\rho c^2}{2\pi^{\frac{1}{2}}r}\left(\frac{1+r}{2}\right)^2 4\pi(\rho a)^{\frac{1}{2}}\left\{1-\frac{1}{1-r^2}\right.$$
$$\left.-\frac{2r}{(r^2-1)^{\frac{1}{2}}}\left[\frac{\pi}{2}-\tan^{-1}\left(\frac{1+r}{1-r}\right)^{\frac{1}{2}}\right]\right\}$$
$$\text{for } r<1. \tag{33}$$

The second-order perturbation energy of the "b" particle with a finite momentum k is given by the following expression:

$$E_k^{(2)}=-\frac{N\{8\pi c[\frac{1}{2}(1+r)]\}^2}{\Omega^2}$$
$$\times\sum_p(u_p-v_p)^2\frac{1}{\omega_p+r[(k-p)^2-k^2]}, \tag{34}$$

where ω_p is given by Eq. (18).
The actual calculation goes

$$E_k^{(2)}=-4\pi\rho\frac{\{8\pi c[\frac{1}{2}(1+r)]\}^2}{8\pi^3}\int_0^\infty dp\frac{p^2}{(p^2+16\pi\rho a)^{\frac{1}{2}}}$$
$$\times\tfrac{1}{2}\int_{-1}^{1}d\mu\frac{1}{(p^2+16\pi\rho a)^{\frac{1}{2}}+r(p-2k\mu)}. \tag{35}$$

After doing the angle integration, we obtain

$$E_k^{(2)}=-32\rho c^2[\tfrac{1}{2}(1+r)]^2\int_0^\infty dp\frac{p^2}{(p^2+16\pi\rho a)^{\frac{1}{2}}}$$
$$\times\frac{1}{4kr}\left\{\ln\frac{[(p^2+16\pi\rho a)^{\frac{1}{2}}+rp]+2kr}{[(p^2+16\pi\rho a)^{\frac{1}{2}}+rp]-2kr}\right\}. \tag{36}$$

Here in (36) we expand the quantity in the curly bracket in small quantity k. We obtain[6]

$$E_k^{(2)}=-32\rho c^2[\tfrac{1}{2}(1+r)]^2$$
$$\times\left\{\int_0^\infty dp\frac{p^2}{(p^2+16\pi\rho a)^{\frac{1}{2}}}\frac{1}{(p^2+16\pi\rho a)^{\frac{1}{2}}+rp}\right.$$
$$+\frac{4r^2}{3}k^2\int_0^\infty dp\frac{p^2}{(p^2+16\pi\rho a)^{\frac{1}{2}}}$$
$$\left.\times\frac{1}{[(p^2+16\pi\rho a)^{\frac{1}{2}}+rp]^3}+O(k^4)\right\}. \tag{37}$$

The first term in (37) is equivalent to $E_g^{(2)}$ in (30) and, after subtracting the linear divergence, we obtain the same result as in (31), (32), and (33). Therefore the difference $E_k^{(2)}-E_g^{(2)}$ is given by

$$E_k^{(2)}-E_g^{(2)}=-Ak^2+O(k^4), \tag{38}$$

where A is calculated from the second term in (37) and is given by

$$A=\frac{4}{3}\frac{(\rho c^4a^{-1})^{\frac{1}{2}}}{\pi^{\frac{1}{2}}}\frac{r^2}{1+r}\left\{\left(\frac{r+1}{r-1}\right)\right.$$
$$+\left[5+2\frac{1+r}{1-r}-3\left(\frac{1+r}{1-r}\right)^2\right]\frac{1+r}{4}$$
$$+\left[3-2\frac{r+1}{1-r}+3\left(\frac{1+r}{1-r}\right)^2\right]\frac{1}{4}\left(\frac{r+1}{r-1}\right)^{\frac{1}{2}}$$
$$\left.\times\ln\left|\frac{(r-1)^{\frac{1}{2}}+(r+1)^{\frac{1}{2}}}{(r-1)^{\frac{1}{2}}-(r+1)^{\frac{1}{2}}}\right|\right\}\quad\text{for } r>1, \tag{39}$$
$$=(64/45\pi^{\frac{1}{2}})(\rho c^4a^{-1})^{\frac{1}{2}}\quad\text{for } r=1, \tag{40}$$
$$=\frac{4}{3}\frac{(\rho c^4a^{-1})^{\frac{1}{2}}}{\pi^{\frac{1}{2}}}\frac{r^2}{1+r}\left\{\left(\frac{r+1}{r-1}\right)\right.$$
$$+\left[5+2\frac{1+r}{1-r}-3\left(\frac{1+r}{1-r}\right)^2\right]\frac{1+r}{4}$$
$$+\left[3-2\frac{1+r}{1-r}+3\left(\frac{1+r}{1-r}\right)^2\right]\frac{1}{2}\left(\frac{r+1}{1-r}\right)^{\frac{1}{2}}$$
$$\left.\times\left[\frac{\pi}{2}-\tan^{-1}\left(\frac{1+r}{1-r}\right)^{\frac{1}{2}}\right]\right\}\quad\text{for } r<1. \tag{41}$$

The coefficient of k^2 term in (38) is not exact, and A gives the lowest-order term in (ρa^3)-type expansion as we discussed in Sec. 2. The neglected terms of the full Hamiltonian contribute $(\rho c^2a)k^2$ and $(\rho c^3)k^2$ or higher-order, similarly, as in the previous section.

We obtain the following expression for the excitation spectrum:

$$\mathrm{E}_k=rk^2-Ak^2+O(\rho c^2ak^2\ \text{ or }\ \rho c^3k^2,k^4), \tag{42}$$

where r appears in the unperturbed energy because of our unit system. This result is in agreement with the calculation of Girardeau[2] for the case of heavy mass of impurity.

One final comment is that the result of (42) is quite a natural one since we calculated the impurity-excited-state energy $E(k)$, and we subtract the ground-state energy which is equal to the zero-momentum excited state $E(0)$. Therefore, the difference $E(k)-E(0)$ does not naturally contain any constant term and starts eventually with the k^2 term, lacking a linear term in k.

ACKNOWLEDGMENTS

The author is much indebted to Professor C. N. Yang for guidance and discussion throughout this work. He is also thankful to Dr. M. D. Girardeau, Jr., for many stimulating discussions. The author acknowledges thanks for the hospitality of the Institute for Advanced Study, Princeton, New Jersey.

[6] The authors of reference 3 derived a similar expression for impurity effective mass as the second term of $E_k^{(2)}$ in Eq. (37), in their Eq. (3.17).

Approximate Calculations of Scattering Phase Shifts*†

WALTER R. SOOY‡
Department of Physics, University of California, Los Angeles, California
(Received 14 August 1963)

The nonrelativistic scattering problem is reformulated by transforming the partial-wave equation into similarity with the corresponding equation for a known or soluble problem. It is shown that if the potentials for the two cases are sufficiently close (within about 10% typically), a particularly simple and accurate expression for the desired phase shifts can be written. Numerical examples for a Gaussian potential are given. The method is also extended to the calculation of bound-state energy levels, but no numerical examples are given.

I. INTRODUCTION

IN THE following, we present a method for the evaluation of scattering phase shifts which retains the essential calculational simplicity of the Born and WKB approximations while improving substantially upon their accuracy. The basic idea is to transform a given radial wave equation into close similarity with another such equation whose phase shift is known. This has the effect of casting part of the analytical burden on the determination of the asymptotic properties of the transformation. In the resulting expression for the exact phase shift, a group of known or readily calculable terms is segregated, which constitutes a good approximation under certain conditions. In addition to the known phase shift, this group contains the difference between the Jeffreys (or Langer, or Born, etc., depending on the transformation) approximate phase shifts. The approach lends itself to iteration, so in principle an arbitrary level of accuracy can be realized without specific reference to wavefunctions.

This formulation represents a generalization of the work done by Saxon[1] and Nodvik,[2] who employed a WKB-based transformation to the free-particle equation, and Hougardy[3] who studied a WKB-based transformation to an arbitrary comparison potential for the one-dimensional case. The potential utility of this approach is greatest where accurate phase-shift calculations must be made on a large number of similar potentials, as in a search routine.

* Based on portions of a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of California, Los Angeles, California (1962).

† This work was supported in part by the National Science Foundation and the Air Force Office of Scientific Research.

‡ Now at Hughes Aircraft Company, Culver City, California.

[1] D. S. Saxon, Phys. Rev. 107, 871 (1957).

[2] J. S. Nodvik, Ph.D. Thesis, University of California, Los Angeles (1957).

[3] R. W. Hougardy, Ph.D. Thesis, University of California, Los Angeles (1961).

II. TRANSFORMATION OF THE RADIAL WAVE EQUATION

We seek the asymptotic phase shift of the solution $\psi_l(r)$, of the partial wave equation

$$d^2\psi_l(r)/dr^2 + \kappa^2(r)\psi_l(r) = 0, \tag{1}$$

where

$$\kappa^2(r) = 1 - U(r) - l(l+1)/r^2, \tag{2}$$

and

$U(r)$ = scattering potential energy/kinetic energy of incident particle,
$r = k\rho$ = wavenumber of incident particle × radius

We assume that the phase shift for the problem

$$d^2\phi_l(x)/dx^2 + \gamma^2(x)\phi_l(x) = 0, \tag{3}$$

$$\gamma^2(x) = 1 - V(x) - l(l+1)/x^2 \tag{4}$$

is known, whether by analytical solution, numerical calculation or experimental measurement. Instead of solving Eq. (1) directly, we will first transform it into similarity with Eq. (3). To preserve the Helmholtz character of the equation it is necessary simultaneously to transform both the field function and the position variable:

$$x = x(r); \quad r = r(x); \quad r' = dr/dx; \quad x' = dx/dr, \tag{5}$$

$$\chi(x) = \psi_l[r(x)][r'(x)]^{-\frac{1}{2}}. \tag{6}$$

Then Eq. (1) becomes

$$d^2\chi(x)/dx^2 + \gamma^2(x)\chi(x) = -\{g(x) + \kappa^2[r(x)][r'(x)]^2 - \gamma^2(x)\}\chi(x), \tag{7}$$

where $g(x)$ is one half of the Schwartzian derivative,

$$g(x) = \frac{1}{2}\frac{r'''(x)}{r'(x)} - \frac{3}{4}\left(\frac{r''(x)}{r'(x)}\right)^2. \tag{8}$$

From Eq. (7) it is possible to go in a number of different directions, depending upon the choice

of $r(x)$. We hold in abeyance this choice, but restrict the transformation to the following asymptotic form:

$$dr/dx \underset{r\to\infty}{\to} 1. \tag{9}$$

Then, as $r \to \infty$,

$$r \to x + \alpha_l; \qquad \alpha_l = \text{constant}, \tag{10}$$

and

$$\chi(x) \underset{r,x\to\infty}{\to} \chi(r - \alpha_l) = \psi_l(r).$$

For $U(r)$ and $V(x)$ short-range (dropping off faster than $1/r$ or $1/x$), $\chi(x)$ has the asymptotic form

$$\chi(x) \underset{x\to\infty}{\to} \sin [x - (l\pi/2) + \delta_l]. \tag{11}$$

So,

$$\psi(r) \underset{r\to\infty}{\to} \sin [r - (l\pi/2) + (\delta_l - \alpha_l)]. \tag{12}$$

The phase shift which we seek is

$$\eta_l(U) = (\delta_l - \alpha_l). \tag{13}$$

Here, δ_l is the phase shift of the transformed wavefunction, $\chi(x)$. An exact expression can be written for δ_l by converting Eq. (7) to integral (Fredholm) form. We note that the regular solution to the homogeneous part of Eq. (7), Eq. (3), asymptotically goes as

$$\phi_l(x) \underset{x\to\infty}{\to} \sin [x - (l\pi/2) + \eta_l(V)], \tag{14}$$

and that the irregular solution can be chosen to go as

$$\bar{\phi}_l(x) \underset{x\to\infty}{\to} -\cos [x - (l\pi/2) + \eta_l(V)] \tag{15}$$

($\eta_l(V)$ is assumed known). Consistent with this choice, the Green's function for Eq. (3) is taken to be

$$G_v(x \mid x^0) = \phi_l(x_<)\bar{\phi}_l(x_>), \tag{16}$$

and the integral equation for $\chi(x)$ is

$$\chi(x) = \phi_l(x) - \int_0^\infty G_v(x \mid x^0) \times \{g(x^0) + \kappa^2[r(x^0)][r'(x^0)]^2 - \gamma^2(x^0)\}\chi(x^0)\, dx^0. \tag{17}$$

Inserting Eqs. (14), (15), and (16) into Eq. (17) to get the asymptotic form of χ, we see that δ_l is

$$\delta_l = \eta_l(V) + \tan^{-1} \int_0^\infty \phi_l(x) \times \{g(x) + \kappa^2[r(x)][r'(x)]^2 - \gamma^2(x)\}\chi(x)\, dx. \tag{18}$$

The form of α_l will depend upon the choice of transformation function $r(x)$, and here we show the consequences of three different choices.

A. Jeffreys Transformation

$$\int_{x_t}^{x} \gamma(x)\, dx = \int_{r_t}^{r} \kappa(r)\, dr; \qquad \frac{dr}{dx} = \frac{\gamma(x)}{\kappa(r)}; \tag{19}$$

the subscript t designates either the zero of the integrand or, if the integrand has no real, positive zero it indicates that the integral extends to the origin. We exclude from consideration problems involving more than one zero (scattering from a strong attractive potential that drops off fast). Then, if the zeros of $\gamma(x)$ and $\kappa(r)$ are of the same order, dr/dx is continuous. Evaluating Eq. (19) as x and r go to ∞, we see that

$$-\alpha_l = \eta_l^{\mathrm{J}}(U) - \eta_l^{\mathrm{J}}(V), \tag{20}$$

where J designates the Jeffreys form of the WKB approximation,

$$\eta_l^{\mathrm{J}}(U) = \int_{r_t}^{\infty} \left[1 - U(r) - \frac{l(l+1)}{r^2}\right]^{\frac{1}{2}} dr - \int_{r_{t^0}}^{\infty} \left[1 - \frac{l(l+1)}{r^2}\right]^{\frac{1}{2}} dr. \tag{21}$$

Then, inserting Eqs. (20) and (18) into Eq. (13), we get

$$\eta_l(U) = \eta_l(V) - \eta_l^{\mathrm{J}}(V) + \eta_l^{\mathrm{J}}(U) + \tan^{-1} \int_0^\infty \phi_l(x)g(x)\chi(x)\, dx. \tag{22}$$

B. Langer Transformation

$$\int_{x_t}^{x} [1 - V(x) - (l+1/2)^2/x^2]^{\frac{1}{2}}\, dx = \int_{r_t}^{r} [1 - U(r) - (l+1/2)^2/r^2]^{\frac{1}{2}}\, dr; \tag{23}$$

then,

$$-\alpha_l = \eta_l^{\mathrm{L}}(U) - \eta_l^{\mathrm{L}}(V), \tag{24}$$

where L designates the Langer form of the WKB approximation,

$$\eta_l^{\mathrm{L}}(U) = \int_{r_t}^{\infty} [1 - U(r) - (l+1/2)^2/r^2]^{\frac{1}{2}}\, dr - \int_{r_{t^0}}^{\infty} [1 - (l+1/2)^2/r^2]^{\frac{1}{2}}\, dr. \tag{25}$$

This gives

$$\eta_l(U) = \eta_l(V) - \eta_l^{\mathrm{L}}(V) + \eta_l^{\mathrm{L}}(U) + \tan^{-1} \int_0^\infty \phi_l(x) \times \left\{g(x) + \frac{1}{4x^2}\left[\left(\frac{x}{r}\frac{dr}{dx}\right)^2 - 1\right]\right\}\chi(x)\, dx. \tag{26}$$

C. Born transformation

$$\int_0^x \{1 - [xj_l(x)]^2 V(x)\}\, dx = \int_0^r \{1 - [rj_l(r)]^2 U(r)\}\, dr; \quad (27)$$

then

$$-\alpha_l = \eta_l^{\mathrm{B}}(U) - \eta_l^{\mathrm{B}}(V), \quad (28)$$

where B designates the Born approximation,

$$\eta_l^{\mathrm{B}}(U) = \int_0^\infty [rj_l(r)]^2 U(r)\, dr. \quad (29)$$

Therefore,

$$\eta_l(U) = \eta_l(V) - \eta_l^{\mathrm{B}}(V) + \eta_l^{\mathrm{B}}(U) + \tan^{-1} \int_0^\infty \phi_l(x) \times \left\{ g(x) + \gamma^2(x) \left[\frac{\kappa^2(r)}{\gamma^2(x)} \frac{1 - [xj_l(x)]^2 V(x)}{1 - [rj_l(r)]^2 U(r)} \right] \right\} \chi(x)\, dx. \quad (30)$$

Here we have written three different expressions for $\eta_l(U)$, Eqs. (22), (26), and (30). All of them are exact, and in each we have separated out a group of known or readily calculable terms which together, hopefully, constitute a good approximation to $\eta_l(U)$. We will call this the comparison approximation. The last term in each case, the arctangent term, which we would like to drop, is difficult to evaluate. It involves an integral over the weighted product of the two generally unknown wavefunctions $\phi_l(x)$ and $\chi(x)$. The weighting function in each expression, the bracketed part of the integrand, plays the role of a residual scattering potential (RSP) and, to a large measure, the importance of the arctangent term depends upon the magnitude of this potential. The transformations we have used here have the effect of eliminating terms of the order of $(U - V)$ from the RSP and leaving only terms of the order of the derivatives or powers of $(U - V)$. This doesn't guarantee that the RSP is small, but it is a strong indication when U and V are similar.

The justification for dropping the arctangent terms from Eqs. (22), (26), and (30) must finally be found in an empirical test of the adequacy of the first three terms on the right-hand sides of these equations. We require, here, that the arctangent terms be small not only compared to the other terms collectively and individually, but also small compared to

$$[\eta_l(V) - \eta_l^{\mathrm{J,L,B}}(V)] \quad \text{and} \quad [\eta_l^{\mathrm{J,L,B}}(U) - \eta_l^{\mathrm{J,L,B}}(V)].$$

This is sufficient to ensure that the comparison approximation is superior to either $\eta_l(U) \cong \eta_l(V)$ or $\eta_l(U) \cong \eta_l^{\mathrm{J,L,B}}(U)$.

It is possible to penetrate, partially, the calculational thicket surrounding the arctangent term by successive applications of the same transformational scheme. This requires that the RSP be evaluated (or approximated) but avoids explicit reference, to whatever order desired, to the wavefunctions. We illustrate this for the next order.

Referring to Eq. (13) we see that the arctangent term in the exact expression for the phase shift, Eq. (22), resulted from the evaluation of the term δ_l [Eq. (18)], the phase shift for the transformed equation. The total effective potential in the transformed equation, Eq. (7), is

$$V_{\mathrm{total}} = V(x) - \mathrm{RSP} = V(x) - g_{\mathrm{A}}(x), \quad (31)$$

where the subscript A designates either Jeffreys, Langer, or Born. Therefore,

$$\delta_l = \eta_l(V - g_{\mathrm{A}}). \quad (32)$$

In neglecting the arctangent term we are using, essentially, the approximation

$$\eta_l(V - g_{\mathrm{A}}) \cong \eta_l(V). \quad (33)$$

However, we can improve upon this by using the comparison approximation on $\eta_l(V - g_{\mathrm{A}})$. Thus, to better accuracy we write:

$$\eta_l(V - g_{\mathrm{A}}) \cong \eta_l(V) - \eta_l^{\mathrm{A}}(V) + \eta_l^{\mathrm{A}}(V - g_{\mathrm{A}}), \quad (34)$$

in which case

$$\eta_l(U) \cong \eta_l(V) - \eta_l^{\mathrm{A}}(V) + \eta_l^{\mathrm{A}}(U) + [\eta_l^{\mathrm{A}}(V - g_{\mathrm{A}}) - \eta_l^{\mathrm{A}}(V)]. \quad (35)$$

The bracketed term here represents an approximation to the arctangent term, but its evaluation involves only the calculational method (Jeffreys, Langer, Born, or other) taken as the basis for the transformation. In principle, this scheme can be repeatedly applied to the problem, much in the spirit of an iterative approach. Equation (7) is treated as a fresh, new, insoluble equation with potential $(V - g_{\mathrm{A}})$, then transformed into similarity with Eq. (3) (still with potential V), creating a new RSP. However, in actuality, $g_{\mathrm{A}}(x)$ is difficult to evaluate, and higher-order schemes would be much worse, so the utility of this method terminates with Eq. (35).

III. COMPARISON PHASE SHIFTS

The approximation

$$\eta_l^{c\mathrm{A}}(U) = \eta_l(V) - \eta_l^{\mathrm{A}}(V) + \eta_l^{\mathrm{A}}(U), \quad (36)$$

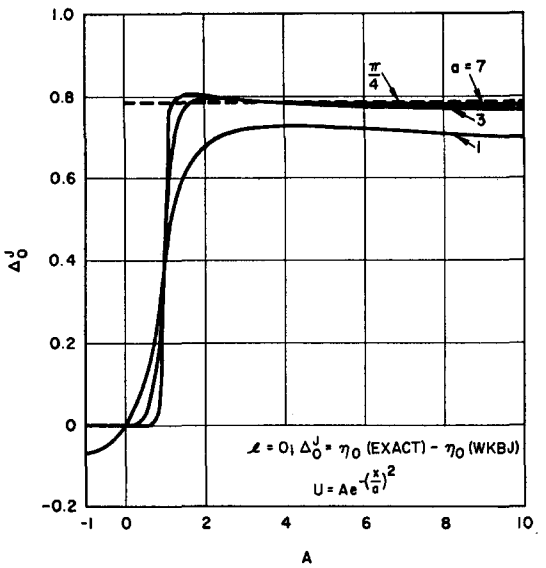


FIG. 1. Phase-shift error in the Jeffrey's approximation.

which we designate the comparison approximation, has followed more or less naturally from the transformation of the radial wave equation. Here the superscript A indicates either the Jeffreys, Langer, or Born approximations, but from the derivational scheme it is apparent that the same form could be written with A indicating many other approximations as well. Basically, the comparison approximation asserts that if U and V are similar, the errors in the η_l^A are about the same, i.e., if we take

$$\Delta_l^A(U) = \eta_l(U) - \eta_l^A(U), \tag{37}$$

then the comparison approximation results if we take

$$\Delta_l^A(U) \cong \Delta_l^A(V). \tag{38}$$

This is manifestly true if U and V are infinitesimally different, and it almost always breaks down if they are radically different. The problem is to determine how "different" U and V can be before the approximation fails. This is not the sort of thing that can be described in a quantitative way, since the accuracy of Eq. (36) varies with the detailed structure of U and V.

In order to make a qualitative assessment of the permissible difference in U and V, we first examine the behavior of $\Delta_l^{J,L}$ in the parameter space of a given potential. The difference in the Δ's at two points is equal to the error in the comparison approximation on the potentials at those points, so this constitutes a test of the method and serves to identify the sensitive regions. A Gaussian potential is used for this purpose. It resembles many of the potentials used in nuclear physics, has two parameters, which gives versatility to the shape without unduly complicating the coverage of the parameter space, and drops off sharply at long range, thus easing the burden of numerical evaluation.

In Fig. 1 we show Δ_0^J in the amplitude interval from $A = -1$ (mildly attractive) to $A = +10$ (strongly repulsive) for range values $a = 1, 3, 7$, roughly short, medium and long range. This indicates that the sensitive region for the Jeffreys based comparison method, S wave, is in the vicinity of $A = +1$. It is at this point, $A = +1$, that the zeros of the integrand in the expression for the Jeffreys phase shift, Eq. (21), first appear on the real axis in the complex r plane. The two zeros (for a Gaussian or Gaussian-type potential) approach the origin along the + and − imaginary axes as A approaches $+1$ from below, and separate along the + and − real axes as A passes beyond $+1$. When they coalesce at $A = +1$ there is a weak divergence in the derivative of η_0^J with respect to A, and this is the source of the sudden jump in Δ_0^J.

The $\frac{1}{4}\pi$ error in the phase shift for large positive A and medium or long range is a consequence of the improper behavior of the free-space Jeffreys wavefunction. For small or zero A, this wavefunction has a cusp at the origin which pulls the wave in by $\frac{1}{4}\pi$. This error is approximately cancelled for small A by subtracting off the asymptotic phase for the free-space wavefunction, which has the same error in it. But for strong repulsive potentials, where the wavefunction is forced into a smooth approach to the origin by the potential shielding, this error doesn't develop in the wavefunction and the subtraction of the free-space asymptotic phase introduces an uncompensated $\frac{1}{4}\pi$ error. Thus we have the ironical situation in which the Jeffreys phase shift is wrong when the Jeffreys wavefunction is correct, and correct when the wavefunction is wrong. At any rate, the $\frac{1}{4}\pi$ error is predictable for this type of potential, and consequently the method is useful. For higher l values, the same mechanism is operative and in general, for strong, long-range potentials of the Gaussian type, the error in η_l^J is

$$\begin{aligned}\eta_l - \eta_l^J &= \tfrac{1}{2}\pi\{(l+\tfrac{1}{2}) - [l(l+1)]^{\frac{1}{2}}\} \\ &= \tfrac{1}{4}\pi \quad \text{for} \quad l = 0, \\ &= 0.135 \quad\quad l = 1, \\ &= 0.079 \quad\quad l = 2.\end{aligned}$$

The effect of the large slope near $A = +1$ is that, with this sort of potential, U and V must have very nearly the same amplitudes at the origin.

The ranges of U and V are not so closely circumscribed and can differ by 10% to 20%.

In Fig. 2 we show the same thing for the Langer approximation. Here the errors do not build up monotonically to a large value, but the region around $A = +1$ is still critical.

It is interesting to note that, for very weak repulsive and weak to moderate attractive potentials of medium or long range, the S-wave Jeffreys approximation is considerably superior to the Langer form (and to the Born also).

As l increases, the variation of Δ with A becomes increasingly smooth everywhere and the region near $A = +1$ becomes less critical. Thus, the S-wave phase shift is the most difficult case, and if that is properly accommodated near $A = +1$, the rest of the potential parameter space is not a serious problem. One consideration encountered at the higher l values is that the important region of the potential, the region contributing most to the scattering, moves out toward the tail. Then it is the similarity of U and V in that region which matters. The comparison approximation shows a surprising tolerance to tail differences, however, as we show in Table I.

The variation of the error in the Jeffreys phase shift with angular momentum is shown in Fig. 3 for a repulsive potential. It is characteristic of this type of potential that the error in the WKB phase shift (both Jeffreys and Langer forms) changes sign in the vicinity of $l = 1.5a$ to $l = 2a$. When a single comparison potential V is used in the evaluation of all of the significant phase shifts for U, the comparison approximation is inferior to the WKB approximations in the vicinity of this crossover point. This results because the crossover points for U and V will generally be slightly different, and cancellation of the errors will not occur for l values thereabouts. Generally, however, the errors in either approximation are small in that region. This effect can also be seen in Table I.

It is possible to conjure up a number of "criteria of similarity" between U and V: equal integrals over the potentials, equal WKB phase shifts, equal integrals over the potentials outside the turning points, equal turning points, etc. However, unless the criterion bears some relationship to the arctangent terms which we are dropping, there is no reason to expect that it will be systematically superior to a subjective estimate of equivalence. It is also true that we are not, in general, free to adjust the parameters of V to best resemble U. Usually we have a set of phase shifts for one or a group of possible

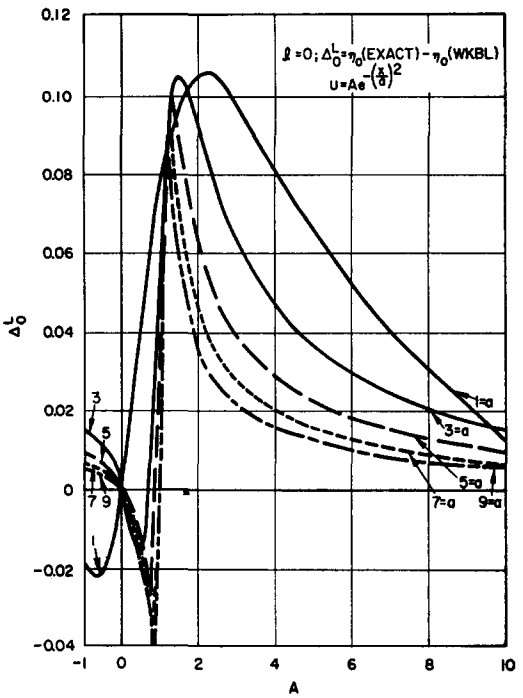


FIG. 2. Phase-shift error in the Langer approximation.

comparison potentials, V, and must use what is available.

Consistent with these usually operative limitations, we illustrate the comparison approximation with a specific example; for U and V we take

$$U(x) = A \exp -(x/a)^2;$$

$$V(x) = A \cosh^{-2}(x/0.8862a).$$

This choice of V satisfies the simplest "criterion of similarity", equal integrals over potentials, and should give us some idea of how far we can trust a criterion which is independent of l. The equal amplitudes at the origin guarantee that the sensitive region near $A = +1$ will be correctly treated. In the main body, the two potentials are quite similar, differing at most by 8% there, but they have much different tail behavior. The difference between U and V is comparable to what might be encountered in a search routine. In Table I we list the exact phase shift, and the Jeffreys, Langer, Jeffreys comparison and Langer comparison errors for the significant l values of a number of different potentials.

The conclusion, drawn from this table and from a larger set of calculations on other potentials, is that, roughly, with a 10% difference between U and V in the main body, the comparison method works well and is significantly more accurate than the basic approximation. This statement holds for

TABLE I. Phase-shift errors for WKB approximations and comparison WKB approximations.

A	l	η_{exact}	Δ_J	Δ_{cJ}	Δ_L	Δ_{cL}
$a = 1.0$						
−1.0	0	+0.31558	−0.07038	0.00241	−0.01836	−0.00016
	1	+0.05033	−0.02208	0.00214	−0.00494	0.00409
	2	+0.00470	0.00359	0.00209	0.00384	0.00119
−0.3	0	+0.08739	−0.03924	−0.00187	−0.01569	−0.00284
+0.3	0	−0.08066	0.06036	0.00418	0.02229	0.00477
+0.9	0	−0.22248	0.29439	0.02279	0.06961	0.01517
+1.0	0	−0.24374	0.38480	0.03083	0.07594	0.01637
	1	−0.04228	0.00931	−0.00493	−0.00144	−0.00491
	2	−0.00443	−0.00334	−0.00198	−0.00358	−0.00118
+1.1	0	−0.26437	0.47111	0.03775	0.08156	0.01735
+2.0	0	−0.42461	0.67701	0.03708	0.10410	0.01857
+10.0	0	−0.99791	0.69575	−0.00941	0.01217	−0.00942
	1	−0.25458	0.00047	−0.03587	−0.04084	−0.03102
	2	−0.03595	−0.02592	−0.01592	−0.02801	−0.01037
$a = 3.0$						
−1.0	0	+1.15501	−0.00285	−0.00039	0.01475	−0.00162
	1	+1.01271	−0.00621	0.00303	0.01085	0.00251
	2	+0.73761	−0.01943	0.00429	−0.00400	0.00512
	4	+0.15843	−0.01246	−0.00248	−0.00738	−0.00201
	6	+0.01630	0.00342	0.00307	0.00378	0.00259
	8	+0.00111	0.00067	−0.00004	0.00068	0.00041
−0.3	0	+0.37952	−0.00034	0.00001	0.00797	−0.00069
+0.3	0	−0.41977	0.00329	−0.00420	−0.01236	−0.00304
+0.9	0	−1.33004	0.22056	−0.00955	0.03067	−0.00563
+1.0	0	−1.47354	0.41209	0.00602	0.05305	−0.00186
	1	−0.92420	0.07671	0.00890	0.03481	0.00540
	2	−0.53143	0.03895	0.00848	0.02191	0.00665
	4	−0.12120	0.00452	−0.00072	0.00130	−0.00064
	6	−0.01507	−0.00292	−0.00268	−0.00324	−0.00229
	8	−0.00107	−0.00062	−0.00007	−0.00064	0.00006
+1.1	0	−1.61035	0.59607	0.01962	0.07328	0.00154
+2.0	0	−2.50688	0.79796	0.01419	0.09284	0.00527
+10.0	0	−4.31456	0.76639	0.00580	0.01514	0.00568
	1	−3.01794	0.11327	0.00460	0.01226	0.00471
	2	−1.98627	0.05306	0.00186	0.00674	0.00240
	4	−0.65227	0.00168	−0.01003	−0.01107	−0.00835
	6	−0.11799	−0.01884	−0.01674	−0.02116	−0.01448
	8	−0.01017	−0.00579	−0.00138	−0.00591	−0.00026
$a = 7.0$						
−1.0	0	+2.70071	−0.00107	−0.00010	0.00651	−0.00061
	1	+2.64015	−0.00136	0.00014	0.00615	−0.00033
	2	+2.51990	−0.00201	0.00057	0.00542	0.00020
	4	+2.11009	−0.00473	0.00160	0.00235	0.00160
	6	+1.51592	−0.01003	0.00042	−0.00380	0.00098
	10	+0.38270	−0.00641	−0.00186	−0.00427	−0.00168
	14	+0.04627	0.00174	0.00177	0.00197	0.00154
−0.3	0	+0.88631	−0.00018	−0.00001	0.00343	−0.00028
+0.3	0	−0.98747	−0.00041	−0.00007	−0.00729	0.00055
+0.9	0	−3.51036	0.10761	−0.01754	−0.01612	−0.00936
+1.0	0	−3.99312	0.40659	−0.00034	0.03019	−0.00466
	1	−3.32345	0.07750	0.00140	0.02294	−0.00112
	2	−2.73577	0.04664	0.00268	0.01861	0.00079
	4	−1.77698	0.02432	0.00367	0.01273	0.00254
	6	−1.07275	0.01391	0.00324	0.00819	0.00262
	10	−0.28488	0.00225	−0.00002	0.00088	0.00001
	14	−0.04249	−0.00144	−0.00143	−0.00164	−0.00124
+1.1	0	−4.45172	0.69650	0.01538	0.07263	−0.00010
+2.0	0	−6.91572	0.79551	0.00430	0.04512	0.00022
+10.0	0	−11.07823	0.77723	0.00245	0.00654	0.00239
	1	−9.62464	0.12639	0.00238	0.00624	0.00233
	2	−8.28747	0.07062	0.00223	0.00582	0.00222
	4	−5.95571	0.03381	0.00161	0.00437	0.00172
	6	−4.06428	0.01844	0.00053	0.00223	0.00083
	10	−1.50002	0.00146	−0.00353	−0.00396	−0.00280
	14	−0.32669	−0.00823	−0.00776	−0.00956	−0.00677

the Born approximation also, which was tested on Gaussian and square-well potentials.

IV. BORN APPROXIMATION

The Born comparison approximation,

$$\eta_l^{cB}(U) = \eta_l(V) - \eta_l^B(V) + \eta_l^B(U) \tag{39}$$

has been derived by the formal application of the calculational procedures of the transformation method. However, there is a much more natural way to derive it. Equation (1) can be rewritten as

$$\psi_l''(x) + [1 - V(x) - l(l+1)/x^2]\psi_l(x) = [U(x) - V(x)]\psi_l(x). \tag{40}$$

In integral form this is

$$\psi_l(x) = \phi_l(x) + \int_0^\infty G_\phi(x \mid x^0)[U(x^0) - V(x^0)]\psi_l(x^0)\,dx^0, \tag{41}$$

where

$$G_\phi(x \mid x^0) = \phi_l(x_<)\bar{\phi}_l(x_>). \tag{42}$$

This yields, for the phase shift,

$$\eta_l(U) = \eta_l(V) - \tan^{-1} \times \int_0^\infty \phi_l(x)[U(x) - V(x)]\psi_l(x)\,dx. \tag{43}$$

Using the Born approximation on the wavefunctions

$$\phi_l(x) \cong \psi_l(x) \cong xj_l(x), \tag{44}$$

and, dropping the arctangent dependence since these are approximations to the physical wavefunctions, we get

$$\eta_l(U) \cong \eta_l(V) - \eta_l^B(V) + \eta_l^B(U). \tag{45}$$

This form of the comparison approximation is correct to first order in the difference in potential amplitudes. The other forms must reduce to this equation when evaluated to this order. The reduction for the case of the Jeffreys approximation can be readily shown. The full expression for the phase shift is

$$\eta_l(U) \cong \eta_l(V) - \eta_l^J(V) + \eta_l^J(U) + \int_0^\infty [xj_l(x)]^2 \tfrac{1}{2}[r'''(x)]\,dx. \tag{46}$$

Here the RSP and the wavefunctions in the correction term to the comparison approximation have been written correct to first order in the potential amplitudes. The correction term, which we label I, becomes, after integrating by parts twice,

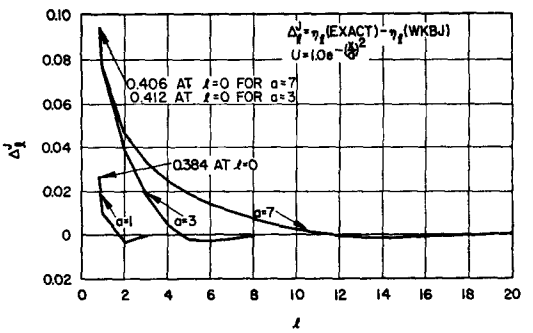


FIG. 3. Phase-shift error in the Jeffrey's approximation as a function of angular momentum.

$$I = \int_0^\infty [r'(x) - 1]\tfrac{1}{2}(d^2/dx^2)[xj_l(x)]^2\,dx. \tag{47}$$

Using the following identity[1]:

$$\tfrac{1}{2}(d^2/dx^2)[xj_l(x)]^2 = 1 - 2[1 - l(l+1)/x^2][xj_l(x)]^2 - 2\int_x^\infty \frac{l(l+1)}{x} j_l^2(x)\,dx, \tag{48}$$

I can be manipulated to yield

$$I = \eta_l^J(V) - \eta_l^J(U) + \eta_l^B(U) - \eta_l^B(V), \tag{49}$$

and the Jeffreys comparison approximation, with corrections, evaluated to first order in the amplitudes becomes

$$\eta_l(U) \cong \eta_l(V) - \eta_l^B(V) + \eta_l^B(U). \tag{50}$$

V. CALCULATION OF ENERGY LEVELS

It is possible to use this same approach in the calculation of bound-state energy levels. This has been done for the one-dimensional case by Miller and Good,[4] and we extend it here to the three-dimensional, spherically symmetric problem. We seek the energy eigenvalue for the equation

$$d^2\psi_l^n/dr^2 + [-K_n^2 - U(r) - l(l+1)/r^2]\psi_l^n = 0, \tag{51}$$

subject to the boundary conditions

$$\psi_l^n(0) = 0; \qquad \psi_l^{n\prime}/\psi_l^n \text{ continuous.} \tag{52}$$

Here n is the radial quantum number. We assume that the energy eigenvalues are known for the equation

$$d^2\phi_l^n/dx^2 + [-k_n^2 - V(x) - l(l+1)/x^2]\phi_l^n = 0. \tag{53}$$

We transform Eq. (53) by

$$\chi_l^n(r) = \phi_l^n[x(r)]/[x'(r)]^{\frac{1}{2}}, \tag{54}$$

[4] S. C. Miller and R. H. Good, Jr., Phys. Rev., **91**, 174 (1953). See also A. Erdélyi, J. Math. Phys. **1**, 16 (1960) for a discussion of the asymptotic properties of Bessel functions, and Hermite and Laguerre polynomials.

and get

$$\frac{d^2\chi_l^n(r)}{dr^2} + \left[\frac{-k_n^2 - V(x) - l(l+1)/x^2}{(r')^2} + g(r)\right] \times \chi_l^n(r) = 0. \quad (55)$$

$g(r)$ is as defined in Eq. (8), and for a Jeffreys-based transformation we take

$$\left(\frac{dr}{dx}\right)^2 = \frac{-k_n^2 - V(x) - l(l+1)/x^2}{-K_n'^2 - U(r) - l(l+1)/r^2}. \quad (56)$$

This yields

$$d^2\chi_l^n(r)/dr^2 + [-K_n'^2 - U(r) + g(r) - l(l+1)/r^2]\chi_l^n(r) = 0. \quad (57)$$

In this, $-K_n'^2$ is the eigenvalue for the potential $(U - g)$. To keep $g(r)$ well behaved, we require that the zeros of the numerator and denominator in Eq. (56) coincide, i.e.,

$$\int_{x_1}^{x_2}\left[-k_n^2 - V(x) - \frac{l(l+1)}{x^2}\right]^{\frac{1}{2}} dx = \int_{r_1}^{r_2}\left[-K_n'^2 - U(r) - \frac{l(l+1)}{r^2}\right]^{\frac{1}{2}} dr, \quad (58)$$

where x_1, x_2, r_1 and r_2 are the interior turning points of the respective integrands.

If we now equate $K_n'^2$ to K_n^2, i.e., neglect the effect of $g(r)$ in Eq. (57), we have the equivalent of the comparison approximation, with Eq. (58) determining the desired eigenvalue, $-K_n^2$.

For U and V similar, this approach should be superior to the standard WKB calculation, which gives

$$\int_{r_1}^{r_2}\left[-K_n^{J^2} - U(r) - \frac{l(l+1)}{r^2}\right]^{\frac{1}{2}} dr = (n + \tfrac{1}{2})\pi. \quad (59)$$

In general, if we were to insert the correct value of the energy $-K_n^2$ into the integral on the left side of Eq. (59), we would get something a bit different from $(n + \frac{1}{2})\pi$. Eq. (58) states that, to a closer approximation, this deviation from $(n + \frac{1}{2})\pi$ is about the same for U and V.

ACKNOWLEDGMENTS

The author wishes to thank Dr. D. S. Saxon who suggested this problem, and who was a constant source of aid and enlightenment throughout. He also thanks the Western Data Processing Center, Graduate School of Business Administration, University of California, Los Angeles, for the use of their IBM 7090 computer, and the Hughes Aircraft Company for the preparation of this manuscript.