JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Transverse Dielectric Tensor for a Free-Electron Gas in a Uniform Magnetic Field

V. ARUNASALAM
Plasma Physics Laboratory, Princeton University, Princeton, New Jersey

(Received 10 August 1968)

An elementary method of calculating that part of the tensor dielectric coefficient which determines the propagation of transverse electromagnetic radiation through a free-electron gas in a uniform external magnetic field is presented. The method presented here is based on a particle-orbit analysis and is somewhat analogous to a generalized version of the Kramers–Heisenberg quantum theory of gaseous dispersion. It is shown that the elements of the transverse dielectric tensor can be obtained from a knowledge of the quantum-mechanical transition probabilities for emission and absorption of photons (that is, from a knowledge of the Einstein A and B coefficients). The formal expression for the dielectric tensor thus obtained is shown to be valid for both the degenerate and the nondegenerate system of electrons. The dielectric tensor thus obtained is shown to reduce in the classical limit to the familiar results of the conventional classical hot plasma kinetic theory. The first quantum correction to the classical hot plasma dielectric tensor is explicitly given and it is shown that, under suitable conditions, this quantum correction will play a significant role in the analysis of the experimental studies of the electrodynamic behavior of "classical electron–hole plasmas" in a uniform external magnetic field.

I. INTRODUCTION

In this paper we are interested in discussing the electrodynamic properties of a free-electron gas in a uniform external magnetic field. In particular, it is our aim to investigate the response of the system to transverse electromagnetic radiations. We shall use the solutions for the particle orbits as the basis for our discussion. In the literature[1–4] there exists an alternative method of handling this problem. There one seeks a kinetic description of the system on the basis of the Vlasov or collisionless Boltzmann equation. The equivalence of the particle-orbit theory and the kinetic description on the basis of the Vlasov equation was first demonstrated by Jeans and is usually referred to as the Jeans theorem. In general, the particle-orbit theory is much more of a physical approach and provides a better insight into the physical processes involved in a particular problem. However, the kinetic description on the basis of the Vlasov equation provides a much more rigorous treatment of complex problems which may not be easily accessible to the analysis of the particle-orbit theory.

In electrodynamics the basic quantity for the description of the collective response of a system is the retarded frequency and wave-vector-dependent dielectric tensor $\mathsf{D}(\omega, \mathbf{k})$. Here ω is the frequency of the electromagnetic wave of wave vector $\mathbf{k}$. In principle, both methods—the kinetic description based on the Vlasov equation and the particle-orbit theory—enable one to calculate the response tensor $\mathsf{D}(\omega, \mathbf{k})$. In this paper we shall be interested in that part of the dielectric tensor which determines the propagation of transverse electromagnetic radiations through the free-electron gas in a uniform external magnetic field.

In the kinetic theory approach one starts from the Vlasov equation which describes the time evolution of the particle velocity distribution function (or one starts from the Liouville theorem, which describes the time evolution of the single-particle density matrix). As a first step one seeks a linearized perturbation solution of the Vlasov equation (or of the equation of motion for the density matrix) in the presence of both the external uniform magnetic field and the radiation field. The uniform magnetic field determines the zero-order trajectory (or the unperturbed Hamiltonian and the corresponding zero-order eigenvalues and eigenstates) of the charged particles. In a coordinate system which follows the zero-order trajectory of the particles, usually referred to as the Lagrangian system of coordinates, one calculates the first-order perturbation to the particle velocity distribution function (or, in the representation in which the unperturbed Hamiltonian is diagonal, one calculates the first-order perturbation to the density matrix). Knowing the first-order perturbed-velocity distribution function (or the density matrix) in terms of the perturbing radiation field, one then calculates the induced macroscopic current density by taking the first moment of the perturbed-velocity distribution function (or by taking the trace of the product of the density matrix and the

[1] J. E. Drummond, *Plasma Physics* (McGraw-Hill Book Company, Inc., New York, 1961), Chap. 2; D. C. Kelly, Phys. Rev. **134A**, 641 (1964); J. J. Quinn and S. Rodriguez, *ibid.* **128**, 2487 (1962); I. B. Bernstein, *ibid.* **109**, 10 (1958); E. N. Adams and T. D. Holstein, J. Phys. Chem. Solids **10**, 254 (1959); N. J. Horing, Ann. Phys. (N.Y.) **31**, 1 (1965); W. R. Chappel, JILA Report No. 35 (University of Colorado, Ph.D. Thesis, 1965).

[2] G. Bekefi, *Radiation Processes In Plasmas* (John Wiley & Sons, Inc., New York, 1966).

[3] T. H. Stix, *The Theory of Plasma Waves* (McGraw-Hill Book Company, Inc., New York, 1962).

[4] D. C. Montgomery and D. A. Tidman, *Plasma Kinetic Theory* (McGraw-Hill Book Company, Inc., New York, 1964).

unperturbed current-density operator). This induced macroscopic current density yields directly the conductivity or the mobility tensor, which, in turn, yields directly the required dielectric tensor $\mathsf{D}(\omega, \mathbf{k})$.

The particle-orbit theory to be presented below is somewhat analogous to (a generalized version of) the Kramers–Heisenberg[5,6] quantum theory of gaseous dispersion. The underlying physical principle behind the Kramers–Heisenberg method is that the dispersive properties of a gas of atoms are simply a manifestation of a balance between the two competing processes of photon emission (spontaneous plus induced or stimulated emission) and photon absorption by the atoms of the gas. It is our aim in this paper to show that the response of a system of "Landau electrons" to transverse electromagnetic radiations is indeed a consequence of such a balance between photon emission and photon absorption.

We will begin by assuming Landau's quantized particle motion of the electrons in a uniform magnetic field[7]; then we calculate the transition probabilities for emission (spontaneous plus the induced emission) and absorption of a photon of momentum $\hbar\mathbf{k}$, energy $\hbar\omega$, and polarization vector $\boldsymbol{\epsilon}_{ks}$ by a "Landau electron." From these we then derive the energy-balance equation by applying the principle of detailed balance for a differential path length along the photon's trajectory in the free-electron gas. We then seek the solution of this energy-balance equation and thus derive equations for the equilibrium photon-number density and the absorption coefficient per unit path length of a photon of definite momentum $\hbar\mathbf{k}$, energy $\hbar\omega$, and polarization vector $\boldsymbol{\epsilon}_{ks}$. We then examine the angular dependence of the absorption coefficient $\tau(\omega, \mathbf{k})$ and thus construct an absorption tensor $\tau_{ij}(\omega, \mathbf{k})$. We then show that the Hermitian part of the absorption tensor $\boldsymbol{\tau}(\omega, \mathbf{k})$ is directly proportional to the anti-Hermitian part of the transverse dielectric tensor $\mathsf{D}(\omega, \mathbf{k})$. We then obtain the transverse dielectric tensor $\mathsf{D}(\omega, \mathbf{k})$ by making use of the fact that the real and imaginary parts of $D_{ij}(\omega, \mathbf{k})$ must satisfy the well-known Kramers–Kronig relations as a consequence of the causality principle.[8] It will be seen that the formal expression for $D_{ij}(\omega, \mathbf{k})$ thus obtained is valid for both the degenerate and the nondegenerate system of electrons. That is, the formal expression for $D_{ij}(\omega, \mathbf{k})$ is unaffected, whether or not one takes account of the Pauli exclusion principle, as long as one uses the appropriate distribution function of the electrons (Fermi–Dirac statistics for a degenerate system of electrons and Maxwell–Boltzmann statistics for a nondegenerate system of electrons). Finally, we will show that the classical limit of the dielectric tensor $D_{ij}(\omega, \mathbf{k})$ thus obtained is in complete agreement with the familiar results of the conventional classical hot plasma kinetic theory (whose starting point is the Vlasov equation), and we will give the first quantum correction to the classical hot plasma dielectric tensor. It will be seen that this quantum correction should prove useful (under some conditions) in examining the electrodynamic behavior of "classical electron–hole plasmas" in a uniform external magnetic field.

II. REVIEW OF BASIC CONCEPTS

We consider the motion of a free electron of charge q and mass μ in a uniform magnetic field $\mathbf{B} = B\mathbf{i}_z$, where $\mathbf{i}_\alpha$ denotes the unit vector along the α axis. (See Fig. 1.) Let

$$\begin{aligned} \mathbf{R} &= x\mathbf{i}_x + y\mathbf{i}_y + z\mathbf{i}_z = \mathbf{r} + z\mathbf{i}_z, \\ \mathbf{V} &= v_x\mathbf{i}_x + v_y\mathbf{i}_y + v_z\mathbf{i}_z = \mathbf{v} + v_z\mathbf{i}_z \end{aligned} \tag{1}$$

be the position and velocity vectors of the electron. One can show that the energy-level spectrum of the electron is given by[7]

$$E_{n,v_z} = (n + \tfrac{1}{2})\hbar\omega_b + \tfrac{1}{2}\mu v_z^2, \tag{2}$$

where $n = 0, 1, 2, \cdots, \infty$ and $\omega_b = (qB/\mu c)$ is the electron-cyclotron frequency. The nonzero matrix elements of the perpendicular velocity and position

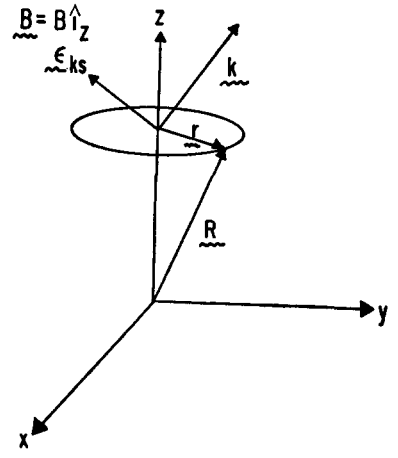

FIG. 1. Motion of an electron in Cartesian coordinates.

[5] H. G. Kuhn, F.R.S., *Atomic Spectra* (Academic Press Inc., New York, 1962), pp. 59–68.

[6] H. A. Kramers, Nature **113**, 673 (1924); **114**, 310 (1924); G. Breit, *ibid.* **114**, 310 (1924).

[7] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1958); V. Arunasalam, Princeton Plasma Physics Laboratory MATT-439, 1966 (unpublished).

[8] C. Kittel, *Elementary Statistical Physics* (John Wiley & Sons, Inc., New York, 1958); N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields*, English transl. by G. M. Volkoff (Interscience Publishers, Inc., New York, 1959), Sec. 46.2.

operators are given by

$$v_{x_{n,n-1}} = (v_{x_{n-1,n}})^* = -iv_{y_{n,n-1}} = (iv_{y_{n-1,n}})^* = (\hbar\omega_b/2\mu)^{\frac{1}{2}} n^{\frac{1}{2}} e^{-i\phi_n}, \quad (3a)$$

$$x_{n,n-1} = (x_{n-1,n})^* = -iy_{n,n-1} = (iy_{n-1,n})^* = -i(\hbar/2\mu\omega_b)^{\frac{1}{2}} n^{\frac{1}{2}} e^{-i\phi_n}, \quad (3b)$$

where ϕ_n is an arbitrary phase factor and the asterisk means the complex conjugate. The matrix elements of $e^{\pm ik_z z}$ are given by

$$\langle v_z' | e^{\pm ik_z z} | v_z \rangle = \delta_{v_z', v_z \pm \hbar k_z/\mu}. \quad (4)$$

The photon–electron interaction Hamiltonian that is responsible for transitions in which only one light quantum is involved is given by

$$H = -\frac{q}{2c}\left[\mathbf{A}\cdot\left(\mathbf{v} - \frac{i\hbar}{\mu}\frac{\partial}{\partial z}\mathbf{i}_z\right) + \left(\mathbf{v} - \frac{i\hbar}{\mu}\frac{\partial}{\partial z}\mathbf{i}_z\right)\cdot\mathbf{A}\right], \quad (5)$$

where the vector potential $\mathbf{A}(\mathbf{R})$ for the radiation field in a box of volume L^3 can be written in terms of the usual creation and annihilation operators as[9,10]

$$\mathbf{A}(\mathbf{R}) = \sum_{\mathbf{k}}\left(\frac{2\pi\hbar c^2}{L^3\omega}\right)^{\frac{1}{2}} \sum_{s=1,2} (\boldsymbol{\epsilon}_{ks} a_{ks} e^{i\mathbf{k}\cdot\mathbf{R}} + \boldsymbol{\epsilon}_{ks}^* a_{ks}^+ e^{-i\mathbf{k}\cdot\mathbf{R}}), \quad (6)$$

where $\mathbf{k}\cdot\boldsymbol{\epsilon}_{ks} = 0$ and $|\boldsymbol{\epsilon}_{ks}|^2 = 1$. According to the golden rule of time-dependent perturbation theory, the transition probability $j(f;i)$ from an initial state $|i\rangle$ of energy E_i to a final state $|f\rangle$ of energy E_f is given by[9]

$$j(f;i) = (2\pi/\hbar)\,|\langle f|\,H\,|i\rangle|^2\,\delta(E_f - E_i). \quad (7)$$

We now consider the two fundamental processes of absorption j_A and emission j_E illustrated in Fig. 2.

From Eqs. (1), (2), and (4)–(7) we get

$$j_A = N_{ks}\left(\frac{4\pi^2 q^2}{L^3\hbar\omega}\right)|\langle n+l|\,\boldsymbol{\epsilon}_{ks}\cdot\mathbf{M}\,|n\rangle|^2\,\delta_{v_z', v_z+\hbar k_z/\mu} \times \delta[\omega - l\omega_b - k_z(v_z + \hbar k_z/2\mu)],$$

$$j_E = (N_{ks}+1)\left(\frac{4\pi^2 q^2}{L^3\hbar\omega}\right)|\langle n|\,\boldsymbol{\epsilon}_{ks}^*\cdot\mathbf{M}^\dagger\,|n+l\rangle|^2\,\delta_{v_z', v_z+\hbar k_z/\mu} \times \delta[\omega - l\omega_b - k_z(v_z + \hbar k_z/2\mu)], \quad (8)$$

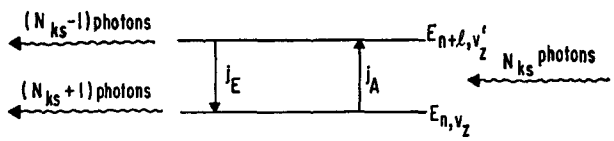

FIG. 2. Emission and absorption of a photon by a "Landau electron."

where

$$\mathbf{M} = [\tfrac{1}{2}(e^{i\mathbf{k}\cdot\mathbf{r}}\mathbf{v} + \mathbf{v}e^{i\mathbf{k}\cdot\mathbf{r}}) + \mathbf{i}_z(v_z + \hbar k_z/2\mu)e^{i\mathbf{k}\cdot\mathbf{r}}]. \quad (9)$$

$\mathbf{M}^\dagger$ is the Hermitian conjugate of $\mathbf{M}$ and is simply obtained by replacing i by $-i$ in the above expression for $\mathbf{M}$ since the perpendicular velocity and position operators $\mathbf{v}$ and $\mathbf{r}$ are Hermitian and N_{ks} represents the number of light quanta all having a momentum $\hbar\mathbf{k}$, energy $\hbar\omega$, and polarization vector $\boldsymbol{\epsilon}_{ks}$. To evaluate the matrix elements of $\mathbf{M}$ and $\mathbf{M}^\dagger$ we can use the simplest form of the multipole expansion

$$e^{\pm i\mathbf{k}\cdot\mathbf{r}} = \sum_{s=0}^{\infty}(1/s!)\{\pm i(k_x x + k_y y)\}^s = \sum_{s=0}^{\infty}(1/s!)\{\pm i(k_+ r_- + k_- r_+)\}^s, \quad (10)$$

where $k_\pm = (2^{-\frac{1}{2}})(k_x \pm ik_y)$ and $r_\pm = (2^{-\frac{1}{2}})(x \pm iy)$ and the nonzero matrix elements of r_+ and r_- are given by

$$r_{+_{n-1,n}} = (r_{-_{n,n-1}})^* = i(\hbar/\mu\omega_b)^{\frac{1}{2}} n^{\frac{1}{2}} e^{i\phi_n}. \quad (11)$$

III. THEORY OF THE DIELECTRIC TENSOR

Our aim now is to derive the energy-balance equation. Let the box of volume L^3 under consideration contain $N_0 F(E_{n,v_z})$ electrons per unit volume in the quantum state $|E_{n,v_z}\rangle$. We assume that the probability function $F(E_{n,v_z})$ is normalized so that

$$\int dv_z \sum_{n=0}^{\infty} F(E_{n,v_z}) = 1. \quad (12)$$

The detailed balance relation between the probabilities per unit volume of emission and absorption for a differential path length $d\chi$ of a photon of momentum $\hbar\mathbf{k}$, energy $\hbar\omega \approx \hbar l\omega_b$ (where the cyclotron harmonic number l may take any one of the values[11] $0, \pm1, \pm2, \cdots, \pm\infty$), and polarization vector $\boldsymbol{\epsilon}_{ks}$ may be written as

$$\left(\frac{\partial N_{ks}}{\partial\chi}\right)_l = \left(\frac{L^3 N_0}{\omega/k}\right)\int dv_z \times \sum_{n=0}^{\infty}[F(E_{n+l,v_z'})j_E - F(E_{n,v_z})j_A]. \quad (13)$$

This equation (13) is usually referred to as the energy-balance equation or as the equation of energy transfer.[12,13] Since j_E and j_A are proportional to $(N_{ks}+1)$

[9] W. Heitler, *The Quantum Theory of Radiation* (Clarendon Press, Oxford, England, 1954).

[10] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1955).

[11] The emission and absorption corresponding to $l = 0$ are essentially the Čerenkov emission and absorption processes. The positive values (that is, $l = +1, +2, \cdots, +\infty$) and the negative values (that is, $l = -1, -2, \cdots, -\infty$) of l correspond to emission and absorption of circularly polarized plane electromagnetic waves (or photons) whose sense of rotation is the same as and opposite to that of the gyrating electrons, respectively.

[12] S. Chandrasekhar, *Stellar Structure* (Dover Publications, Inc., New York, 1957).

[13] G. Bekefi and S. C. Brown, Am. J. Phys. **29**, 404 (1961).

and N_{ks}, respectively, the solution of Eq. (13) for a uniform homogeneous system is

$$N_{ks} = N_{ks}^{(0)}[1 - e^{-\tau^{(l)}(\omega,\mathbf{k})\chi}] \tag{14}$$

if we assume that $N_{ks} = 0$ at $\chi = 0$. Here,

$$\tau^{(l)}(\omega, \mathbf{k}) = \left(\frac{L^3 N_0}{\omega/k}\right)\int dv_z \sum_{n=0}^{\infty}[F(E_{n,v_z})(j_A/N_{ks}) - F(E_{n+l,v_{z'}})(j_E/N_{ks} + 1)] \tag{15}$$

is the absorption coefficient per unit path length for a photon of momentum $\hbar\mathbf{k}$, energy $\hbar\omega \approx l\hbar\omega_b$, and polarization vector $\boldsymbol{\epsilon}_{ks}$, and

$$N_{ks}^{(0)} = \left[\left(\frac{L^3 N_0}{\omega/k}\right)\int dv_z \times \sum_{n=0}^{\infty} F(E_{n+l,v_{z'}})(j_E/N_{ks} + 1)\right] \Big/ \tau^{(l)}(\omega, \mathbf{k}) \tag{16}$$

is the number of photons at steady state. Since the absorption coefficient $\tau^{(l)}(\omega, \mathbf{k})$ is real, by making use of Eqs. (8) and (9), Eq. (15) could be written in the form

$$\tau^{(l)}(\omega, \mathbf{k}) = \sum_i \sum_j \mathrm{Re}\,[\tau_{ij}^{(l)}(\omega, \mathbf{k})(\boldsymbol{\epsilon}_{ks}^*)_i(\boldsymbol{\epsilon}_{ks})_j], \tag{17}$$

where $i, j = x, y$, and z; Re stands for the real part, and

$$\tau_{ij}^{(l)}(\omega, \mathbf{k}) = \left(\frac{N_0 4\pi^2 q^2}{\hbar\omega(\omega/k)}\right)\int dv_z \times \sum_{n=0}^{\infty}\{[F(E_{n,v_z}) - F(E_{n+l,v_z+\hbar k_z/\mu})] \times \Pi_{ij}^{(l)}\delta[\omega - l\omega_b - k_z(v_z + \hbar k_z/2\mu)]\}, \tag{18}$$

where

$$\Pi_{ij}^{(l)} = [\langle n|\, M_i^+ \,|n + l\rangle][\langle n + l|\, M_j \,|n\rangle]. \tag{19}$$

In deriving the above expression for $\tau_{ij}^{(l)}(\omega, \mathbf{k})$ we have used the convention $|Z|^2 = [Z^*][Z]$ for any complex number Z.

It is relatively easy to show that (after Fourier analysis in space and time) Maxwell's electromagnetic field equations for plane waves of the form

$$\mathbf{E}(\mathbf{R}, t) = \boldsymbol{\epsilon}_{ks} \exp\,[i(\mathbf{k}\cdot\mathbf{R} - \omega t)]$$

in a medium of dielectric tensor $\mathsf{D}(\omega, \mathbf{k})$ give[2,3]

$$\mathbf{k} \times (\mathbf{k} \times \boldsymbol{\epsilon}_{ks}) + (\omega^2/c^2)\mathsf{D}\cdot\boldsymbol{\epsilon}_{ks} = 0. \tag{20}$$

Letting $k = \mathrm{Re}\,k + i\,\mathrm{Im}\,k$ and setting $\mathbf{k}\cdot\boldsymbol{\epsilon}_{ks} = 0$ for transverse electromagnetic waves, from Eq. (20) we obtain

$$2\,\mathrm{Im}\,k = (\omega^2/c^2\,\mathrm{Re}\,k)\,\mathrm{Im}\,[\boldsymbol{\epsilon}_{ks}^*\cdot\mathsf{D}\cdot\boldsymbol{\epsilon}_{ks}]. \tag{21}$$

Since $\mathbf{E}(\mathbf{R}, t) = \boldsymbol{\epsilon}_{ks}\exp\,[i(\mathbf{k}\cdot\mathbf{R} - \omega t)]$, we get the absorption coefficient per unit path length $\tau(\omega, \mathbf{k})$ for a primary electromagnetic beam of wave vector $\mathbf{k}$, frequency ω, and polarization $\boldsymbol{\epsilon}_{ks}$:

$$\tau(\omega, \mathbf{k}) = 2\,\mathrm{Im}\,k = (\omega^2/c^2\,\mathrm{Re}\,k)\sum_i\sum_j \mathrm{Im}\,[D_{ij}(\boldsymbol{\epsilon}_{ks}^*)_i(\boldsymbol{\epsilon}_{ks})_j]. \tag{22}$$

We now write

$$D_{ij} = \delta_{ij} + \sum_l D_{ij}^{(l)}, \tag{23}$$

where $\mathsf{D}^{(l)}(\omega, \mathbf{k})$ is the retarded frequency and wave-vector-dependent dielectric tensor appropriate for the description of the collective response of the system under study to transverse electromagnetic radiation of wave vector $\mathbf{k}$ and frequency $\omega \approx l\omega_b$ (where l may take any one of the values $0, \pm1, \pm2, \cdots, \pm\infty$). From Eqs. (17), (22), and (23) we get the following identity:

$$\sum_i\sum_j \mathrm{Re}\,[\tau_{ij}^{(l)}(\boldsymbol{\epsilon}_{ks}^*)_i(\boldsymbol{\epsilon}_{ks})_j] \equiv (\omega^2/c^2\,\mathrm{Re}\,k)\sum_i\sum_j \mathrm{Im}\,[D_{ij}^{(l)}(\boldsymbol{\epsilon}_{ks}^*)_i(\boldsymbol{\epsilon}_{ks})_j] \tag{24}$$

for all values of $i, j = x, y$, and z. Hence

$$(2i)^{-1}[\mathsf{D}^{(l)} - \mathsf{D}^{(l)+}] = (\tfrac{1}{2})(c^2\,\mathrm{Re}\,k/\omega^2)[\boldsymbol{\tau}^{(l)} + \boldsymbol{\tau}^{(l)+}], \tag{25}$$

where $\mathsf{D}^{(l)+}$ and $\boldsymbol{\tau}^{(l)+}$ are the Hermitian conjugates of $\mathsf{D}^{(l)}$ and $\boldsymbol{\tau}^{(l)}$, respectively. Thus, from Eq. (18), we obtain the retarded frequency and wave-vector-dependent dielectric tensor

$$D_{ij}^{(l)}(\omega, \mathbf{k}) = \lim_{\gamma\to+0}\left(\frac{\mu\omega_p^2}{\hbar\omega^2}\right)\int dv_z \times \sum_{n=0}^{\infty}\frac{[F(E_{n+l,v_z+\hbar k_z/\mu}) - F(E_{n,v_z})]\Pi_{ij}^{(l)}}{\omega - l\omega_b - k_z(v_z + \hbar k_z/2\mu) + i\gamma}, \tag{26}$$

where

$$\omega_p^2 = (4\pi N_0 q^2/\mu), \tag{27}$$

and we have used the symbolic identity

$$\lim_{\gamma\to+0}\frac{1}{\omega - l\omega_b - k_z(v_z + \hbar k_z/2\mu) + i\gamma} = \left\{P\frac{1}{\omega - l\omega_b - k_z(v_z + \hbar k_z/2\mu)} - i\pi\delta[\omega - l\omega_b - k_z(v_z + \hbar k_z/2\mu)]\right\}, \tag{28}$$

where P denotes the principal value. It is clear from this symbolic identity that the real and imaginary parts of $D_{ij}^{(l)}(\omega, \mathbf{k})$ will satisfy the well-known Kramers–Kronig relations[8] in accordance with the causality principle (which states that the effect should not precede the cause). Equations (23) and (26) complete our formal theory of the transverse dielectric tensor for a free-electron gas in a uniform external magnetic field.

IV. COMPARISON WITH THE CLASSICAL HOT PLASMA THEORY

It is physically instructive to compare the classical limit of the dielectric tensor thus obtained with the familiar results of the conventional hot plasma theory.[2–4] We may do this with the aid of the following relations:

$$E_n = (n + \tfrac{1}{2})\hbar\omega_b \to E_\perp = \tfrac{1}{2}\mu v^2,$$

$$F(E_{n,v_z}) \to F(E_\perp, v_z),$$

$$\sum_{n=0}^{\infty} F(E_{n,v_z})[\ ] \to \int_0^\infty dE_\perp F(E_\perp, v_z)[\ ]$$

$$= \int_0^\infty 2\pi v\, dv f(v, v_z)[\ ],$$

$$[F(E_{n+l,v_z+\hbar k_z/\mu}) - F(E_{n,v_z})]$$

$$\to \left[l\hbar\omega_b \frac{\partial}{\partial E_\perp} + \frac{\hbar k_z}{\mu}\frac{\partial}{\partial v_z}\right] F(E_\perp, v_z)$$

$$= \left(\frac{\hbar}{\mu}\right)\left[\frac{l\omega_b}{v}\frac{\partial}{\partial v} + k_z\frac{\partial}{\partial v_z}\right] f(v, v_z), \quad (29)$$

where $E_\perp = \mu v^2/2$ is the kinetic energy of the electrons in a plane perpendicular to the uniform magnetic field $\mathbf{B} = B\mathbf{i}_z$ and $f(v, v_z)$ is the electron velocity distribution function.

On making use of Eqs. (23), (26), and (29), one finds that the dielectric tensor appropriate for the description of a classical hot plasma may be written

$$D_{ij}^{(c)}(\omega, \mathbf{k}) = \delta_{ij} + \sum_{l=-\infty}^{\infty}\left[\lim_{\hbar\to 0} D_{ij}^{(l)}(\omega, \mathbf{k})\right]$$

$$= \delta_{ij} + \sum_{l=-\infty}^{\infty}\left\{\lim_{\gamma\to+0}\frac{\omega_p^2}{\omega^2}\int_{-\infty}^{\infty} dv_z \right.$$

$$\left. \times \int_0^\infty 2\pi v\, dv\left(\frac{S_{ij}^{(l)}}{\omega - l\omega_b - k_z v_z + i\gamma}\right)\right\}, \quad (30)$$

where

$$S_{ij}^{(l)} = \left(\lim_{\hbar\to 0}\Pi_{ij}^{(l)}\right)\left[\frac{l\omega_b}{v}\frac{\partial}{\partial v} + k_z\frac{\partial}{\partial v_z}\right] f(v, v_z). \quad (31)$$

By making use of Eqs. (3a), (3b) or (11), (9), and (10) in Eq. (19), one can readily obtain the various elements of the tensor $\mathbf{\Pi}^{(l)}$ as a function of $(n + \frac{1}{2})\hbar\omega_b$ and $(v_z + \hbar k_z/2\mu)$ for any given value of the harmonic number l. Then, taking the classical limit, one obtains $\lim_{\hbar\to 0}\Pi_{ij}^{(l)}$ as a function of v and v_z for all values of $i, j = x, y$, and z. Our results of Eqs. (30) and (31) are equivalent to the results of Eqs. (7.13) to (7.17) found in the book by Bekefi.[2] It should be noted that we have written the tensor $\mathsf{S}^{(l)}$ in a Cartesian coordinate system which manifestly exhibits the cylindrical symmetry of the system under consideration, while Bekefi in his Eq. (7.15) has written the corresponding tensor S in a Cartesian coordinate system in which the propagation vector $\mathbf{k}$ of the radiation field is assumed to lie in the $x - z$ plane (that is, $k_\perp = k_x$ and $k_y = 0$). Thus the tensor $\mathsf{S}^{(l)}$ of Eq. (31) is related to the corresponding tensor S of Eq. (7.15) found in the book by Bekefi through a similarity transformation. Furthermore, since the frequency ω must satisfy the classical resonance condition $\omega - l\omega_b - k_z v_z = 0$, it is relatively easy to show that, in the nonrelativistic limit, the functions U and W as given by Eq. (7.17) of the book by Bekefi are such that

$$\frac{W}{v_z} = \frac{U}{v} = \left[\frac{l\omega_b}{v}\frac{\partial}{\partial v} + k_z\frac{\partial}{\partial v_z}\right] f(v, v_z). \quad (32)$$

Finally, in the Appendix we show that the classical dielectric tensor $\mathsf{D}^{(c)}(\omega, \mathbf{k})$ of Eq. (30) is in complete agreement with the familiar results of the conventional hot plasma theory.[2–4]

V. THE FIRST QUANTUM CORRECTION TO THE CLASSICAL HOT PLASMA THEORY

It is our aim now to evaluate the dielectric tensor $D_{ij}(\omega, \mathbf{k})$ to the first order in the Planck constant $\hbar$. On making use of Eqs. (23), (26), (30), and (A21), one finds that, to the first order in $\hbar$, the dielectric tensor may be written

$$D_{ij}(\omega, \mathbf{k}) = D_{ij}^{(c)}(\omega, \mathbf{k}) + \Delta D_{ij}(\omega, \mathbf{k}), \quad (33)$$

where the first quantum correction $\Delta D_{ij}(\omega, \mathbf{k})$ may be written

$$\Delta D_{ij}(\omega, \mathbf{k}) = \sum_{l=-\infty}^{\infty}\left[\lim_{\gamma\to+0}\frac{\omega_p^2}{\omega^2}\int_{-\infty}^{\infty} dv_z \int_0^\infty 2\pi v\, dv\Theta_{ij}^{(l)}\right], \quad (34)$$

where

$$\Theta_{ij}^{(l)} = \left\{\frac{\Delta S_{ij}^{(l)}}{\omega - l\omega_b - k_z v_z + i\gamma}\right\} + \left(\frac{\hbar}{2\mu}\right)\left\{\frac{k_z^2 S_{ij}^{(l)}}{(\omega - l\omega_b - k_z v_z + i\gamma)^2} + \left[\left(\frac{S_{ij}^{(l)}}{Q_l}\right)\left(\frac{l\omega_b}{v}\frac{\partial}{\partial v} + k_z\frac{\partial}{\partial v_z}\right)\left(\frac{l\omega_b}{v}\frac{\partial}{\partial v} + k_z\frac{\partial}{\partial v_z}\right) f(v, v_z)\right] \Big/ \omega - l\omega_b - k_z v_z + i\gamma\right\}, \quad (35)$$

where Q_l is given by Eq. (A11), $S_{ij}^{(l)}$ is given by Eq. (31), and the first quantum correction $\Delta S_{ij}^{(l)}$ is given by

$$\Delta S_{ij}^{(l)} = \hbar Q_l \lim_{\hbar\to 0}\left[\left(\Pi_{ij}^{(l)} - \lim_{\hbar\to 0}\Pi_{ij}^{(l)}\right)\Big/\hbar\right]. \quad (36)$$

Let us now choose a Cartesian coordinate system in which the x axis lies in the plane of the vectors $\mathbf{k}$ and $\mathbf{B}$ (that is, $k_\perp = k_x$ and $k_y = 0$). In such a coordinate system, the elements of the tensor $\mathsf{S}^{(l)}$ are given by Eq. (A10), and from Eqs. (9), (19), (36), (A18), (A19), and (A20) we obtain the elements of the tensor $\Delta\mathsf{S}^{(l)}$ as

$$\Delta S_{xx}^{(l)} = v^2 Q_l\left(\frac{\hbar\omega_b}{\mu v^2}\right)\left(\frac{lJ_l}{\lambda}\right) \times \left\{J_l' + (\operatorname{sgn} l)\left[(l^2+1)\left(\frac{lJ_l}{\lambda}\right) - 2lJ_l'\right]\right\},$$

$$\Delta S_{yy}^{(l)} = v^2 Q_l\left(\frac{\hbar\omega_b}{\mu v^2}\right) \times J_l'\left\{\left(\frac{lJ_l}{\lambda}\right) + (\operatorname{sgn} l)\left[(l^2+1)J_l' - 2l\left(\frac{lJ_l}{\lambda}\right)\right]\right\},$$

$$\Delta S_{zz}^{(l)} = v_z^2 Q_l J_l^2\{(\hbar k_z/\mu v_z) + (\operatorname{sgn} l)(l^2\hbar\omega_b/\mu v^2)\},$$

$$\Delta S_{xy}^{(l)} = -\Delta S_{yx}^{(l)} = iv^2 Q_l\left(\frac{\hbar\omega_b}{2\mu v^2}\right) \times \left\{\left(\frac{lJ_l}{\lambda}\right)^2 + (J_l')^2 + (\operatorname{sgn} l) \times \left[2(l^2+1)\left(\frac{lJ_l}{\lambda}\right)J_l' - 2l\left(\frac{lJ_l}{\lambda}\right)^2 - 2l(J_l')^2\right]\right\},$$

$$\Delta S_{xz}^{(l)} = \Delta S_{zx}^{(l)} = v_z v Q_l \times \left\{\left(\frac{\hbar k_z}{2\mu v_z}\right)\left(\frac{lJ_l^2}{\lambda}\right) + \left(\frac{\hbar\omega_b}{2\mu v^2}\right)\left(J_l J_l' + (\operatorname{sgn} l) \times \left[(2l^2+1)\left(\frac{lJ_l^2}{\lambda}\right) - 2lJ_l J_l'\right]\right)\right\},$$

$$\Delta S_{yz}^{(l)} = -\Delta S_{zy}^{(l)} = -iv_z v Q_l \times \left\{\left(\frac{\hbar k_z}{2\mu v_z}\right)J_l J_l' + \left(\frac{\hbar\omega_b}{2\mu v^2}\right)\left(\left[\frac{lJ_l^2}{\lambda}\right] + (\operatorname{sgn} l) \times \left[(2l^2+1)J_l J_l' - 2l\left(\frac{lJ_l^2}{\lambda}\right)\right]\right)\right\}, \quad (37)$$

where sgn l is $+1$ for positive values of l and -1 for negative values of l, and λ is given by Eq. (A4).

It is physically instructive to examine the order of magnitude of the ratio $[\Delta D_{ij}(\omega, \mathbf{k})/D_{ij}^{(c)}(\omega, \mathbf{k})]$. Since the leading contribution to $\Delta D_{ij}(\omega, \mathbf{k})$ comes from the $\Delta S_{ij}^{(l)}$ term of Eq. (35), we may, for example, examine the ratio

$$\left(\frac{\Delta D_{xx}}{D_{xx}^{(c)}}\right) \approx \left(\frac{\Delta S_{xx}^{(l)}}{S_{xx}^{(l)}}\right) \approx \left\langle\frac{\hbar\omega_b}{\mu v^2}\right\rangle \approx \left(\frac{\hbar\omega_b}{2\kappa T}\right), \quad (38)$$

where we have made use of Eqs. (37) and (A10), the angular brackets refer to a statistical average over the electron states, T is the kinetic temperature of the electrons, and κ is the Boltzmann constant. For a gaseous plasma of electrons and ions whose electron temperature $T \approx 1$ eV in an external magnetic field of 5×10^4 gauss, $(\hbar\omega_b/2\kappa T) \approx 3 \times 10^{-4}$. Thus one can ignore the quantum corrections to the classical dielectric tensor for most laboratory gaseous plasma studies. However, if one is interested in studying the electrodynamic behavior of a "classical plasma" formed by electrons and holes in semiconductors at not too high carrier densities and not too low temperatures T (that is, under conditions where the electrons and holes obey classical statistics), in the presence of an external magnetic field B, then, for some values of T and B, one has to use the first quantum correction to the classical dielectric tensor. For example, for InSb (for which the effective electron mass is about 0.015 times the free-electron mass) at liquid nitrogen temperatures (that is, $T \approx 77°$K) in an external magnetic field of 10^4 gauss, $(\hbar\omega_b/2\kappa T) \approx 0.6$. Thus, although the electron–hole plasma is "classical" in the sense of statistical mechanics, the quantum corrections (to the classical dielectric tensor) arising from the "Landau quantization" of the electron orbits in a uniform magnetic field are significant. The results of this section will therefore prove useful for the analysis of the experimental studies of the electrodynamic behavior of such "solid-state classical plasmas" in external magnetic fields of moderate strengths.

VI. SOME REMARKS

In what has been presented so far we have assumed that the system of electrons in the box of volume L^3

under consideration is nondegenerate. If the system of electrons is degenerate, then one has to take account of the Pauli exclusion principle,[10] and in applying the principle of detailed balance we must bear in mind that a transition can only take place to an electron state which is vacant. Thus, we can use the above theory for a degenerate system of electrons, provided we make the following modification: We must replace $F(E_{n+l,v_z'})$ and $F(E_{n,v_z})$ by $F(E_{n+l,v_z'})\{1 - F(E_{n,v_z})\}$ and $F(E_{n,v_z})\{1 - F(E_{n+l,v_z'})\}$, respectively, in Eqs. (13), (15), and (16). Since

$$[F(E_{n,v_z}) - F(E_{n+l,v_z+\hbar k_z/\mu})] = [F(E_{n,v_z})\{1 - F(E_{n+l,v_z+\hbar k_z/\mu})\} - F(E_{n+l,v_z+\hbar k_z/\mu})\{1 - F(E_{n,v_z})\}],$$

Eqs. (18) and (26) remain valid for both the degenerate and the nondegenerate system of electrons.

Finally, it may be noted that the method presented here is somewhat analogous to the Kramers–Heisenberg[5,6] quantum theory of gaseous dispersion. The basic philosophy behind the Kramers–Heisenberg dispersion theory is that, starting from a knowledge of the fundamental quantum-mechanical transition probabilities for emission and absorption of photons (that is, from a knowledge of the Einstein A and B coefficients), one can obtain the equation of energy transfer or the energy balance equation by a straightforward application of the principle of detailed balance. From the solution of this energy-balance equation one can obtain information about the absorption of electromagnetic energy by the system under consideration. The dissipation or the absorption of energy by any system is, in general, represented by the anti-Hermitian part of the dielectric tensor appropriate for the description of the system. Since the real and imaginary parts of any dielectric tensor must satisfy the well-known Kramers–Kronig relations as a consequence of the causality principle, one can thus obtain the entire dielectric tensor from a knowledge of its anti-Hermitian part. Thus, in conclusion, it is interesting to find that the original ideas of Kramers and Heisenberg suffice to examine the dispersion properties of a hot plasma which is customarily done by a kinetic description based on the Vlasov equation.[1–4]

ACKNOWLEDGMENTS

It is a pleasure to express my thanks to Dr. R. Ellis, Dr. J. Dow, Professor T. H. Stix, Dr. W. Hooke, Dr. P. Rutherford, and Professor E. Frieman for useful discussions.

This work was performed under the auspices of the U.S. Atomic Energy Commission, Contract No. AT(30-1)-1238.

APPENDIX

From Eqs. (10) and (11) we get

$$\begin{aligned}(e^{\pm i\mathbf{k}\cdot\mathbf{r}})_{m,n} = \Bigg(&\delta_{m,n} \pm \left(\frac{\hbar}{\mu\omega_b}\right)^{\frac{1}{2}}[k_+(n+1)^{\frac{1}{2}}\delta_{m,n+1} - k_-\sqrt{n}\,\delta_{m,n-1}] \\ &+ \frac{1}{2!}\left(\frac{\hbar}{\mu\omega_b}\right)\{k_+^2[(n+2)(n+1)]^{\frac{1}{2}}\delta_{m,n+2} - k_+k_-2(n+\tfrac{1}{2})\delta_{m,n} \\ &+ k_-^2[n(n-1)]^{\frac{1}{2}}\delta_{m,n-2}\} \pm \frac{1}{3!}\left(\frac{\hbar}{\mu\omega_b}\right)^{\frac{3}{2}}\{k_+^3[(n+3)(n+2)(n+1)]^{\frac{1}{2}}\delta_{m,n+3} \\ &- k_+^2k_-3(n+1)[n+1]^{\frac{1}{2}}\delta_{m,n+1} + k_+k_-^2 3n\sqrt{n}\,\delta_{m,n-1} - k_-^3[n(n-1)(n-2)]^{\frac{1}{2}}\delta_{m,n-3}\} \\ &+ \frac{1}{4!}\left(\frac{\hbar}{\mu\omega_b}\right)^2\{k_+^4[(n+4)(n+3)(n+2)(n+1)]^{\frac{1}{2}}\delta_{m,n+4} - k_+^3k_-(4n+6)[(n+2)(n+1)]^{\frac{1}{2}}\delta_{m,n+2} \\ &+ k_+^2k_-^2[3n^2 + 3(n+1)^2]\delta_{m,n} - k_+k_-^3(4n-2)[n(n-1)]^{\frac{1}{2}}\delta_{m,n-2} \\ &+ k_-^4[n(n-1)(n-2)(n-3)]^{\frac{1}{2}}\delta_{m,n-4}\} \pm \frac{1}{5!}\left(\frac{\hbar}{\mu\omega_b}\right)^{\frac{5}{2}}\{k_+^5[(n+5)(n+4)(n+3)(n+2)(n+1)]^{\frac{1}{2}} \\ &\times \delta_{m,n+5} - k_+^4k_-5(n+2)[(n+3)(n+2)(n+1)]^{\frac{1}{2}}\delta_{m,n+3} + k_+^3k_-^2 5(2n^2+4n+3)[n+1]^{\frac{1}{2}} \\ &\times \delta_{m,n+1} - k_+^2k_-^3 5(2n^2+1)\sqrt{n}\,\delta_{m,n-1} + k_+k_-^4 5(n-1)[n(n-1)(n-2)]^{\frac{1}{2}}\delta_{m,n-3} \\ &- k_-^5[n(n-1)(n-2)(n-3)(n-4)]^{\frac{1}{2}}\delta_{m,n-5}\} + \cdots\Bigg),\end{aligned} \tag{A1}$$

where we have evaluated the first six terms of Eq. (10). Grouping together the terms that belong to the same Kronecker δ's, the classical limit of Eq. (A1) may be written

$$\begin{aligned}\lim_{\hbar\to 0}(e^{\pm i\mathbf{k}\cdot\mathbf{r}})_{m,n} = \Big\{&\delta_{m,n}\Big[1-\Big(\frac{k_+k_-v^2}{2\omega_b^2}\Big)+\frac{1}{2!\,2!}\Big(\frac{k_+k_-v^2}{2\omega_b^2}\Big)^2+\cdots\Big]\\ &\pm\delta_{m,n+1}\Big[\Big(\frac{k_+v}{\sqrt{2}\,\omega_b}\Big)-\frac{1}{2!}\Big(\frac{k_+^2k_-v^3}{2\sqrt{2}\,\omega_b^3}\Big)+\frac{1}{2!\,3!}\Big(\frac{k_+^3k_-^2v^5}{4\sqrt{2}\,\omega_b^5}\Big)+\cdots\Big]\\ &\mp\delta_{m,n-1}\Big[\Big(\frac{k_-v}{\sqrt{2}\,\omega_b}\Big)-\frac{1}{2!}\Big(\frac{k_+k_-^2v^3}{2\sqrt{2}\,\omega_b^3}\Big)+\frac{1}{2!\,3!}\Big(\frac{k_+^2k_-^3v^5}{4\sqrt{2}\,\omega_b^5}\Big)+\cdots\Big]\\ &+\delta_{m,n+2}\Big[\frac{1}{2!}\Big(\frac{k_+^2v^2}{2\omega_b^2}\Big)-\frac{1}{3!}\Big(\frac{k_+^3k_-v^4}{4\omega_b^4}\Big)+\cdots\Big]\\ &+\delta_{m,n-2}\Big[\frac{1}{2!}\Big(\frac{k_-^2v^2}{2\omega_b^2}\Big)-\frac{1}{3!}\Big(\frac{k_+k_-^3v^4}{4\omega_b^2}\Big)+\cdots\Big]\\ &\pm\delta_{m,n+3}\Big[\frac{1}{3!}\Big(\frac{k_+^3v^3}{2\sqrt{2}\,\omega_b^3}\Big)-\frac{1}{4!}\Big(\frac{k_+^4k_-v^5}{4\sqrt{2}\,\omega_b^5}\Big)+\cdots\Big]\\ &\mp\delta_{m,n-3}\Big[\frac{1}{3!}\Big(\frac{k_-^3v^3}{2\sqrt{2}\,\omega_b^3}\Big)-\frac{1}{4!}\Big(\frac{k_+k_-^4v^5}{4\sqrt{2}\,\omega_b^5}\Big)+\cdots\Big]\\ &+\delta_{m,n+4}\Big[\frac{1}{4!}\Big(\frac{k_+^4v^4}{4\omega_b^4}\Big)+\cdots\Big]+\delta_{m,n-4}\Big[\frac{1}{4!}\Big(\frac{k_-^4v^4}{4\omega_b^4}\Big)+\cdots\Big]\\ &\pm\delta_{m,n+5}\Big[\frac{1}{5!}\Big(\frac{k_+^5v^5}{4\sqrt{2}\,\omega_b^5}\Big)+\cdots\Big]\mp\delta_{m,n-5}\Big[\frac{1}{5!}\Big(\frac{k_-^5v^5}{4\sqrt{2}\,\omega_b^5}\Big)+\cdots\Big]+\cdots\Big\}. \qquad \text{(A2)}\end{aligned}$$

In order to compare our results with those found in Sec. 7.2 of the book by Bekefi,[2] let us now choose the coordinate system such that the propagation vector $\mathbf{k}$ lies in the $x-z$ plane of the Cartesian frame (that is, $k_\perp = k_x$ and $k_y = 0$). Then

$$k_+ = k_- = (k_\perp/\sqrt{2}). \qquad \text{(A3)}$$

Let

$$\lambda = (k_\perp v/\omega_b). \qquad \text{(A4)}$$

Then it is apparent that one can show that the complete form of Eq. (A2) may be written

$$\lim_{\hbar\to 0}(e^{\pm i\mathbf{k}\cdot\mathbf{r}})_{m,n} = \sum_{l=-\infty}^{\infty}\delta_{m,n+l}(\pm 1)^{|l|}J_l(\lambda), \qquad \text{(A5)}$$

where $J_l(\lambda)$ is the Bessel function[14] of order l. Let

$$v_\pm = (2)^{-\frac{1}{2}}(v_x \pm iv_y). \qquad \text{(A6)}$$

From Eq. (3a) it is relatively easy to show that

$$\lim_{\hbar\to 0}(v_\pm)_{m,n} = \delta_{m,n\mp 1}(v(2)^{-\frac{1}{2}}). \qquad \text{(A7)}$$

From Eqs. (A5) and (A7) we get

$$\lim_{\hbar\to 0}(e^{\pm i\mathbf{k}\cdot\mathbf{r}})_{n+l,n} = (\pm 1)^{|l|}J_l(\lambda) \qquad \text{(A8)}$$

and

$$\lim_{\hbar\to 0}[\tfrac{1}{2}(e^{\pm i\mathbf{k}\cdot\mathbf{r}}v_\pm + v_\pm e^{\pm i\mathbf{k}\cdot\mathbf{r}})]_{n+l,n} = (\pm 1)^{|l\pm 1|}J_{l\pm 1}(\lambda)\Big(\frac{v}{(2)^{\frac{1}{2}}}\Big). \qquad \text{(A9)}$$

Here the plus and minus signs in (± 1) correspond to those in $\exp(\pm i\mathbf{k}\cdot\mathbf{r})$ and $l \pm 1$ correspond to $v_\pm$. From Eqs. (9), (19), (31), (A6), (A8), and (A9) we get

$$\mathsf{S}^{(l)} = \begin{vmatrix} v^2Q_l\Big(\dfrac{lJ_l}{\lambda}\Big)^2 & iv^2Q_l\dfrac{lJ_lJ_l'}{\lambda} & v_zvQ_l\dfrac{lJ_l^2}{\lambda} \\ -iv^2Q_l\dfrac{lJ_lJ_l'}{\lambda} & v^2Q_l(J_l')^2 & -iv_zvQ_lJ_lJ_l' \\ v_zvQ_l\dfrac{lJ_l^2}{\lambda} & iv_zvQ_lJ_lJ_l' & v_z^2Q_lJ_l^2 \end{vmatrix}, \qquad \text{(A10)}$$

where

$$Q_l = \Big[\frac{l\omega_b}{v}\frac{\partial}{\partial v}+k_z\frac{\partial}{\partial v_z}\Big]f(v, v_z), \qquad \text{(A11)}$$

and we have used the relations[14]

$$\tfrac{1}{2}[J_{l-1}(\lambda)+J_{l+1}(\lambda)] = (l/\lambda)J_l(\lambda) \qquad \text{(A12)}$$

and

$$\tfrac{1}{2}[J_{l-1}(\lambda)-J_{l+1}(\lambda)] = J_l'(\lambda). \qquad \text{(A13)}$$

[14] I. N. Sneddon, *Special Functions of Mathematical Physics and Chemistry* (Oliver and Boyd Ltd., London, England, 1961).

Here the prime denotes differentiation with respect to λ. We wish to emphasize that the tensor $\mathsf{S}^{(l)}$ of Eq. (A10) is written in a Cartesian coordinate system in which the z axis is directed along the external magnetic field **B** and the x axis lies in the plane of the vectors **k** and **B**. Thus one sees readily that our results of the dielectric tensor agree in the classical limit with the corresponding results found in Sec. 7.2 of the book by Bekefi.[2]

We now wish to show how one obtains the first quantum corrections to the results of the classical hot plasma theory. Since $(n+\frac{1}{2})\hbar\omega_b \to \mu v^2/2$, it is relatively easy to show that, for $s \geq 0$,

$$\left(\frac{\hbar}{\mu\omega_b}\right)^{s/2}[(n+s)(n+s-1)\cdots(n+1)]^{\frac{1}{2}} \to \frac{1}{2^{s/2}}\left(\frac{v}{\omega_b}\right)^s\left[1+\frac{1}{2}\frac{s^2\hbar\omega_b}{\mu v^2}\right] \quad \text{(A14)}$$

and

$$\left(\frac{\hbar}{\mu\omega_b}\right)^{s/2}[n(n-1)\cdots(n-s+1)]^{\frac{1}{2}} \to \frac{1}{2^{s/2}}\left(\frac{v}{\omega_b}\right)^s\left[1-\frac{1}{2}\frac{s^2\hbar\omega_b}{\mu v^2}\right] \quad \text{(A15)}$$

to the first order in $\hbar$. On making use of Eqs. (A14) and (A15), one can easily show that

$$(e^{\pm i\mathbf{k}\cdot\mathbf{r}})_{m,n} = \sum_{l=-\infty}^{\infty}\delta_{m,n+l}(\pm 1)^{|l|} \times J_l(\lambda)[1+(\operatorname{sgn} l)(l^2\hbar\omega_b/2\mu v^2)] \quad \text{(A16)}$$

and

$$(v_\pm)_{m,n} = \delta_{m,n\mp 1}(v(2)^{-\frac{1}{2}})[1 \mp (\hbar\omega_b/2\mu v^2)] \quad \text{(A17)}$$

to the first order in $\hbar$. Here $\operatorname{sgn} l$ is $+1$ for positive values of l and -1 for negative values of l. From Eqs. (A6), (A12), (A13), (A16), and (A17) we get

$$(v_z + \hbar k_z/2\mu)(e^{\pm i\mathbf{k}\cdot\mathbf{r}})_{n+l,n} = (\pm 1)^{|l|}v_z J_l(\lambda)[1+(\hbar k_z/2\mu v_z)+(\operatorname{sgn} l)(l^2\hbar\omega_b/2\mu v^2)], \quad \text{(A18)}$$

$$[\tfrac{1}{2}(e^{\pm i\mathbf{k}\cdot\mathbf{r}}v_x + v_x e^{\pm i\mathbf{k}\cdot\mathbf{r}})]_{n+l,n} = (\pm 1)^{|l+1|}(v)\left(\left[\frac{lJ_l(\lambda)}{\lambda}\right] + \left(\frac{\hbar\omega_b}{2\mu v^2}\right)\left\{J_l'(\lambda) + (\operatorname{sgn} l)\left[\frac{(l^2+1)lJ_l(\lambda)}{\lambda} - 2lJ_l'(\lambda)\right]\right\}\right), \quad \text{(A19)}$$

and

$$[\tfrac{1}{2}(e^{\pm i\mathbf{k}\cdot\mathbf{r}}v_y + v_y e^{\pm i\mathbf{k}\cdot\mathbf{r}})]_{n+l,n} = (\pm 1)^{|l+1|}(iv)\left(J_l'(\lambda) + \left(\frac{\hbar\omega_b}{2\mu v^2}\right)\left\{\left[\frac{lJ_l(\lambda)}{\lambda}\right] + (\operatorname{sgn} l)\left[(l^2+1)J_l'(\lambda) - \frac{2llJ_l(\lambda)}{\lambda}\right]\right\}\right) \quad \text{(A20)}$$

to the first order in $\hbar$. Furthermore, one can easily show that

$$\frac{(\mu/\hbar)[F(E_{n+l,v_z+\hbar k_z/\mu}) - F(E_{n,v_z})]}{\omega - l\omega_b - k_z(v_z+\hbar k_z/2\mu)+i\gamma} \to \left[\frac{\left(\frac{l\omega_b}{v}\frac{\partial}{\partial v}+k_z\frac{\partial}{\partial v_z}\right)f(v,v_z)}{\omega - l\omega_b - k_z v_z + i\gamma} + \frac{\hbar}{2\mu}\left\{\frac{\left(\frac{l\omega_b}{v}\frac{\partial}{\partial v}+k_z\frac{\partial}{\partial v_z}\right)\left(\frac{l\omega_b}{v}\frac{\partial}{\partial v}+k_z\frac{\partial}{\partial v_z}\right)f(v,v_z)}{\omega - l\omega_b - k_z v_z + i\gamma} + \frac{k_z^2\left(\frac{l\omega_b}{v}\frac{\partial}{\partial v}+k_z\frac{\partial}{\partial v_z}\right)f(v,v_z)}{(\omega - l\omega_b - k_z v_z + i\gamma)^2}\right\}\right] \quad \text{(A21)}$$

to the first order in $\hbar$. In this way one can easily obtain the first quantum corrections to the results of the classical hot plasma theory as given by Eqs. (33)–(37).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Singularities of Conformal-Invariant Scattering Amplitudes*

DARRYL D. COON
Department of Physics, University of Washington, Seattle, Washington

(Received 22 November 1968)

It is shown that conformal invariance alone, with no specific dynamics, provides severe constraints on the location of singularities of the off-mass-shell amplitude for elastic scattering of spinless particles. Kinematical and dynamical singularities are correlated in an interesting fashion with the singular and general solutions of differential equations which follow from the momentum-space requirements of conformal invariance. Explicit expressions are found for the location of singularities which are independent of t, the square of the momentum transfer. The possibility of having an asymptotic behavior of t^α for large t together with singularities specified by the function α is discussed.

I. INTRODUCTION

The conformal group[1–3] is a higher space–time symmetry group in which the Poincaré transformations form a subgroup. Conformal transformations map light cones into light cones and conserve angles defined locally. Maxwell's equations[1] and various massless field theories[2] including electrodynamics and $\lambda\phi^4$ theory are conformal invariant.

Exact conformal invariance implies vanishing masses or continuous mass spectra.[2] However, a symmetry may be useful, as is SU_3, even though its unbroken form has unrealistic mass constraints. Mack and Salam[4] discussed a scheme for broken conformal symmetry in Lagrangian field theory. In S-matrix theory, constraints on mass spectra manifest themselves as restrictions on the singularity structure of scattering amplitudes. We will show that the interesting singularity structure of conformal-invariant amplitudes is off the mass shell with singularities converging toward zero energy as the mass shell is approached. The way in which singularities are located might suggest reasonable symmetry-breaking schemes within the framework of S-matrix theory. We make use of some results of Bali, Coon, and Katz,[5] who recently investigated the restrictions imposed by conformal invariance on the off-mass-shell amplitude for scattering of massless spinless particles.

It has been conjectured that the conformal group may provide an approximate description of some processes[6] such as scattering at high energy and high momentum transfer. Integral representations[5] for conformal-invariant elastic scattering amplitudes involve arbitrary functions of only two variables rather than the six implied by invariance under the Poincaré subgroup alone. One might think that some of this information could remain relevant, at least in an asymptotic region, when the symmetry is broken.

We begin, in Sec. II, with the momentum-space requirements of conformal invariance given in Ref. 5. From these requirements, three second-order partial differential equations are derived with six Lorentz scalars as independent variables. The off-mass-shell elastic scattering amplitude must be a dilatation-invariant solution of these equations. Singularities of solutions of such partial differential equations lie along characteristic surfaces which are investigated in Sec. III. Attention is focused on singularities whose positions are independent of t, the square of the momentum transfer. In Sec. IV, it is shown that, for a conformal-invariant amplitude to have an asymptotic behavior of t^α for large t, then α must satisfy the differential equations which determine characteristic surfaces.

II. CONFORMAL INVARIANCE AND SCATTERING AMPLITUDES

The conformal group[3] is made up of Poincaré transformations, the inverse radius transformation

$$x_\mu \to x_\mu/x^2, \tag{1}$$

and dilatations

$$x_\mu \to bx_\mu, \tag{2}$$

where b is an arbitrary parameter. If an inverse radius transformation, a translation, and another inverse radius transformation are performed successively, the

* Work supported by the U.S. Atomic Energy Commission AT(45-1)-1388.

[1] E. Cunningham, Proc. London Math. Soc. **8**, 77 (1910); H. Bateman, *ibid.* **8**, 223 (1910).

[2] J. Wess, Nuovo Cimento **18**, 1086 (1960).

[3] Discussion of the conformal group and many references to the other articles can be found in: F. Gürsey, Nuovo Cimento **3**, 988 (1956); T. Fulton, F. Rohrlich, and L. Witten, Rev. Mod. Phys. **34**, 442 (1962); H. A. Kastrup, Ann. Physik **7**, 388 (1962).

[4] G. Mack and A. Salam, ICTP Preprint IC/68/68, Trieste, 1968. This paper includes references to more recent work.

[5] N. F. Bali, D. D. Coon, and A. Katz, University of Washington Preprint No. 539, Seattle, Washington, 1968; J. Math. Phys. (to be published).

[6] H. A. Kastrup, Nucl. Phys. **58**, 561 (1964); Phys. Rev. **142**, 1060 (1965); **143**, 1041 (1966); **150**, 1183 (1966).

resulting transformation has the form

$$x_\mu \to \frac{x_\mu + a_\mu x^2}{1 + 2a \cdot x + a^2 x^2}. \tag{3}$$

The corresponding infinitesimal generator is

$$K_\mu \equiv -i(x^2 \partial_\mu - 2x_\mu x \cdot \partial). \tag{4}$$

We now consider elastic scattering of spinless massless particles with momenta p_i, where $i = 1, 2, 3, 4$. By virtue of the Poincaré subgroup, which implies

$$p_1 + p_2 + p_3 + p_4 = 0, \tag{5}$$

the usual off-mass-shell amplitude can be expressed as a function of the six Lorentz scalars

$$s \equiv (p_1 + p_2)^2, \quad t \equiv (p_1 + p_3)^2, \tag{6}$$

and

$$y_i \equiv p_i^2, \quad i = 1, 2, 3, 4. \tag{7}$$

It has recently been shown[5] that, for the S matrix to be conformal-invariant, the off-mass-shell elastic scattering amplitude M must be invariant under dilatations

$$p_i^\mu \to b^{-1} p_i^\mu \tag{8}$$

and must satisfy the equation

$$\sum_{i=1}^{3} L_i^\mu M = 0, \tag{9}$$

where

$$L_i^\mu = \Box_i p_i^\mu - 2\partial_i^\mu \partial_i \cdot p_i + 6\partial_i^\mu \tag{10a}$$

$$= p_i^\mu \Box_i - 2p_i \cdot \partial_i \partial_i^\mu - 2\partial_i^\mu. \tag{10b}$$

Here ∂_i^μ and $\Box_i$ indicate differentiation with respect to p_i. Except for the last term in Eq. (10a), the differential operators L_i^μ are just the momentum-space operators corresponding to the coordinate-space operator K^μ. The last term in Eq. (10a) is present because $\delta(p_1 + p_2 + p_3 + p_4)M$ is the Fourier transform of a conformal scalar density rather than a conformal scalar. The only dependence on the weight[5] of this scalar density is in the coefficient of the first derivative terms in Eqs. (10).

Feynman diagrams of massless $\lambda\phi^4$ theory provide examples of formal expressions which satisfy the above requirements. This is most easily seen in coordinate space where divergences are avoided.

From Eq. (9) we obtain the following three partial differential equations for M in terms of Lorentz scalars:

$$y_1 \frac{\partial^2 M}{\partial y_1^2} - y_4 \frac{\partial^2 M}{\partial s \partial t} + R + S + T = 0, \tag{11}$$

$$y_2 \frac{\partial^2 M}{\partial y_2^2} - y_3 \frac{\partial^2 M}{\partial s \partial t} + R + S = 0, \tag{12}$$

$$y_3 \frac{\partial^2 M}{\partial y_3^2} - y_2 \frac{\partial^2 M}{\partial s \partial t} + R + T = 0, \tag{13}$$

where

$$R \equiv 2\frac{\partial}{\partial y_4}\left[s\frac{\partial M}{\partial s} + t\frac{\partial M}{\partial t} + \sum_{i=1}^{4} y_i \frac{\partial M}{\partial y_i}\right] + y_1 \frac{\partial^2 M}{\partial s \partial t} - y_4 \frac{\partial^2 M}{\partial y_4^2}, \tag{14}$$

$$S \equiv \frac{\partial}{\partial s}\left[s\frac{\partial M}{\partial s} + t\frac{\partial M}{\partial t} + 2y_1 \frac{\partial M}{\partial y_1} + 2y_2 \frac{\partial M}{\partial y_2}\right], \tag{15}$$

$$T \equiv \frac{\partial}{\partial t}\left[s\frac{\partial M}{\partial s} + t\frac{\partial M}{\partial t} + 2y_1 \frac{\partial M}{\partial y_1} + 2y_3 \frac{\partial M}{\partial y_3}\right]. \tag{16}$$

Making use of the differential statement of dilatation invariance

$$s\frac{\partial M}{\partial s} + t\frac{\partial M}{\partial t} + \sum_{i=1}^{4} y_i \frac{\partial M}{\partial y_i} = 0, \tag{17}$$

we can write three linear combinations of Eqs. (11)–(13) in the more compact forms

$$y_1 \frac{\partial^2 M}{\partial y_1^2} - y_2 \frac{\partial^2 M}{\partial y_2^2} + (y_3 - y_4)\frac{\partial^2 M}{\partial s \partial t} + \left[y_1 \frac{\partial}{\partial y_1} - y_2 \frac{\partial}{\partial y_2} + y_3 \frac{\partial}{\partial y_3} - y_4 \frac{\partial}{\partial y_4}\right]\frac{\partial M}{\partial t} = 0, \tag{18}$$

$$y_3 \frac{\partial^2 M}{\partial y_3^2} - y_4 \frac{\partial^2 M}{\partial y_4^2} + (y_1 - y_2)\frac{\partial^2 M}{\partial s \partial t} + \left[y_1 \frac{\partial}{\partial y_1} - y_2 \frac{\partial}{\partial y_2} + y_3 \frac{\partial}{\partial y_3} - y_4 \frac{\partial}{\partial y_4}\right]\frac{\partial M}{\partial t} = 0, \tag{19}$$

$$y_1 \frac{\partial^2 M}{\partial y_1^2} - y_3 \frac{\partial^2 M}{\partial y_3^2} + (y_2 - y_4)\frac{\partial^2 M}{\partial s \partial t} + \left[y_1 \frac{\partial}{\partial y_1} + y_2 \frac{\partial}{\partial y_2} - y_3 \frac{\partial}{\partial y_3} - y_4 \frac{\partial}{\partial y_4}\right]\frac{\partial M}{\partial s} = 0. \tag{20}$$

The off-mass-shell amplitude M must be a dilatation-invariant solution of these linear partial differential equations. The coefficients only involve $y_i \equiv p_i^2$. Thus, derivatives of M with respect to s and t must also be homogeneous solutions of Eqs. (18)–(20).

III. CHARACTERISTIC SURFACES

Singularities of solutions of a linear analytic second-order partial differential equation lie along characteristic surfaces.[7] Characteristic surfaces are surfaces along which the partial differential equation represents an interior differential equation. The differential equation does not determine higher normal derivatives

[7] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (English transl.: Interscience Publishers, Inc., New York, 1962), Vol. II, pp. 170–180, 486–488, 552–574, and 635. Our characteristic surfaces are in the six-dimensional s, t, p_i^2 space.

from given initial data (Cauchy data) on such a surface. Characteristic surfaces are found by solving a first-order partial differential equation[8] which is related to the original second-order equation.

We now let

$$\Phi(s, t, y_1, y_2, y_3, y_4) = 0 \tag{21}$$

specify a characteristic surface. The characteristic equations[8] associated with Eqs. (18)–(20) are

$$y_1\Phi_1^2 - y_2\Phi_2^2 + (y_3 - y_4)\Phi_s\Phi_t + [y_1\Phi_1 - y_2\Phi_2 + y_3\Phi_3 - y_4\Phi_4]\Phi_t = 0, \tag{22}$$

$$y_3\Phi_3^2 - y_4\Phi_4^2 + (y_1 - y_2)\Phi_s\Phi_t + [y_1\Phi_1 - y_2\Phi_2 + y_3\Phi_3 - y_4\Phi_4]\Phi_t = 0, \tag{23}$$

$$y_1\Phi_1^2 - y_3\Phi_3^2 + (y_2 - y_4)\Phi_s\Phi_t + [y_1\Phi_1 + y_2\Phi_2 - y_3\Phi_3 - y_4\Phi_4]\Phi_s = 0, \tag{24}$$

where we use the notation

$$\Phi_i \equiv \frac{\partial\Phi}{\partial y_i}, \quad \Phi_s \equiv \frac{\partial\Phi}{\partial s}, \quad \text{etc.} \tag{25}$$

The same function Φ appears in all three equations because we wish to determine the characteristic surfaces which are common to Eqs. (18)–(20). That is, we are interested in combining the constraints of all three equations on the location of singularities.

There are two cases[8] to be considered:

(1) If Φ satisfies Eqs. (22)–(24) identically in s, t, and y_i, then $\Phi = 0$ is a characteristic surface which belongs to a one-parameter family of characteristic surfaces given by $\Phi = \text{const}$.

(2) If Φ satisfies Eqs. (22)–(24) only on the surface $\Phi = 0$, then $\Phi = 0$ is a characteristic surface which is not a member of such a family of characteristic surfaces. In this case, Eqs. (22)–(24) are not interpreted as differential equations since they are not satisfied identically in s, t, and y_i.

If Eq. (21) is solved for s so that

$$s = \Psi(t, y_1, y_2, y_3, y_4), \tag{26}$$

then we have another way of specifying the same surface. Dilatation invariance implies that Φ depends only on ratios of variables and that Ψ is homogeneous of degree one in the variables t and y_i.

Equations (22)–(24) correspond to the following equations for Ψ:

$$y_1\Psi_1^2 - y_2\Psi_2^2 - (y_3 - y_4)\Psi_t + [y_1\Psi_1 - y_2\Psi_2 + y_3\Psi_3 - y_4\Psi_4]\Psi_t = 0, \tag{27}$$

$$y_3\Psi_3^2 - y_4\Psi_4^2 - (y_1 - y_2)\Psi_t + [y_1\Psi_1 - y_2\Psi_2 + y_3\Psi_3 - y_4\Psi_4]\Psi_t = 0, \tag{28}$$

$$y_1\Psi_1^2 - y_3\Psi_3^2 - (y_2 - y_4)\Psi_t - [y_1\Psi_1 + y_2\Psi_2 - y_3\Psi_3 - y_4\Psi_4] = 0, \tag{29}$$

with subscripts denoting derivatives as in (25). These equations are satisfied identically in the variables t and y_i. Thus, both cases considered above now involve solutions of differential equations.[8] Again there are two possibilities:

(1) That Ψ is obtained by finding general solutions of each of the differential equations encountered in combining the homogeneity condition and Eqs. (27)–(29), or

(2) that Ψ involves the singular solutions of at least one differential equation.[9]

In the following, we will solve equations for Φ and Ψ separately in order to illustrate the relation between the above statements concerning Φ and Ψ.

t-Independent Singularities

Since many significant singularities of the amplitude are independent of one kinematic invariant, we now find the t-independent dilatation-invariant solutions of the characteristic equations. (The s- or u-independent solutions can be found from these by interchanging indices of the y_i.) In this case, Eqs. (22), (23), (27), and (28) simplify greatly. Together with dilatation invariance they imply

$$\Phi = \Phi\left[\frac{m_1 + \sigma m_2}{s^{\frac{1}{2}}}, \frac{m_3 + \omega m_4}{s^{\frac{1}{2}}}\right] \tag{30}$$

and

$$\Psi = (m_1 + \sigma m_2)^2 F\left[\frac{m_3 + \omega m_4}{m_1 + \sigma m_2}\right], \tag{31}$$

where

$$m_i \equiv y_i^{\frac{1}{2}} = (p_i^2)^{\frac{1}{2}}. \tag{32}$$

Also, $\sigma = \pm 1$ and $\omega = \pm 1$ independently. The signs associated with m_1 and m_3 are chosen to be positive. Here Φ and F are arbitrary functions of their arguments.

Since Φ now depends on only two variables, Eq. (24) can be reduced to the form of a partial differential equation with two independent variables. The equation can be factored into two linear partial differential forms. To find those Φ which satisfy these equations identically, we simply interpret the equations as partial differential equations. The solutions are

$$\Phi = \Phi\left[\frac{\{s - (m_1 + m_2)^2\}^{\frac{1}{2}} \pm \{s - (m_3 + m_4)^2\}^{\frac{1}{2}}}{m_1 + m_2 + m_3 + m_4}\right], \tag{33}$$

where Φ is an arbitrary function of its argument and

[8] See Ref. 7, pp. 552–558.

[9] When dilatation invariance is combined with a characteristic partial differential equation (for Ψ) which possesses no singular solution, we may obtain a differential equation which does have singular solutions.

each m_i can be replaced by $-m_i$. This replacement is possible because Eqs. (22)–(24) only involve $y_i = m_i^2$. Thus, we have found the one-parameter family of characteristic surfaces given by $\Phi =$ constant.

Dilatation-invariant functions Φ which satisfy the characteristic equations only on the surface $\Phi = 0$ can most easily be found from the differential equations for Ψ. By combining Eq. (31) with the remaining equation (29) for Ψ, we obtain the ordinary differential equation

$$(z^2 - 1)\left[\frac{dF}{dz}\right]^2 - 4z(F - 1)\frac{dF}{dz} + 4F(F - 1) = 0, \tag{34}$$

where

$$z \equiv (m_3 + \omega m_4)/(m_1 + \sigma m_2). \tag{35}$$

This nonlinear equation can be solved by introducing the quantities

$$2A \equiv (z - 1)\frac{dF}{dz} - 2(F - 1) \tag{36}$$

and

$$-2B \equiv (z + 1)\frac{dF}{dz} - 2(F - 1) \tag{37}$$

and writing Eq. (34) as

$$F = AB + 1. \tag{38}$$

Equations (36)–(38) can be combined so that we need to solve only a single linear differential equation for A. Thus, we obtain the general solution

$$F = (1/4c)[(c + 1)^2z^2 + 2(c^2 - 1)z + (c + 1)^2], \tag{39}$$

where c is an arbitrary constant. The singular solutions[10]

$$F = 1 \quad \text{and} \quad F = z^2 \tag{40}$$

can be obtained directly from Eq. (34).

By examining the differential equation (34), we discover that if $F(z)$ is a solution, then $z^2F(1/z)$ is a solution. The symmetry is related to the interchange of incoming and outgoing particles. The solutions (39) and (40) exhibit this symmetry. We also observe that $F(-z)$ must be a solution. This corresponds to the replacement of c by $1/c$ in Eq. (39).

We now have found all the t-independent solutions of the characteristic equations (27)–(29) for Ψ. In terms of the variables $m_i \equiv (p_i^2)^{\frac{1}{2}}$ and the arbitrary constant c, the characteristic surfaces $s = \Psi$ are given by

$$s = \frac{1}{4c}[(c + 1)^2(m_1 + \sigma m_2)^2 + 2(c^2 - 1)(m_1 + \sigma m_2) \times (m_3 + \omega m_4) + (c + 1)^2(m_3 + \omega m_4)^2], \tag{41}$$

$$s = (m_1 \pm m_2)^2, \tag{42}$$

$$s = (m_3 \pm m_4)^2, \tag{43}$$

where $\sigma = \pm 1$ and $\omega = \pm 1$ independently. The family of characteristic surfaces (41) is the same as the family of surfaces $\Phi =$ const associated with Eq. (33). This can be verified by solving Eq. (41) for the constant c. Equations (42) and (43) are obtained from the singular solutions (40) and have no arbitrary constants. They can be put into correspondence with dilatation-invariant Φ's which satisfy all characteristic equations (22)–(24) only for $\Phi = 0$.

Integral representations[5] for dilatation-invariant solutions of equations such as Eqs. (18)–(20) are quite complicated. However, in some special cases the representations reduce to elementary functions:

$$\frac{s - m_1^2 - m_2^2}{\{[s - (m_1 + m_2)^2][s - (m_1 - m_2)^2]\}^{\frac{1}{2}}}, \tag{44}$$

$$m_1^2m_2^2/s^2, \tag{45}$$

$$m_1^2m_2^2m_3^2m_4^2/s^4. \tag{46}$$

The singularities of (44) are given by Eq. (42) and the singularity at $s = 0$ is given by Eq. (41) with $c = -1$.

No change in dynamics can affect the positions of singularities associated with Eqs. (42) and (43). Such singularities are purely kinematical. If the amplitude has dynamical singularities, they must be associated with Eq. (41). This is reasonable since the constant c in Eq. (41) can take on a set of values each of which might depend on some coupling constant. Of course, kinematical singularities may also be associated with Eq. (41).

IV. ASYMPTOTIC BEHAVIOR

In order to see what kind of restrictions conformal invariance might impose on the asymptotic behavior of amplitudes, we assume that some amplitude behaves like

$$\beta t^\alpha, \tag{47}$$

where

$$\alpha = \alpha(s, p_1^2, p_2^2, p_3^2, p_4^2), \tag{48}$$

$$\beta = \beta(s, p_1^2, p_2^2, p_3^2, p_4^2), \tag{49}$$

for large t. Dilatation invariance of the amplitude implies that α is dilatation invariant. We also assume that the asymptotic behaviors of the derivatives in Eqs. (18)–(20) are given by the derivatives of the

[10] The manner in which these singular solutions arise is stated in Footnote 9.

asymptotic behavior (47).[11] The leading terms on the left-hand side in Eqs. (18)–(20) then behave like $(\ln t)^2 t^\alpha$. These terms are generated by derivatives acting on the α in t^α. From the requirement that the sum of the coefficients of the leading terms in each equation must vanish, we obtain three first-order partial differential equations for α. In fact, these equations are precisely the characteristic equations (22)–(24) for a t-independent characteristic surface. The solution of these equations has already been found in Sec. III. From Eq. (33), we see that

$$\alpha = \alpha\left(\frac{\{s-(m_1+m_2)^2\}^{\frac{1}{2}} \pm \{s-(m_3+m_4)^2\}^{\frac{1}{2}}}{m_1+m_2+m_3+m_4}\right), \tag{50}$$

where α is an arbitrary function of its argument.

It is interesting that α must identically satisfy the characteristic equations because this admits the possibility that an amplitude with asymptotic behavior (47) may have singularities along surfaces given by $\alpha = \text{const}$. An example of a function with a set of singularities along such surfaces is $1/\sin \pi\alpha$. Of course, Eqs. (18)–(20) provide restrictions on the nature of the singularities as well as their location.

It is also interesting that α is not related to the characteristic surfaces (42) and (43) which are associated with strictly kinematical singularities.

V. DISCUSSION

We have given partial differential equations for the surfaces along which the off-mass-shell amplitude can have singularities and we have found the possible locations (41)–(43) of t-independent singularities. Some features which are, perhaps, suggestive are the appearance of singularities (42) and (43) resembling thresholds and pseudothresholds and the dependence of other singularities (41) on arbitrary constants which might be related to coupling constants. However, there is no guarantee that such singularities will actually be present in the Feynman diagrams or the amplitudes of a theory with conformal-invariant dynamics. The formulas (27)–(29) and (41)–(43) derived here constitute severe constraints, but conformal invariance does not require the amplitude to have singularities along any given characteristic surface. This is obvious since a constant "amplitude" would trivially satisfy all the requirements. A constant could represent the Born term in $\lambda\phi^4$ theory.

[11] This is compatible with and possibly indicative of next leading terms behaving like $t^\alpha/\ln t$, $t^\alpha/(\ln t)^2$, etc., and $t^{\alpha-1}$, $t^{\alpha-2}$, etc.

The purely kinematical singularities (42) and (43), which resemble thresholds and pseudothresholds, originate from the singular solutions of a differential equation (34), while the other class of singularities (41), which could encompass dynamical singularities, originates from the general solution of the differential equation (34). It is very interesting that there is a natural way for the differential conditions of a higher space–time symmetry to provide two such categories of singularities.

Since the singularities (41)–(43) converge to $s = 0$ as the mass shell is approached, there can be no production of particles with mass. The differential equations (18)–(20) for the amplitude also provide restrictions on the nature of possible singularities. It should be noted that all of the formulas presented here are consistent with crossing symmetry.

One might wonder if any of the singularities (41)–(43) other than $s = 0$ could arise in individual Feynman diagrams. An interesting possibility is that they might arise in the process of making conformal-invariant subtractions or cut-offs. A diagram with a logarithmic ultraviolet divergence could develop an infrared divergence if we perform a subtraction at $s = 0$. Performing a subtraction at $s = \text{const} \neq 0$ would destroy dilatation invariance. Perhaps a resolution of the difficulty is to let the subtraction point depend on the p_i^2 and insist that the resulting amplitude satisfy the conditions (17)–(20) of conformal invariance. This conformal-invariant subtraction procedure could introduce singularities off the mass shell and not at $s = 0$. Because of the mathematical subtleties involved, this question will require careful investigation.

The restrictions of conformal symmetry apply to the whole amplitude as well as individual Feynman diagrams so that singularities connected with summing an infinite number of Feynman diagrams should also be restricted by Eqs. (27)–(29) and (41)–(43).

Finally, we observe that if the $(p_i^2)^{\frac{1}{2}}$ are replaced by nonvanishing masses, then the singularity structure (41)–(43) and its possible connection with the asymptotic form (47) are in accord with the usual features of S-matrix theory. The reasons for, and the significance of, this phenomenon are not clear.

ACKNOWLEDGMENTS

The author wishes to thank Professor Marshall Baker, Professor Naren Bali, and Professor Amnon Katz for many enlightening conversations.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Exact Static-Model Bootstrap Solutions for Arbitrary 2×2 Crossing Matrices*

JAMES T. CUSHING
Department of Physics, University of Notre Dame, Notre Dame, Indiana

(Received 6 January 1969)

Explicit, exact solutions, satisfying a bootstrap criterion in the form of Levinson's theorem and having real coupling constants, are exhibited for the two-channel Low equation for arbitrary real values of the parameter appearing in the 2×2 crossing matrix. This disproves a claim made recently in the literature that the bootstrap criterion will restrict values of the crossing-matrix parameter to those corresponding to the internal-symmetry group SU_2. It is demonstrated that the previous alleged proof is inconclusive.

I. INTRODUCTION AND SUMMARY OF RESULTS

There have been several attempts[1–5] to bootstrap internal symmetries in exactly soluble two-channel static models. To date, only Cunningham[4] has claimed that a bootstrap requirement in the form of Levinson's theorem, first suggested for this purpose by Huang and Low,[6] along with the usual requirements of unitarity, crossing, analyticity, and the reality of coupling constants, will restrict the crossing-matrix parameter to the integer values which correspond to the internal-symmetry group SU_2. This result seemed remarkable in light of the negative results of several previous attempts with similar models to bootstrap this internal symmetry. Warnock[5] has shown that solutions exist for the n-channel Low equation with an arbitrary $n \times n$ crossing matrix. Warnock's solutions have no subtractions, one bound state, and one CDD pole in one channel, with no bound states or CDD poles in the other channels. The present author[3] had studied a model with coupled inelastic channels in a two-dimensional space–time and found that bootstrap solutions exist for continuous real values of the crossing-matrix parameter. Finally, there is the impressive paper of Blankenbecler, Coon, and Roy,[7] which demonstrated that the assumption of diagonalizability of the S-matrix by a constant real orthogonal matrix rigorously and with no approximations implies SU_2 as a symmetry group for a system of pions as an essentially kinematic result independent of the strong-interaction dynamics. In this proof the existence of the third crossed channel plays a crucial roll.

Since Cunningham's proof assumes a static model with only two crossed channels, one subtraction, and no CDD poles, the papers mentioned above do not conclusively imply that his result cannot be correct. However, we shall demonstrate that Cunningham's arguments show only that a certain representation of the solutions to the two-channel problem does not exist for arbitrary values of the crossing-matrix parameter, not that no solutions exist. The source of the trouble comes from using Rothleitner's special solutions,[8] which have an infinite number of poles and zeros for arbitrary real values of the crossing-matrix parameter and then attempting to find an infinite-product representation of an arbitrary multiplicative function which will cancel all but a finite number of these poles and zeros in order to produce S-matrix elements with only a finite number of poles and zeros. It turns out that no infinite-product representation of the form assumed exists. However, we are able to show that a form of solution given by Martin and McGlinn[1] does allow S-matrix elements having all of the desired properties. We shall exhibit one class of bootstrap solutions. It is not claimed that these are all of the solutions nor that they are unique.

II. STATEMENT OF THE PROBLEM

A mathematical statement of the two-channel static model, describing the scattering of mesons of mass m from fixed baryons, is the following: If we let $z = x + iy$ be the complex energy variable, then, in units $\hbar = c = m = 1$, there are two S-matrix elements, $S_\alpha(z)$, $\alpha = 1, 2$, which are uncoupled from each other in the physical region (i.e., z real and above threshold). The complex z plane has the direct-channel cut running from $+1$ to $+\infty$ along the positive real axis and the crossed-channel cut running from -1 to $-\infty$ along the negative real axis. Unitarity implies that

$$S_\alpha(x + i\epsilon) = \exp\,[2i\delta_\alpha(x)], \qquad (1)$$

* Work supported in part by the U.S. Atomic Energy Commission.

[1] A. W. Martin and W. D. McGlinn, Phys. Rev. **136**, B1515 (1964).
[2] K. Huang and A. H. Mueller, Phys. Rev. **140**, B365 (1965).
[3] J. T. Cushing, Phys. Rev. **148**, 1558 (1966).
[4] A. W. Cunningham, J. Math. Phys. **8**, 716 (1967).
[5] R. L. Warnock, Phys. Rev. **170**, 1323 (1968).
[6] K. Huang and F. E. Low, J. Math. Phys. **6**, 795 (1965).
[7] R. Blankenbecler, D. D. Coon, and S. M. Roy, Phys. Rev. **156**, 1624 (1967).
[8] J. Rothleitner, Z. Physik **177**, 287 (1964).

where $x \geq 1$ and $\delta_\alpha(x)$ is real on this cut. Also, we have

$$S_\alpha(z^*) = S_\alpha^*(z). \tag{2}$$

The continuation of $S_\alpha(z)$ onto the second sheet through the unitarity cut, denoted by $S_\alpha^{(2)}(z)$, is given as

$$S_\alpha^{(2)}(z) = 1/S_\alpha(z). \tag{3}$$

Finally, the statement of crossing symmetry on the first sheet is

$$S_\alpha(-z) = \sum_{\beta=1}^{2} A_{\alpha\beta} S_\beta(z). \tag{4}$$

Since crossing twice must bring us back to the initial channel and since both channels are physically identical in this problem, the crossing-matrix elements must satisfy

$$\sum_{\gamma=1}^{2} A_{\alpha\gamma} A_{\gamma\beta} = \delta_{\alpha\beta}, \tag{5}$$

$$\sum_{\beta=1}^{2} A_{\alpha\beta} = 1. \tag{6}$$

The most general 2×2 crossing matrix satisfying these requirements is[1,2]

$$A = \frac{1}{(2t+1)} \begin{pmatrix} -1 & 2t+2 \\ 2t & 1 \end{pmatrix}. \tag{7}$$

The subset of these 2×2 crossing matrices corresponding to SU_2 has $t = n$, where n is zero or an integer. General solutions to the problem as defined above have been given for arbitrary values of t by Martin and McGlinn[1] and by Rothleitner.[8] We shall discuss both solutions in the following sections.

Our criterion for a bootstrap solution will be that suggested by Huang and Low,[6] namely, Levinson's theorem with no CDD poles,

$$\Delta\delta_\alpha \equiv \delta_\alpha(\infty) - \delta_\alpha(1) = -\pi n_\alpha, \tag{8}$$

where $\Delta\delta_\alpha$ is the change in phase of the phase shift $\delta_\alpha(z)$ of Eq. (1) between threshold and infinity along the physical cut and n_α is the number of bound states in channel α. As in Huang and Low,[6] the coupling constants squared are introduced via the Low equation for the function $h_\alpha(z)$ defined as

$$S_\alpha(z) \equiv 1 + 2iq(z)v(z)h_\alpha(z), \tag{9}$$

where we are scattering S-wave mesons and

$$q(z) \equiv (z^2 - 1)^{\frac{1}{2}}, \tag{10}$$

and the cutoff function $v(z)$ has the form

$$v(z) = \kappa^{2c}/(q^2 + \kappa^2)^c, \tag{11}$$

with $\kappa > 1$. For reasons that have been fully discussed in Ref. 6, in order to be able to have bootstrap solutions we must choose $c = 1$ and allow for one subtraction in the Low equation. (These are also Cunningham's choices.) The once-subtracted Low equation is

$$h_\alpha(z) = P_\alpha(z) + C_\alpha + \frac{z}{\pi} \int_1^\infty \frac{dx'}{x'} q(x')v(x') \times \left[\frac{|h_\alpha(x')|^2}{x' - z} - \sum_{\beta=1}^{2} A_{\alpha\beta} \frac{|h_\beta(x')|^2}{x' + z} \right], \tag{12}$$

where the subtraction was made at $z = 0$ and the C_α are subtraction constants. Here $P_\alpha(z)$ contains the crossing-symmetric combination of poles appearing in the strip $-1 < x < +1$,

$$P_\alpha(z) = \sum_i \left[\frac{\lambda_{i\alpha}}{x_i - z} + \sum_{\beta=1}^{2} A_{\alpha\beta} \frac{\lambda_{i\beta}}{x_i + z} \right], \tag{13}$$

where the x_i are the location of the bound states (i.e., $|x_i| < 1$) and the $\lambda_{i\alpha}$ are the coupling constants squared (i.e., $\lambda_{i\alpha} \geq 0$). The only use we shall have for the Low equation is to connect the residues of the $S_\alpha(z)$ at bound-state poles with the $\lambda_{i\alpha}$ of Eq. (13). If we consider the special case of one bound state in each channel (as will be sufficient for our purposes later), and if $\mathrm{p}_\alpha(z)$ denote the bound-state pole terms of the $S_\alpha(z)$ with effective coupling constants Λ_α defined as

$$\Lambda_\alpha \equiv 2v(x_\alpha)(1 - x_\alpha^2)^{\frac{1}{2}} \lambda_\alpha, \tag{14}$$

where x_α is the bound-state energy in channel α, then we see from Eqs. (7), (9), (13), and (14) that

$$\mathrm{p}_1(z) = \frac{-\Lambda_{11}}{x_1 - z} + \frac{1}{(2t+1)} \left[\frac{\Lambda_{11}}{x_1 + z} - \frac{2(t+1)\Lambda_{12}}{x_1 + z} \right] - \frac{\Lambda_{21}}{x_2 - z} + \frac{1}{(2t+1)} \left[\frac{\Lambda_{21}}{x_2 + z} - \frac{2(t+1)\Lambda_{22}}{x_2 + z} \right], \tag{15a}$$

$$\mathrm{p}_2(z) = \frac{-\Lambda_{12}}{x_1 - z} - \frac{1}{(2t+1)} \left[\frac{2t\Lambda_{11}}{x_1 + z} + \frac{\Lambda_{12}}{x_1 + z} \right] - \frac{\Lambda_{22}}{x_2 - z} - \frac{1}{(2t+1)} \left[\frac{2t\Lambda_{21}}{x_2 + z} + \frac{\Lambda_{22}}{x_2 + z} \right]. \tag{15b}$$

The two-channel S-matrix problem as defined in Eqs. (1)–(7) is solved by defining two auxiliary functions which have simple properties under unitarity and crossing,[1,8,9] of which the first is

$$B(z) = \frac{tS_1(z) + (t+1)S_2(z)}{S_2(z) - S_1(z)}. \tag{16}$$

[9] G. Wanders, Nuovo Cimento **23**, 817 (1962).

Crossings on the first-sheet and the second-sheet continuation follow from Eqs. (4) and (3) and are, respectively,

$$B(-z) = -B(z), \tag{17}$$

$$B^{(2)}(z) = -B(z) + 1. \tag{18}$$

The most general solution for $B(z)$ is[8,9]

$$B(z) = \frac{1}{2} + \frac{i}{\pi}\ln\left[z + (z^2-1)^{\frac{1}{2}}\right] + \frac{iz}{(z^2-1)^{\frac{1}{2}}}\beta(z)$$

$$= \frac{1}{\pi}\sin^{-1}(z) + \frac{iz}{(z^2-1)^{\frac{1}{2}}}\beta(z), \tag{19}$$

where $\beta(z)$ is an arbitrary real [i.e., $\beta(z^*) = \beta^*(z)$] even meromorphic function. For our purposes we shall choose

$$\beta(z) = \beta_0, \tag{20}$$

where β_0 is a real positive constant (cf. Refs. 4 and 6). We shall also need to know some of the properties of $q(z)$ defined in Eq. (10). They are

$$q^{(2)}(z) = -q(z), \tag{21a}$$

$$q^*(z) = -q(z^*), \tag{21b}$$

$$q(-z) = q(z), \tag{21c}$$

$$\operatorname{Im} q(z) \gtrsim 0, \tag{21d}$$

where the last statement holds on the first sheet only. [These properties of $q(z)$ are most easily proved by use of the mapping $z = \cosh(w)$, $w = u + iv$.]

We shall now state separately the solutions given by Rothleitner[8] and by Martin and McGlinn,[1] since these forms shall be important for what follows. Rothleitner defines a second function

$$U(z) = S_2(z) \tag{22}$$

for which crossing and unitarity become

$$U(-z) = U(z)\left[\frac{B(z) - t}{B(z) + t}\right], \tag{23}$$

$$U^{(2)}(z) = 1/U(z). \tag{24}$$

It is easy to see that the most general solution to these functional equations is

$$U(z) = U_0(z)D(z), \tag{25}$$

where $U_0(z)$ is any solution to Eqs. (23) and (24) and $D(z)$ is an arbitrary real even function satisfying

$$D^{(2)}(z) = 1/D(z). \tag{26}$$

Rothleitner constructs a $U_0(z)$ as

$$U_0(z) = \frac{\tan\left[\frac{\pi}{2}B(z)\right]\Gamma\left[\frac{B(z)-t}{2}\right]\Gamma\left[\frac{B(z)+t+1}{2}\right]}{\tan\left[\frac{\pi}{2}(B(z)+t)\right]\Gamma\left[\frac{B(z)+t}{2}\right]\Gamma\left[\frac{B(z)-t+1}{2}\right]}. \tag{27}$$

Finally, the S-matrix elements are given as

$$S_1(z) = U_0(z)D(z)\left[\frac{B(z)-t-1}{B(z)+t}\right]$$

$$= S_2(z)\left[\frac{B(z)-t-1}{B(z)+t}\right], \tag{28a}$$

$$S_2(z) = U_0(z)D(z). \tag{28b}$$

Martin and McGlinn write their S-matrix elements as

$$S_1(z) = A(z)[B(z) - t - 1]$$

$$= S_2(z)\left[\frac{B(z)-t-1}{B(z)+t}\right], \tag{29a}$$

$$S_2(z) = A(z)[B(z) + t], \tag{29b}$$

where $B(z)$ is defined by Eq. (16) and $A(z)$ is an arbitrary antisymmetric real analytic function in the cut z plane. Equations (29) automatically satisfy crossing. The second-sheet continuation for $A(z)$ is, in our notation,

$$A^{(2)}(z)A(z) = -1/[B(z) - t - 1][B(z) + t], \tag{30}$$

the most general solution of which is

$$A(z) = A_0(z)D(z), \tag{31}$$

where $A_0(z)$ is any particular solution of Eq. (30) and $D(z)$ is an arbitrary real even function satisfying Eq. (26). Martin and McGlinn constructed an $A_0(z)$ having no zeros or poles (away from both branch cuts), except for a simple pole at the origin. In our notation this particular solution is

$$A_0(z) = \frac{[1 - i(z^2-1)^{\frac{1}{2}}]}{(t + \frac{1}{2})z} \times \exp\left\{\frac{i(z^2-1)^{\frac{1}{2}}}{\pi}\int_1^\infty \frac{\ln[1 + F^2(x')]x'\,dx'}{(x'^2-1)^{\frac{1}{2}}(x'^2 - z^2)}\right\}, \tag{32}$$

where, for $1 \leq x < \infty$,

$$F(x) = \frac{-1}{(t + \frac{1}{2})} \times \left[\frac{1}{\pi}\ln[x + (x^2-1)^{\frac{1}{2}}] + \frac{x}{(x^2-1)^{\frac{1}{2}}}\beta(x)\right]. \tag{33}$$

Here, $\beta(x)$ is the arbitrary function in $B(z)$, Eq. (19). We shall make the choice $\beta(x) = \beta_0$, Eq. (20).

III. CUNNINGHAM'S PROOF

One possible way of attempting to construct bootstrap solutions is to begin with Rothleitner's solutions, Eqs. (27) and (28). Let us first examine $U_0(z)$, Eq. (27), when t is zero or a positive integer.

If t is zero or an even integer,

$$t = 2n, \quad n = 0, 1, 2, \cdots, \tag{34}$$

then, since

$$\tan [\tfrac{1}{2}\pi(B(z) + 2n)] = \tan [\tfrac{1}{2}\pi B(z)] \tag{35}$$

and

$$\Gamma(z + 1) = z\Gamma(z), \tag{36}$$

we find

$$U_0(z) = \frac{\Gamma\left[\frac{B(z)}{2} - n\right]\Gamma\left[\frac{B(z)+1}{2} + n\right]}{\Gamma\left[\frac{B(z)}{2} + n\right]\Gamma\left[\frac{B(z)+1}{2} - n\right]}, \tag{37}$$

which will be simply a ratio of two finite-order polynomials in $B(z)$. If t is an odd integer

$$t = 2n + 1, \quad n = 0, 1, 2, \cdots, \tag{38}$$

then, since

$$\tan \{\tfrac{1}{2}\pi[B(z) + 2n + 1]\} = -1/\tan [\tfrac{1}{2}\pi B(z)], \tag{39}$$

we have

$$U_0(z) = -\tan^2 [\tfrac{1}{2}\pi B(z)] \times \frac{\Gamma[\frac{1}{2}(B(z) - 1) - n]\Gamma[\frac{1}{2}B(z) + n + 1]}{\Gamma[\frac{1}{2}(B(z) - 1) + n + 1]\Gamma[\frac{1}{2}B(z) - n]}. \tag{40}$$

However, since $\tan^2 [\frac{1}{2}\pi B(z)]$ has the properties of a $D(z)$, Eq. (26), this term can be put into $D(z)$ so that $U_0(z)$ is again effectively a ratio of polynomials in $B(z)$.

Although Huang and Low[6] have investigated the roots of

$$B(z) = \gamma, \quad \gamma \text{ real}, \tag{41}$$

for arbitrary $\beta(z)$ of Eq. (19), we shall explicitly study the case when $\beta(z) = \beta_0 > 0$. In this case the substitution

$$z = \cosh (w), \qquad w = u + iv \tag{42}$$

shows that when z ranges over the entire (cut) first sheet, w is restricted to the range, $-\infty < u < +\infty$, $0 \leq v \leq \pi$. Then, since

$$\begin{aligned} B(z) &\equiv \frac{1}{2} + \frac{i}{\pi} \ln [z + q(z)] + \frac{iz}{q(z)}\beta_0 \\ &= \frac{1}{2} - \frac{v}{\pi} + \frac{\beta_0}{2} \frac{\sin (2v)}{\sinh^2 u \cos^2 v + \cosh^2 u \sin^2 v} \\ &\quad + i\left(\frac{u}{\pi} + \frac{\beta_0}{2} \frac{\sinh (2u)}{\sinh^2 u \cos^2 v + \cosh^2 u \sin^2 v}\right), \end{aligned} \tag{43}$$

Eq. (41) can have a root only when the imaginary part of Eq. (43) vanishes. This can happen only when

$$u = 0, \tag{44}$$

so that

$$z = \cos v, \tag{45}$$

which implies that the roots of Eq. (41) lie on the real axis between the cuts (i.e., $-1 \leq x \leq +1$). For x in this range Eq. (19) becomes

$$B(x) = \frac{1}{\pi} \sin^{-1} (x) + \frac{x}{(1 - x^2)^{\frac{1}{2}}} \beta_0, \tag{46}$$

which makes it evident that Eq. (41) has two real, symmetrically placed roots for any given value of γ. Therefore, we see from Eqs. (37), (40), and (28) that the $S_\alpha(z)$ can be constructed to have a finite number of poles and zeros if we choose the $D(z)$ to have only a finite number of poles and zeros (aside from the $\tan^2 [\frac{1}{2}\pi B(z)]$ factor when $t = 2n + 1$) whenever t is zero or an integer.

However, when t is not an integer, $U_0(z)$, Eq. (27), will have infinitely many poles and zeros. This is easily seen since the Γ functions in Eq. (27) have poles only when their arguments are 0 or negative integers, whereas the tangents have poles and zeros whenever their arguments are odd and even multiples of $\pi/2$, respectively. Unless t is an integer, there will be infinitely many of these poles and zeros which will not cancel each other. The form of $B(x)$ given in Eq. (46) shows that $B(x)$ will take on any arbitrarily large real value as x approaches threshold from below. Furthermore, when t is not an integer, $U_0(z)$ has both an even and an odd part, each of which has infinitely many zeros and poles. Therefore, we cannot simply factor out the troublesome term as we did the $\tan^2 [\frac{1}{2}\pi B(z)]$ when t was an odd integer. If we want the S-matrix elements of Eqs. (28) to have only a finite number of poles and zeros and if we use the $U_0(z)$ of Eq. (27), we must construct a $D(z)$ which will exactly cancel all but a finite number of the poles and zeros of $U_0(z)$.

Now Huang and Low[6] have shown that if $D(z)$ has only a *finite* number of zeros and poles, it can be represented as the finite product

$$D(z) = \prod_m \frac{1 - ir_m q}{1 + ir_m q} \prod_n \frac{(1 - a_n q)(1 + a_n^* q)}{(1 + a_n q)(1 - a_n^* q)}, \tag{47}$$

where $q(z)$ has been defined in Eq. (10). Cunningham[4] assumed that this could be extended to an infinite product if $D(z)$ had infinitely many poles and zeros. We shall show that this infinite product never converges. From Eq. (46) we see that the roots of

$$B(x_N) = N \tag{48}$$

are, for large N,

$$x_N^2 \to 1 - \beta_0^2/N^2, \tag{49}$$

so that these roots accumulate at $x = 1^-$ (and at -1^+). If t is not an integer, $D(z)$ must contain a factor

$$\prod_{m=0}^{\infty}\left(\frac{1 - ir_m q}{1 + ir_m q}\right) = \prod_{m=0}^{\infty}\left(1 + \frac{2r_m(1 - x^2)^{\frac{1}{2}}}{1 - r_m(1 - x^2)^{\frac{1}{2}}}\right), \tag{50}$$

where

$$r_m \xrightarrow[m\to\infty]{} m/\beta_0. \tag{51}$$

The necessary and sufficient condition that the infinite product in Eq. (50) converge is that

$$\sum_{m=0}^{\infty}\frac{r_m}{1 - r_m(1 - x^2)^{\frac{1}{2}}} \tag{52}$$

also converge.[10] A necessary and sufficient condition that this series converge is that if we find an $f(y)$ such that

$$f(y = m) = \frac{r_m}{1 - r_m(1 - x^2)^{\frac{1}{2}}}, \tag{53}$$

then

$$\int_c^{\infty} f(y)\, dy \tag{54}$$

must exist.[11] From Eq. (51) we see that

$$\frac{r_m}{1 - r_m(1 - x^2)^{\frac{1}{2}}} \xrightarrow[m\to\infty]{} \frac{m}{\beta_0 - m(1 - x^2)^{\frac{1}{2}}} \tag{55}$$

so that

$$f(y) = \frac{y}{\beta_0 - y(1 - x^2)^{\frac{1}{2}}}. \tag{56}$$

Since the integral in Eq. (54) diverges, the infinite product in $D(z)$, Eq. (47), also diverges. Therefore, all Cunningham[4] should have concluded was that the required $D(z)$ cannot be represented as the infinite product of Eq. (47), not that no such $D(z)$ exists. Furthermore, Levinson's theorem has not even been relevant here.

IV. EXPLICIT BOOTSTRAP SOLUTIONS

We shall now demonstrate that the S-matrix elements represented by Eqs. (29), (31), and (32) satisfy Levinson's theorem (8) and provide once-subtracted solutions to the Low equation (12). For this we shall need to know the analytic structure of the exponential in Eq. (32), namely,

$$I(z) \equiv \frac{i(z^2 - 1)^{\frac{1}{2}}}{\pi}\int_1^{\infty}\frac{\ln\,[1 + F^2(x')]x'\, dx'}{(x'^2 - 1)^{\frac{1}{2}}(x'^2 - z^2)}, \tag{57}$$

where $F(x)$ is given in Eq. (33). The detailed examination of this function will be left to the Appendix. The $B(z)$ we shall use is that given in Eqs. (19) and (20). Since the nonsingular part of the integrand in Eq. (57),

$$\frac{\ln\,[1 + F^2(x)]x}{(x^2 - 1)^{\frac{1}{2}}}, \tag{58}$$

is continuous and has a continuous first derivative on any open interval $1 < x < R$, where R is large but finite and positive, the function in Eq. (58) will satisfy a Hölder condition[12] here. Therefore, the integral in Eq. (57), and even its principal value, will exist[12] for all finite values of z^2, except possibly $z^2 = 1$. The behavior of $I(z)$ in the neighborhoods of $z = 1$ and $z = \infty$ is studied in the Appendix.

In order to verify that Levinson's theorem can be satisfied, we define, for the range $1 \leq x < \infty$,

$$A_0(x) = |A_0(x)| \exp\,[i\phi(x)], \tag{59a}$$

$$D(x) = \exp\,[i\theta(x)], \tag{59b}$$

$$B(x) - t - 1 = |B(x) - t - 1| \exp\,[i\psi(x)], \tag{59c}$$

$$B(x) + t = |B(x) + t| \exp\,[i\chi(x)], \tag{59d}$$

so that, from Eqs. (8) and (29), we have

$$\Delta\delta_1 = \tfrac{1}{2}(\Delta\phi + \Delta\theta + \Delta\psi), \tag{60a}$$

$$\Delta\delta_2 = \tfrac{1}{2}(\Delta\phi + \Delta\theta + \Delta\chi). \tag{60b}$$

From Eqs. (A5) and (A13) we see that

$$\Delta\phi = -\tfrac{1}{2}\pi - (-\tfrac{1}{2}\pi) = 0, \tag{61}$$

while Eqs. (19) and (20) show that

$$\Delta\psi = 0 = \Delta\chi. \tag{62}$$

Therefore,

$$\Delta\delta_1 = \Delta\delta_2 = \tfrac{1}{2}\Delta\theta. \tag{63}$$

Since Eqs. (9) and (11) imply that the cutoff poles must appear in each $S_\alpha(z)$, since Eq. (32) shows that $A_0(z)$ has a simple pole at $z = 0$, and since Eqs. (63) and (8) tell us to arrange for the same number of bound states in each channel, a reasonable form for $D(z)$ would be

$$D(z) = \frac{(1 - iq/\kappa)(1 - iq/s_0)}{(1 + iq/\kappa)(1 + iq/s_0)}, \tag{64}$$

where

$$B(x_0) - t - 1 = 0, \quad 0 < x_0 < 1,$$

$$s_0 \equiv (1 - x_0^2)^{\frac{1}{2}}, \quad 0 < s_0 < 1. \tag{65}$$

We see from Eqs. (29) that the pole at $z = x_0$ will appear only in $S_2(z)$ but not in $S_1(z)$. Therefore,

[10] E. C. Titchmarsh, *The Theory of Functions* (Oxford University Press, London, 1939), 2nd ed., p. 14.

[11] W. Kaplan, *Advanced Calculus* (Addison-Wesley Publ. Co., Reading, Mass., 1952), p. 315.

[12] N. I. Muskhelishvili, *Singular Integral Equations* (P. Noordhoff Ltd., Gröningen, The Netherlands, 1953), pp. 11 and 26.

Eq. (64) implies that

$$\tan\left[\tfrac{1}{2}\theta(x)\right] = -\left[\frac{q(x)(\kappa + s_0)}{\kappa s_0 - q^2(x)}\right] \tag{66}$$

or

$$\tfrac{1}{2}\Delta\theta = -\pi, \tag{67}$$

so that

$$\Delta\delta_1 = \Delta\delta_2 = -\pi. \tag{68}$$

Therefore, if we consider the target a bound state in channel 1 at $z = 0$ and another bound state in channel 2 at $z = x_0$, then Eqs. (15), with $\Lambda_{12} = 0 = \Lambda_{21}$, require that

$$\frac{2(t+1)}{(2t+1)}\Lambda_{11} = \text{Res}\,[S_1(z)]\big|_{z=0}, \tag{69a}$$

$$\Lambda_{22} = \text{Res}\,[S_2(z)]\big|_{z=x_0}, \tag{69b}$$

where, from Eq. (8), $n_1 = n_2 = 1$.

We must now verify that Λ_{11} and Λ_{22} are positive in order that there be no ghosts. If we define the manifestly positive constant C as [cf. Eq. (32)]

$$C = (t + \tfrac{1}{2}) \lim_{z\to 0}[zA_0(z)], \tag{70}$$

then Eqs. (28a), (64), and (69a) imply that

$$\Lambda_{11} = C\left(\frac{\kappa+1}{\kappa-1}\right)\left(\frac{1+s_0}{1-s_0}\right) > 0, \tag{71}$$

since $\kappa > 1$, Eq. (11), and $s_0 < 1$, Eq. (65). Similarly, we find that

$$\Lambda_{22} = 2(2t+1)\frac{s_0^2}{x_0}\left(\frac{1 - s_0/\kappa}{1 + s_0/\kappa}\right)A_0(x_0) > 0. \tag{72}$$

Finally, we see from the representation given in Eqs. (29) and (32) that the S-matrix elements have no poles on the first sheet, aside from the bound-state poles just discussed, only a finite number of zeros, and, from Eq. (A20),

$$\lim_{|z|\to\infty} \frac{|S_\alpha(z)|}{[\ln|z|]^{\frac{1}{2}}} = 0 \tag{73}$$

along any ray on the first sheet so that one subtraction is sufficient for the Low equation.

ACKNOWLEDGMENT

The author wishes to thank Professor W. D. McGlinn for several valuable discussions and criticisms of the work contained in this paper.

APPENDIX

In this appendix we shall study the threshold and asymptotic behaviors of

$$I(z) = \frac{i(z^2-1)^{\frac{1}{2}}}{\pi}\int_1^\infty \frac{\ln[1 + F^2(x')]x'\,dx'}{(x'^2-1)^{\frac{1}{2}}(x'^2 - z^2)}, \tag{A1}$$

where

$$F(x) = \frac{-1}{(t+\frac{1}{2})} \times \left\{\frac{1}{\pi}\ln[x + (x^2-1)^{\frac{1}{2}}] + \frac{x}{(x^2-1)^{\frac{1}{2}}}\beta_0\right\}. \tag{A2}$$

We begin by computing $I(z)$ as z approaches threshold. We shall show that the most singular part of the integrand of Eq. (A1) yields a finite result at threshold, so that the less singular terms yield nothing. We simply state a result which is easily verified by an application of Cauchy's integral theorem,

$$-\frac{1}{\pi}\int_1^\infty \frac{\ln(y-1)\,dy}{(y-1)[y-(z-1)]} = \frac{\ln(e^{i\pi}z)}{i(z)^{\frac{1}{2}}}, \tag{A3}$$

where the phase of the argument of the logarithm has been chosen to produce a cut for $0 < x < \infty$ and a real logarithm for $-\infty < x < 0$. Then, since $F(x)$ is dominated by the second term in the square brackets as z approaches 1,

$$I(z) \xrightarrow[z\to 1]{} \frac{i(z^2-1)^{\frac{1}{2}}}{2\pi} \times \int_1^\infty \frac{\ln\{[\beta_0^2/(t+\frac{1}{2})^2][1/(y-1)]\}\,dy}{(y-1)^{\frac{1}{2}}(y-z^2)} = \ln\left[\frac{(t+\frac{1}{2})}{\beta_0}\frac{(z^2-1)^{\frac{1}{2}}}{i}\right], \tag{A4}$$

where we have taken the upper limit on the integral to be $1 + \epsilon$ and then added on a piece which vanishes in the limit when z approaches 1. Therefore, from Eq. (32),

$$A_0(z) \xrightarrow[z\to 1]{} (z^2-1)^{\frac{1}{2}}/i\beta_0.$$

In terms of the phase $\phi(x)$, defined in Eq. (59a), this implies that

$$\phi(1) = -\tfrac{1}{2}\pi. \tag{A5}$$

Also, Eqs. (19), (20), (29), (32), and (64) show that

$$S_\alpha(z) \xrightarrow[z\to 1]{} 1, \tag{A6}$$

so that the S-matrix elements have the proper threshold behavior.

It is somewhat more difficult to find the asymptotic behavior of $I(z)$ as $|z|$ approaches infinity along an arbitrary direction in the z plane. Even though $I(z)$ in Eq. (A1) is a function of z^2 and the whole z^2 plane is only the upper half z plane, the reality property of $A_0(z)$ [i.e., $A_0(z^*) = A_0^*(z)$] and of the $S_\alpha(z)$ will imply that there are no essential singularities anywhere on the first sheet of the cut z plane if there are none on

the first sheet of the cut z^2 plane. If we let R be an arbitrarily large but fixed constant and choose $z > R$, then the dominant asymptotic behavior of $I(z)$ is given as

$$\begin{aligned} I(z) &\xrightarrow[|z|\to\infty]{} \frac{i(z^2-1)^{\frac{1}{2}}}{\pi}\int_R^\infty \ln\left[\frac{\ln^2(x)}{(t+\frac{1}{2})^2\pi^2}\right]\frac{x\,dx}{(x^2-1)^{\frac{1}{2}}(x^2-z^2)} \\ &\approx \frac{i(z^2-1)^{\frac{1}{2}}}{\pi}\left\{-2\ln[\pi(t+\tfrac{1}{2})]\int_1^\infty \frac{dx}{(x^2-z^2)} + \int_R^\infty \frac{\ln[\ln^2(x)]x\,dx}{(x^2-1)^{\frac{1}{2}}(x^2-z^2)}\right\} \\ &\to \ln[\pi(t+\tfrac{1}{2})]\left[1-\frac{2i}{z}\right] + \frac{2iz}{\pi}\int_R^\infty \frac{\ln[\ln(x)]\,dx}{(x^2-z^2)} \\ &= \ln[\pi(t+\tfrac{1}{2})]\left(1-\frac{2i}{z}\right) - \ln(\ln R) - \frac{i}{\pi}\int_R^\infty \frac{dx}{x\ln(x)}\ln\left(\frac{x-z}{x+z}\right), \end{aligned} \tag{A7}$$

where we have again added or dropped terms which vanish when the limit is taken. Let

$$\begin{aligned} h(z) &= -\frac{i}{\pi}\int_R^\infty \frac{dx}{x\ln(x)}\ln\left(\frac{x-z}{x+z}\right) \\ &= \frac{i}{\pi}\int_R^\infty \frac{dx}{x\ln(x)}\ln\left(\frac{x+z}{x-z}\right). \end{aligned} \tag{A8}$$

If we let z be real and positive, then

$$\begin{aligned} h(z) &= \ln(\ln R) - \ln(\ln z) \\ &\quad + \frac{i}{\pi}\int_R^z \frac{dx}{x\ln(x)}\ln\left(\frac{x+z}{z-x}\right) \\ &\quad + \frac{i}{\pi}\int_z^\infty \frac{dx}{x\ln(x)}\ln\left(\frac{x+z}{x-z}\right) \\ &\xrightarrow[z\to\infty]{} -\ln(\ln z) + \frac{i}{\pi}\left[\int_{R/z}^1 \frac{dy}{y\ln(zy)}\right. \\ &\quad \left.\times \ln\left(\frac{1+y}{1-y}\right) + \int_0^1 \frac{dy}{y\ln(z/y)}\ln\left(\frac{1+y}{1-y}\right)\right]. \end{aligned} \tag{A9}$$

Since a standard definite integral is

$$\int_0^1 \ln\left(\frac{1+x}{1-x}\right)\frac{dx}{x} = \frac{\pi^2}{4}, \tag{A10}$$

we see that the two integrals in Eq. (A9) have the respective bounds

$$\frac{\pi^2}{4}\frac{1}{\ln(R)}, \tag{A11a}$$

$$\frac{\pi^2}{4}\frac{1}{\ln(z)}. \tag{A11b}$$

Therefore, as z approaches infinity along the positive real axis,

$$I(z) \xrightarrow[z\to\infty]{} \ln\left[\frac{\pi(t+\frac{1}{2})}{\ln z}\right] + i\frac{\pi}{4}\frac{1}{\ln R}, \tag{A12}$$

where the last term can be made as small as we please. From Eqs. (32) and (59a) we see that

$$\phi(\infty) = -\tfrac{1}{2}\pi, \tag{A13}$$

while Eqs. (19), (20), (29), and (64) show that, on the real z axis,

$$|S_\alpha(z)| \xrightarrow[z\to\infty]{} 1, \tag{A14}$$

as must be the case in order to satisfy unitarity, Eq. (1).

Finally, we must examine $I(z)$ as z approaches infinity along any ray on the first sheet. If we let $z = \rho e^{i\phi}$, then

$$\begin{aligned} \ln\left(\frac{x+z}{x-z}\right) &= \ln\left\{\frac{[(x^2-\rho^2)^2 + 4x^2\rho^2\sin^2\phi]^{\frac{1}{2}}}{x^2+\rho^2-2x\rho\cos\phi}\right\} \\ &\quad + i\tan^{-1}\left(\frac{2x\rho\sin\phi}{x^2-\rho^2}\right), \end{aligned} \tag{A15}$$

where $0 < \phi < \pi$, since we have already studied the case when z is real. Also, in order to recover the proper expression for real positive z (i.e., $\phi = 0$), the inverse tangent in Eq. (A15) ranges from π to $\pi/2$ as its argument goes from zero to infinity, when $\rho > x$, but from 0 to $\pi/2$ if $\rho < x$. Furthermore, since we shall need only $|S_\alpha(z)|$, and therefore $|A_0(z)|$, we can examine just the real part of $h(z)$, Eq. (A8), since the imaginary part is bounded:

$$\begin{aligned} \mathrm{Re}\, h(z) &= -\frac{1}{\pi}\int_R^\infty \frac{dx}{x\ln(x)}\tan^{-1}\left(\frac{2x\rho\sin\phi}{x^2-\rho^2}\right) \\ &= -\frac{1}{\pi}\left\{\int_R^\rho \frac{dx}{x\ln(x)}\tan^{-1}\left[\frac{2(x/\rho)\sin\phi}{(x/\rho)^2-1}\right]\right. \\ &\quad \left.+\int_\rho^\infty \frac{dx}{x\ln(x)}\tan^{-1}\left[\frac{2(\rho/x)\sin\phi}{1-(\rho/x)^2}\right]\right\} \\ &= -\frac{1}{\pi}\left[\int_{R/\rho}^1 \frac{dy}{y\ln(\rho y)}\tan^{-1}\left(\frac{2y\sin\phi}{-1+y^2}\right)\right. \\ &\quad \left.+\int_0^1 \frac{dy}{y\ln(\rho/y)}\tan^{-1}\left(\frac{2y\sin\phi}{1-y^2}\right)\right]. \end{aligned} \tag{A16}$$

The last term in Eq. (A16) vanishes as $\rho \to \infty$ since

the integral is bounded by

$$\frac{1}{\ln \rho}\int_0^1 \frac{dy}{y}\tan^{-1}\left(\frac{2y\sin\phi}{1-y^2}\right), \tag{A17}$$

and the integral in (A17) clearly exists [cf. the statement about the range of the inverse tangent following Eq. (A15)]. The first integral in Eq. (A16) has the following asymptotic bounds, again because of the range of the inverse tangent:

$$\ln[(\ln\rho)^{\frac{1}{2}}] < \frac{1}{\pi}\int_{R/\rho}^1 \frac{dy}{y\ln(\rho y)} \times \tan^{-1}\left(\frac{2y\sin\phi}{-1+y^2}\right) < \ln(\ln\rho). \tag{A18}$$

Because of the minus sign in Eq. (A16), we see from Eqs. (A1), (32), (29), and (19) that, asymptotically for z not real,

$$|A_0(z)| < \pi(t+\tfrac{1}{2})(\ln|z|)^{-\frac{1}{2}} \tag{A19}$$

so that

$$|S_\alpha(z)| < (\ln|z|)^{\frac{1}{2}}, \tag{A20}$$

which implies that one subtraction will be sufficient for the Low equation. That one subtraction is necessary for bootstrap solutions has been shown by Huang and Low.[6]

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Asymptotic Fields of the Lee Model*†

STANLEY JERNOW‡ AND EMIL KAZES
The Pennsylvania State University, University Park, Pennsylvania

(Received 20 November 1968)

The field equations for the Lee model are solved by constructing eigenfields from the most general possible combination of bare fermion and boson operators. These solutions are found to require an infinite number of boson terms, the coefficients of which obey the integral equations of scattering theory. It is discovered that the algebra of the Lee-model fermion eigenfields is not that of free-particle operators. The anticommutator of the ith fermion eigenfield A_i^+ has the form $\{A_i, A_i^+\}_+ = 1 - \sum_{j\neq i} A_j^+ A_j$, where the sum is over all other fermion eigenfields. An algebra of this type is not peculiar to the Lee model. It will hold for all fields which obey a general Pauli exclusion principle and which are orthogonal to one another. The explanation for this algebra lies in the fact that these eigenfields represent operators which create states and not particles. Several other models exhibiting this same behavior (including the harmonic oscillator) are presented; for these models boson fields may also be constructed which have the above algebra. The asymptotic convergence of these eigenfields is not examined. However, it is found that all such Lee-model fermion fields, including those constructed by the Yang–Feldman method, must satisfy the above algebra and do not enjoy a free-fermion canonical algebra.

1. INTRODUCTION

Asymptotic fields are essential to the LSZ formulation of field theory and have been studied intensely by many authors. In order further to understand the nature of asymptotic fields, it is useful to examine them for simple soluble models; in-fields for a separable potential model have already been constructed,[1] while Ezawa[2] has previously studied the Lee model[3,4] in its lowest sector.

One common way to define an asymptotic field is through the Yang–Feldman construction.[5,6] This construction yields fields which automatically obey the equation

$$[H, A_i^+]_- = E_i A_i^+ \tag{1.1}$$

in the Heisenberg picture, where E_i is an energy, H is the relevant Hamiltonian, and A_i^+ is the asymptotic field creation operator. In the present work the solutions to Eq. (1.1) are examined for the Lee model in momentum space. Here the field operators are constructed, not by the Yang–Feldman method, but by the most general possible expansion in terms of bare fields.[7] This procedure allows us to examine the in-fields in greater detail than Ezawa has done.

* Supported in part by the U.S. Atomic Energy Commission.
† Based on part of a doctoral thesis submitted by Stanley Jernow to the graduate school of The Pennsylvania State University.
‡ Present address: Knolls Atomic Power Laboratory, Schenectady, New York.
[1] E. Kazes, Phys. Rev. **135**, B477 (1964).
[2] H. Ezawa, Ann. Phys. (N.Y.) **24**, 46 (1963).
[3] T. D. Lee, Phys. Rev. **95**, 1329 (1954).
[4] G. Källén and W. Pauli, Kgl. Danske Vidensk. Selsk. Mat.-Fys. Medd. **30**, No. 7 (1955).
[5] C. N. Yang and D. Feldman, Phys. Rev. **79**, 972 (1950).
[6] W. Zimmermann, Nuovo Cimento **10**, 597 (1958).
[7] H. Ezawa, K. Kikkawa, and H. Umezawa [Nuovo Cimento **25**, 1141 (1962)] have given a similar construction for in- and out-fields.

the integral is bounded by

$$\frac{1}{\ln \rho}\int_0^1 \frac{dy}{y}\tan^{-1}\left(\frac{2y\sin\phi}{1-y^2}\right), \tag{A17}$$

and the integral in (A17) clearly exists [cf. the statement about the range of the inverse tangent following Eq. (A15)]. The first integral in Eq. (A16) has the following asymptotic bounds, again because of the range of the inverse tangent:

$$\ln[(\ln\rho)^{\frac{1}{2}}] < \frac{1}{\pi}\int_{R/\rho}^1 \frac{dy}{y\ln(\rho y)} \times \tan^{-1}\left(\frac{2y\sin\phi}{-1+y^2}\right) < \ln(\ln\rho). \tag{A18}$$

Because of the minus sign in Eq. (A16), we see from Eqs. (A1), (32), (29), and (19) that, asymptotically for z not real,

$$|A_0(z)| < \pi(t+\tfrac{1}{2})(\ln|z|)^{-\frac{1}{2}} \tag{A19}$$

so that

$$|S_\alpha(z)| < (\ln|z|)^{\frac{1}{2}}, \tag{A20}$$

which implies that one subtraction will be sufficient for the Low equation. That one subtraction is necessary for bootstrap solutions has been shown by Huang and Low.[6]

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Asymptotic Fields of the Lee Model*†

STANLEY JERNOW‡ AND EMIL KAZES
The Pennsylvania State University, University Park, Pennsylvania

(Received 20 November 1968)

The field equations for the Lee model are solved by constructing eigenfields from the most general possible combination of bare fermion and boson operators. These solutions are found to require an infinite number of boson terms, the coefficients of which obey the integral equations of scattering theory. It is discovered that the algebra of the Lee-model fermion eigenfields is not that of free-particle operators. The anticommutator of the ith fermion eigenfield A_i^+ has the form $\{A_i, A_i^+\}_+ = 1 - \sum_{j\neq i} A_j^+ A_j$, where the sum is over all other fermion eigenfields. An algebra of this type is not peculiar to the Lee model. It will hold for all fields which obey a general Pauli exclusion principle and which are orthogonal to one another. The explanation for this algebra lies in the fact that these eigenfields represent operators which create states and not particles. Several other models exhibiting this same behavior (including the harmonic oscillator) are presented; for these models boson fields may also be constructed which have the above algebra. The asymptotic convergence of these eigenfields is not examined. However, it is found that all such Lee-model fermion fields, including those constructed by the Yang–Feldman method, must satisfy the above algebra and do not enjoy a free-fermion canonical algebra.

1. INTRODUCTION

Asymptotic fields are essential to the LSZ formulation of field theory and have been studied intensely by many authors. In order further to understand the nature of asymptotic fields, it is useful to examine them for simple soluble models; in-fields for a separable potential model have already been constructed,[1] while Ezawa[2] has previously studied the Lee model[3,4] in its lowest sector.

One common way to define an asymptotic field is through the Yang–Feldman construction.[5,6] This construction yields fields which automatically obey the equation

$$[H, A_i^+]_- = E_i A_i^+ \tag{1.1}$$

in the Heisenberg picture, where E_i is an energy, H is the relevant Hamiltonian, and A_i^+ is the asymptotic field creation operator. In the present work the solutions to Eq. (1.1) are examined for the Lee model in momentum space. Here the field operators are constructed, not by the Yang–Feldman method, but by the most general possible expansion in terms of bare fields.[7] This procedure allows us to examine the in-fields in greater detail than Ezawa has done.

* Supported in part by the U.S. Atomic Energy Commission.
† Based on part of a doctoral thesis submitted by Stanley Jernow to the graduate school of The Pennsylvania State University.
‡ Present address: Knolls Atomic Power Laboratory, Schenectady, New York.
[1] E. Kazes, Phys. Rev. **135**, B477 (1964).
[2] H. Ezawa, Ann. Phys. (N.Y.) **24**, 46 (1963).
[3] T. D. Lee, Phys. Rev. **95**, 1329 (1954).
[4] G. Källén and W. Pauli, Kgl. Danske Vidensk. Selsk. Mat.-Fys. Medd. **30**, No. 7 (1955).
[5] C. N. Yang and D. Feldman, Phys. Rev. **79**, 972 (1950).
[6] W. Zimmermann, Nuovo Cimento **10**, 597 (1958).
[7] H. Ezawa, K. Kikkawa, and H. Umezawa [Nuovo Cimento **25**, 1141 (1962)] have given a similar construction for in- and out-fields.

It is discovered that the fermion eigenfield solutions to Eq. (1.1) must obey certain stringent conditions: they must have the noncanonical algebra

$$\{A_i, A_i^+\}_+ = 1 - \sum_{j\neq i} A_j^+ A_j = \prod_{j\neq i}(1 - A_j^+ A_j), \quad (1.2)$$

where the sum (or product) is over all other eigenfields. This algebra is explained by the fact that these fields represent operators which create states and not particles; it is proved that any eigenfield which obeys orthogonality and a general Pauli exclusion principle (as defined in Sec. 3) has the algebra of Eq. (1.2). In Sec. 3 it is shown that the harmonic oscillator can be solved to yield fields having this same algebra. Another model which displays similar behavior is also given.

The asymptotic convergence in time of these eigenfield solutions has not been examined; thus a question remains as to whether the field operators are truly asymptotic fields or rather some linear combination of in- and out-fields. In partial resolution of this question, it should be noted that the mathematical expressions which occur in these solutions do contain appropriate $\pm i\epsilon$ factors, and the fields thus refer to incoming or outgoing waves (just as do the Yang–Feldman retarded or advanced Green's functions; see Ref. 1 on this point). It is for these reasons that the words asymptotic field and in-field are used freely in the following sections. The eigenfield solutions, furthermore, create scattering states for the sector in which they are examined, and are thus in-fields in that sector; they are identical to Ezawa's[2] solution in the N–θ sector. Finally, the fact that the most general fermion solutions to Eq. (1.1) are found to obey the anticommutation relations of Eq. (1.2) implies that any Yang–Feldman Lee-model fermion in-field must also exhibit this algebra.

2. ASYMPTOTIC FIELDS OF THE LEE MODEL

The Lee model describes three particles, two static fermions N and V, and a lightboson θ which may move relativistically. The Hamiltonian is

$$H = H_\theta + m_N N^+N + m_v V^+V + \lambda G^+N^+V + \lambda G V^+N, \quad (2.1)$$

where

$$H_\theta \equiv \int d\mathbf{k}\omega_k \theta^+(k)\theta(k),$$

$$G^+ \equiv \int d\mathbf{k} f(\omega_k)\theta^+(k),$$

$$\omega_k = (\mu^2 + k^2)^{\frac{1}{2}},$$

and

$$\{V, V^+\}_+ = \{N, N^+\}_+ = +1,$$

$$\{V^+, V^+\}_+ = \{N^+, N^+\}_+ = \{V, N\}_+ = \{V, N^+\}_+ = 0, \quad (2.2)$$

$$[\theta(k), \theta^+(p)]_- = \delta(\mathbf{p} - \mathbf{k}),$$

$$[\theta(k), \theta(p)]_- = 0.$$

All bare fermion operators commute with all bare boson operators. The real function $f(\omega_k)$ is assumed smooth and to have a form sufficient to ensure the convergence of all the integrals which will be encountered. Vector symbols for momentum indices of boson operators have been suppressed. In Eq. (2.1) all symbols are bare. Renormalized quantities will be carefully defined when they appear. Often creation and destruction operators for the boson will be symbolically referred to in the text as θ^+ and θ, respectively.

There are two constants of the motion for this Hamiltonian: the operators

$$Q_1 \equiv N^+N + V^+V \qquad \text{(eigenvalue } q_1\text{)}$$

and

$$Q_2 \equiv \int \theta^+(p)\theta(p)\, d\mathbf{p} + V^+V \quad \text{(eigenvalue } q_2\text{)} \quad (2.3)$$

both commute with H. The eigenstates of these operators thus separate into mutually orthogonal groups (sectors) of common q_1 and q_2. If an operator is described as being known to within a certain sector, it means that the matrix elements of the operator between states belonging to that sector and to sectors of lower q_1 and q_2 are known exactly; if this is not the case, it will be made clear in the text. That sector with $q_1 = q_2 = 1$ is known as the N–θ sector.

Corresponding to the three elementary particle fields of the model, we expect to be able to construct three asymptotic field operators obeying the eigenvalue equations

$$[H, \theta_{\text{in}}^+(p)]_- = \omega_p \theta_{\text{in}}^+(p), \quad (2.4)$$

$$[H, V_{\text{in}}^+]_- = M_v V_{\text{in}}^+, \quad (2.5)$$

$$[H, N_{\text{in}}^+]_- = m_N N_{\text{in}}^+, \quad (2.6)$$

where the energies of $\theta_{\text{in}}^+(p)$ and N_{in}^+ are bare and M_v is physical.

To solve the above equations the in-fields will be expressed as linear combinations of all fermion terms of the correct quantum numbers with corresponding

boson coefficients:

$$\theta^+_{\rm in}(p) \equiv F^+_1(p) + F^+_2(p)V^+N + F^+_3(p)N^+V + F^+_4(p)V^+V + F^+_5(p)N^+N + F^+_6(p)N^+NV^+V, \quad (2.7)$$

$$V^+_{\rm in} \equiv L^+_1V^+ + L^+_2V^+N^+N + L^+_3N^+V^+V + L^+_4N^+, \quad (2.8)$$

$$N^+_{\rm in} \equiv P^+_1V^+ + P^+_2V^+N^+N + P^+_3N^+V^+V + P^+_4N^+. \quad (2.9)$$

The quantities $F^+_1(p) \to F^+_6(p)$, $L^+_1 \to L^+_4$, $P^+_1 \to P^+_4$, which are defined by Eqs. (2.7), (2.8), and (2.9), are boson operators containing θ-particle creation and destruction operators; therefore, they commute with all fermion operators, Note that the fermion terms, because of their quantum numbers, are the only ones which may appear in the in-fields (no matter how these in-fields are constructed). The in-fields of Eqs. (2.7), (2.8), and (2.9) are thus the most general asymptotic fields possible. Obviously, $F^+_1(p)$, $F^+_4(p)$, $F^+_5(p)$, $F^+_6(p)$, and L^+_3 and L^+_4 all have the quantum numbers of a θ^+ operator; $F^+_3(p)$ has the quantum numbers of $\theta^+\theta^+$; P^+_1 and P^+_2 have θ quantum numbers; and $F^+_2(p)$, L^+_1, L^+_2, P^+_3, and P^+_4 have the quantum numbers of the vacuum. Once these boson coefficients are known, the problem is solved.

The determining equations for these boson operators are found by substituting Eqs. (2.7)–(2.9) into Eq. (2.4)–(2.6). By equating coefficients of linearly independent terms—that is, coefficients of similar normal ordered fermion operators—the following equations of motion result:

$$\omega_p F^+_1(p) = [H_\theta, F^+_1(p)]_-, \quad (2.10a)$$

$$(\omega_p + m_N - m_v)F^+_2(p) = [H_\theta, F^+_2(p)]_- + \lambda[G, F^+_1(p)]_- - \lambda F^+_4(p)G + \lambda GF^+_5(p), \quad (2.10b)$$

$$(\omega_p + m_v - m_N)F^+_3(p) = [H_\theta, F^+_3(p)]_- + \lambda[G^+, F^+_1(p)]_- + \lambda G^+F^+_4(p) - \lambda F^+_5(p)G^+, \quad (2.10c)$$

$$\omega_p F^+_4(p) = [H_\theta, F^+_4(p)]_- - \lambda F^+_2(p)G^+ + \lambda GF^+_3(p), \quad (2.10d)$$

$$\omega_p F^+_5(p) = [H_\theta, F^+_5(p)]_- - \lambda F^+_3(p)G + \lambda G^+F^+_2(p), \quad (2.10e)$$

$$\omega_p F^+_6(p) = [H_\theta, F^+_6(p)]_- - \lambda[G, F^+_3(p)]_- - \lambda[G^+, F^+_2(p)]_-, \quad (2.10f)$$

and

$$(M_v - m_v)L^+_1 = [H_\theta, L^+_1]_- + \lambda GL^+_4, \quad (2.11a)$$

$$(M_v - m_v)L^+_2 = [H_\theta, L^+_2]_- - \lambda[G, L^+_4]_- + \lambda L^+_3G, \quad (2.11b)$$

$$(M_v - m_N)L^+_3 = [H_\theta, L^+_3]_- - \lambda[G^+, L^+_1]_- + \lambda L^+_2G^+, \quad (2.11c)$$

$$(M_v - m_N)L^+_4 = [H_\theta, L^+_4]_- + \lambda G^+L^+_1. \quad (2.11d)$$

The equations for $P^+_1 \to P^+_4$ are the same as Eqs. (2.11) with m_N substituted for M_v (the solutions P^+_n differ from L^+_n by their quantum numbers).

These equations may be solved by constructing the most general $F^+_1(p) \to F^+_6(p)$, $L^+_1 \to L^+_4$, $P^+_1 \to P^+_4$ in terms of θ^+ and θ operators. An example is

$$F^+_n(p) = \int \mathscr{f}_{n,1}(p; p_1)\theta^+(p_1) + \int \mathscr{f}_{n,2}(p; p_1, p_2; k_1)\theta^+(p_1)\theta^+(p_2)\theta(k_1) + \int \mathscr{f}_{n,3}(p; p_1, p_2, p_3; k_1, k_2) \times \theta^+(p_1)\theta^+(p_2)\theta^+(p_3)\theta(k_1)\theta(k_2) + \cdots, \quad n = 1, 4, 5, 6. \quad (2.12)$$

Repeated variables are integrated over; corresponding $d\mathbf{p}_i$, $d\mathbf{k}_j$ have been suppressed. For ease of recognition momentum variables corresponding to a θ^+ operator have been labeled $\mathbf{p}_i$, while those corresponding to a θ have been labeled $\mathbf{k}_j$. The coefficients have been made symmetric separately in all p_i and in all k_j; for example,

$$\mathscr{f}_{4,3}(p; p_1, p_2, p_3; k_1, k_2) = \mathscr{f}_{4,3}(p; p_1, p_2, p_3; k_2, k_1) = \mathscr{f}_{4,3}(p; p_1, p_3, p_2; k_1, k_2) = \mathscr{f}_{4,3}(p; p_3, p_2, p_1; k_1, k_2) = \mathscr{f}_{4,3}(p; p_1, p_3, p_2; k_2, k_1) = \text{etc.}$$

Similar constructions are true for $F^+_2(p)$, $F^+_3(p)$, all L^+_n, and all P^+_n.

The normal-ordered expressions for $F^+_n(p)$, L^+_n, and P^+_n contain an infinite number of terms, as required by the equations of motion. That is, after substitution of the $F^+_n(p)$, L^+_n, or P^+_n into the corresponding equations of motion, those equations must be normal ordered in order that coefficients of like linearly independent θ-particle operators may be equated. This process would lead to equations of the symbolic form

$$\theta^+ + \theta^+\theta^+\theta = \theta^+ + \theta^+\theta^+\theta + \theta^+\theta^+\theta^+\theta\theta$$

if a finite number of terms were used for $F^+_n(p)$, L^+_n, or P^+_n. The expansion to an infinite number of terms for the boson operators avoids this difficulty and, furthermore, assures that the solution of the equations of motion will be solutions to all sectors. We will, of course, only be able to solve for the lowest sectors, the lowest terms of the series.

Consider $\theta^+_{\rm in}(p)$. Examination of Eqs. (2.10) shows that Eqs. (2.10a) and (2.10f) decouple from the rest. In fact, Eq. (2.10a) is completely separate and allows solution for $F^+_1(p)$ by itself. It is obvious that substituting Eq. (2.12) for $F^+_1(p)$ results in an infinite series of

the symbolic form

$$F_1^+(p) \sim \theta^+ + \theta^+\theta^+\theta + \theta^+\theta^+\theta^+\theta\theta + \cdots, \quad (2.13)$$

each term of which must have a total energy ω_p. But in the limit $\lambda \to 0$, we expect $F_1^+(p) = \theta^+(p)$, and all other $F_n^+(p) = 0$. Therefore, as an initial condition at $\lambda = 0$, the value

$$F_1^+(p) = \theta^+(p) \quad (2.14)$$

shall be imposed. Equation (2.10a) is independent of λ, however; therefore, Eq. (2.14) represents $F_1^+(p)$ for all λ.

Note that Eqs. (2.10d), (2.10e), and (2.10f) may be added together to give

$$\omega_p(F_4^+(p) + F_5^+(p) + F_6^+(p)) = [H_\theta, F_4^+(p) + F_5^+(p) + F_6^+(p)]_-. \quad (2.15)$$

Setting initial conditions in a manner similar to the treatment of $F_1^+(p)$ above,

$$F_4^+(p) + F_5^+(p) + F_6^+(p) = 0 \quad (2.16)$$

will be demanded for all λ. It is thus possible to concentrate on $F_2^+(p) \to F_5^+(p)$.

Substituting the general forms for $F_2^+(p) \to F_5^+(p)$ into Eqs. (2.10), using Eq. (2.14), and taking appropriate matrix elements yields for the lowest coefficients

$$f_{2,1}(p) = \frac{\lambda f(\omega_p)}{D^+(\omega_p - M_v + m_N)} \quad (2.17)$$

and

$$f_{5,1}(p; p_1) = \frac{\lambda^2 f(\omega_p) f(\omega_{p_1})}{D^+(\omega_p - M_v + m_N)(\omega_p - \omega_{p_1} + i\epsilon)}, \quad (2.18)$$

where

$$D^+(\omega_p) = \left(\omega_p + M_v - m_v + \lambda^2 \times \int \frac{d\mathbf{q} f^2(\omega_q)}{(\omega_q + m_N - M_v - \omega_p - i\epsilon)}\right) \quad \text{in the limit } \epsilon \to 0. \quad (2.19)$$

In this derivation, $+i\epsilon$ was added to ω_p in the realization that $\theta_{\text{in}}^+(p)$ will be used in the construction of the N–θ scattering state. It is, of course, the same $+i\epsilon$ which is inherent in the retarded Green's function of a Yang–Feldman definition of the in-field.[8] Note that because of Eqs. (2.4) and (2.6), $\theta_{\text{in}}^+(p)N_{\text{in}}^+ |0\rangle$ is the N–θ scattering state. Since the scattering amplitude is proportional to $\langle \theta^+ N^+ \mid \theta_{\text{in}}^+ N_{\text{in}}^+ \rangle$, we expect the amplitude to be contained in the constructed $\theta_{\text{in}}^+(p)$; and, indeed, $f_{5,1}(p; p_1)$ is (aside from kinematical factors) just this amplitude. The $\delta(\omega_p - \omega_{p_1})$ term which would ordinarily appear on the left-hand side of Eq. (2.18) has been dropped. The equations of motion, Eqs. (2.10), do not yield a unique solution for the coefficient of this term; setting it equal to zero is consistent with the normalization

$$\langle N_{\text{in}}^+ \theta_{\text{in}}^+(p) \mid \theta_{\text{in}}^+(k) N_{\text{in}}^+ \rangle = \delta(\mathbf{p} - \mathbf{k}). \quad (2.20)$$

It was not found possible to proceed any further with the formation of $\theta_{\text{in}}^+(p)$; the equations of motion appear too complicated simply to separate and allow solution of higher terms of $F_2^+(p)$, $F_3^+(p)$, $F_4^+(p)$, or $F_5^+(p)$. When the equations are decoupled, extremely involved integral equations appear. But considering

$$\theta_{\text{in}}^+ \sim \theta^+ + V^+N + \theta^+N^+N,$$

it is clear that $\theta_{\text{in}}^+(p)$ is known completely within the N–θ sector.

Now consider the V_{in}^+ field; because Eqs. (2.11a) and (2.11d) decouple from the rest, it is possible to solve for L_1^+ and L_4^+ alone. Substituting the constructions analogous to Eq. (2.12) into these equations, normal ordering, and, where necessary, taking matrix elements, we get

$$M_v - m_v = \lambda^2 \int \frac{f^2(\omega_q)\, d\mathbf{q}}{M_v - m_N - \omega_q} \quad (2.21)$$

and

$$\ell_{4,1}(p) = \frac{\lambda \ell_{1,1} f(\omega_p)}{M_v - m_N - \omega_p}. \quad (2.22)$$

In Eq. (2.22), $\ell_{1,1}$ is arbitrary. An appropriate $\ell_{1,1}$ may be found by requiring

$$\langle 0 | V_{\text{in}} V_{\text{in}}^+ |0\rangle = 1. \quad (2.23)$$

This yields

$$|\ell_{1,1}|^2 = \left(1 + \lambda^2 \int \frac{d\mathbf{q} f^2(\omega_q)}{(M_v - m_N - \omega_q)^2}\right)^{-1}. \quad (2.24)$$

Both the constant $\ell_{1,1}$ of Eq. (2.24) and the eigenvalue equation for M_v, Eq. (2.21), are expressions which would also result from the usual calculation of the physical V-particle state $|V\rangle_p$ by the solution of

$$H |V\rangle_p = M_v |V\rangle_p. \quad (2.25)$$

Demanding normalization of $|V\rangle_p$ above gives the renormalization constant Z, which is just $|\ell_{1,1}|^2$. Indeed, the usual calculation yields

$$V_{\text{in}}^+ |0\rangle = |V\rangle_p. \quad (2.26)$$

This result was to be expected, of course, from Eq. (2.5). Similarly, $\theta_{\text{in}}^+(p) |0\rangle = |\theta^+(p)\rangle$, and, as will appear, $N_{\text{in}}^+ |0\rangle = |N^+\rangle$. In what follows $M_v \equiv m_N$ for convenience.

By taking matrix elements of Eqs. (2.11a) and (2.11d), it is possible to write the general equations for

[8] See Ref. 1.

the nth terms of L_1^+ and L_4^+. These are

$$n\ell_{4,n}(p_1,\cdots,p_n;k_1,\cdots,k_{n-1}) = \lambda(\omega_{k_1}+\cdots+\omega_{k_{n-1}}-\omega_{p_1}-\cdots-\omega_{p_n}+i\epsilon)^{-1}$$
$$\times[\ell_{1,n}(p_1,\cdots,p_{n-1};k_1,\cdots,k_{n-1})f(\omega_{p_n})+\ell_{1,n}(p_1,\cdots,p_{n-2},p_n;k_1,\cdots,k_{n-1})f(\omega_{p_{n-1}}$$
$$+\cdots+\ell_{1,n}(p_2,p_3,\cdots,p_n;k_1,\cdots,k_{n-1})f(\omega_{p_1})]+C(\omega_{k_1},\cdots,\omega_{k_{n-1}};\omega_{p_1},\cdots,\omega_{p_n})$$
$$\times\delta(\omega_{k_1}+\cdots+\omega_{k_{n-1}}-\omega_{p_1}-\cdots-\omega_{p_n}) \quad (2.27)$$

and the integral equation

$$\ell_{1,n}(p_1,\cdots,p_{n-1};k_1,\cdots,k_{n-1}) = [\lambda/(n-1)][D^+(\omega_{k_1}+\cdots+\omega_{k_{n-1}}-\omega_{p_1}-\cdots-\omega_{p_{n-1}})]^{-1}$$
$$\times[\ell_{4,n-1}(p_1,\cdots,p_{n-1};k_1,\cdots,k_{n-2})f(\omega_{k_{n-1}})+\cdots+\ell_{4,n-1}(p_1,\cdots,p_{n-1};k_2,\cdots,k_{n-1})f(\omega_{k_1})]$$
$$-\lambda^2[D^+(\omega_{k_1}+\cdots+\omega_{k_{n-1}}-\omega_{p_1}-\cdots-\omega_{p_{n-1}})]^{-1}$$
$$\times\int\frac{d\mathbf{q}f(\omega_q)[\ell_{1,n}(p_1,\cdots,p_{n-2},q;k_1,\cdots,k_{n-1})f(\omega_{p_{n-1}})+\cdots+\ell_{1,n}(q,p_2,\cdots,p_{n-1};k_1,\cdots,k_{n-1})f(\omega_{p_1})}{(\omega_q+\omega_{p_1}+\cdots+\omega_{p_{n-1}}-\omega_{k_1}-\cdots-\omega_{k_{n-1}}-i\epsilon)}$$
$$+\lambda\int d\mathbf{q}C(\omega_{k_1},\cdots,\omega_{k_{n-1}};\omega_{p_1},\cdots,\omega_{p_{n-1}},\omega_q)\delta(\omega_{k_1}+\cdots+\omega_{k_{n-1}}-\omega_{p_1}-\cdots-\omega_{p_{n-1}}-\omega_q)f(\omega q). \quad (2.28)$$

Here [similar to Eq. (2.18)] $+i\epsilon$ was added to M_v because of the future use of V_{in}^+ to construct the scattering states. The coefficient $C(\omega_{k_1},\cdots,\omega_{k_{n-1}};\omega_{p_1},\cdots,\omega_{p_n})$ may be set equal to zero. For $n=2$ Eq. (2.28) reduces to

$$\ell_{1,2}(p;k) = \frac{-\lambda^2Z^{\frac{1}{2}}f(\omega_p)f(\omega_k)}{\omega_pD^+(\omega_k-\omega_p)}-\frac{\lambda^2f(\omega_p)}{D^+(\omega_k-\omega_p)}$$
$$\times\int\frac{d\mathbf{q}f(\omega_q)\ell_{1,2}(q;k)}{(\omega_q-\omega_k+\omega_p-i\epsilon)}, \quad (2.29)$$

which is exactly that integral equation which arises (by use of the Lippmann–Schwinger equation) for the amplitude for V–θ scattering. While Eq. (2.29) has been solved by several authors,[9-12] we shall not now be concerned with its solution. The present work shall be restricted to the N–θ sector.

A better understanding of Eq. (2.28) is gained by observing that, for $n=3$, the solution represents that part of the V–$\theta\theta$ scattering amplitude which is independent of the V–θ scattering. The appearance of scattering amplitudes in the V_{in}^+ field is not surprising, for such amplitudes are proportional to

$$\langle V^+\theta_1^+\cdots\theta_n^+|\,V_{\text{in}}^+\theta_{\text{in}_1}^+\cdots\theta_{\text{in}_n}^+\rangle \quad (\text{for } V\text{–}n\theta \text{ scattering}).$$

Hence, V_{in}^+ should intrinsically contain scattering amplitudes; we expect that the equations for higher-order coefficients will be just those which would arise from a consideration of higher-sector scattering.

In solving for V_{in}^+ and for N_{in}^+, the symmetry

$$L_1^+=-L_2^+,\quad P_1^+=-P_2^+,$$
$$L_3^+=-L_4^+,\quad P_3^+=-P_4^+ \quad (2.30)$$

is required to all sectors. A detailed proof will be given in the Appendix. The N_{in}^+ field may be solved for in exactly the same manner as V_{in}^+. Here, again, relevant scattering amplitudes appear.

All the in-fields can now be written:

$$\theta_{\text{in}}^+(p) = \theta^+(p)+\frac{\lambda f(\omega_p)V^+N}{D^+(\omega_p)}+\frac{\lambda^2f(\omega_p)}{D^+(\omega_p)}$$
$$\times\int\frac{f(\omega_{p_1})\theta^+(p_1)N^+N}{\omega_p-\omega_{p_1}+i\epsilon}, \quad (2.31)$$

$$V_{\text{in}}^+ = (Z)^{\frac{1}{2}}\Big\{V^+-V^+N^+N+\lambda\int\frac{f(\omega_{p_1})\theta^+(p_1)N^+V^+V}{\omega_{p_1}}$$
$$-\lambda\int\frac{f(\omega_{p_1})\theta^+(p_1)N^+}{\omega_{p_1}}\Big\}, \quad (2.32)$$

$$N_{\text{in}}^+ = \lambda\int\frac{f(\omega_k)\theta(k)V^+}{D^+(\omega_k)}-\lambda\int\frac{f(\omega_k)\theta(k)V^+N^+N}{D^+(\omega_k)}$$
$$-\left(1+\lambda^2\int\frac{f(\omega_p)f(\omega_k)\theta^+(p)\theta(k)}{D^+(\omega_k)(\omega_k-\omega_p+i\epsilon)}\right)N^+V^+V$$
$$+\left(1+\lambda^2\int\frac{f(\omega_p)f(\omega_k)\theta^+(p)\theta(k)}{D^+(\omega_k)(\omega_k-\omega_p+i\epsilon)}\right)N^+, \quad (2.33)$$

where repeated variables are integrated over.

These fields are all known to the N–θ sector; indeed, because

$$V_{\text{in}}^+ = L_1^+V^+-L_1^+V^+N^+N-L_4^+N^+V^+V+L_4^+N^+,$$
$$N_{\text{in}}^+ = P_1^+V^+-P_1^+V^+N^+N-P_4^+N^+V^+V+P_4^+N^+, \quad (2.34)$$

[9] R. P. Kenschaft and R. D. Amado, J. Math. Phys. **5**, 1340 (1964).
[10] A. Pagnamenta, J. Math. Phys. **6**, 995 (1965).
[11] C. M. Sommerfield, J. Math. Phys. **6**, 1170 (1965).
[12] E. Kazes, J. Math. Phys. **6**, 1772 (1965).

the two central terms of V_{in}^+ and N_{in}^+ are known to even higher sectors. This form is demanded by the equations of motion.

One surprising result of Eq. (2.34) is the inability of these in-fields to create a state with the $|N^+V^+\rangle$ quantum numbers. In fact, from Eq. (2.34),

$$N_{\text{in}}^+ V_{\text{in}}^+ = V_{\text{in}}^+ N_{\text{in}}^+ = 0. \tag{2.35}$$

And such a state may not be created by the combination $\theta_{\text{in}}^+ N_{\text{in}}^+ N_{\text{in}}^+$, since $N_{\text{in}}^+ N_{\text{in}}^+ = 0$. Thus the operators N_{in}^+, V_{in}^+, and θ_{in}^+ do not create a complete set of states.

What has been overlooked is the fact that N^+V^+ itself is an eigenfield satisfying

$$[H, N^+V^+]_- = (m_N + m_v)N^+V^+. \tag{2.36}$$

This in-field may, of course, be multiplied by any suitable quantum-numberless combination of θ^+ and θ operators. The resulting in-field

$$(N^+V^+)_{\text{in}} \equiv RN^+V^+ \tag{2.37}$$

would obey

$$[H, RN^+V^+] = MRN^+V^+, \tag{2.38}$$

where R is a boson operator with the quantum numbers of the vacuum and M is an energy. Investigation of Eq. (2.38) shows that the equation of motion for R is independent of λ. Thus, similar to the discussion concerning Eq. (2.14), R may be set equal to 1 for all λ; with this choice, Eq. (2.36) holds.

It may be noted that Eq. (2.35) should follow independently of the equations of motion or of the explicit form either of the Hamiltonian or of V_{in}^+ or N_{in}^+. This is because, regardless of the properties of any other field, N^+V^+ exists as an in-field [obeying Eq. (2.36)] for the Lee model. Therefore, the state $|N^+V^+\rangle$ is a physical state with an energy of $m_N + m_v$. Were $N_{\text{in}}^+ V_{\text{in}}^+$ (or $V_{\text{in}}^+ N_{\text{in}}^+$) nonzero, then the state $|N_{\text{in}}^+ V_{\text{in}}^+\rangle$ (or $|V_{\text{in}}^+ N_{\text{in}}^+\rangle$) would also exist, but with an energy $m_N + M_v$. In the limit of $\lambda \to 0$, of course, only the first state would be present. There seems to be no physical reason why this state should, upon the introduction of an interaction, split into two states of the same quantum numbers but different energies. The impossibility, then, of having two such states is reflected in Eq. (2.35). Furthermore, the state $|N_{\text{in}}^+ V_{\text{in}}^+\rangle$ could be expected not to exist, since its energy would be just the sum of the energies of the two physical particles which make it up, implying the absence of an interaction energy.

Now it is possible to express the Lee-model Hamiltonian in terms of in-fields. Using Eqs. (2.24), (2.19), (2.31), (2.32), and (2.33), along with the dispersion relation for $D(z)$,

$$\frac{1}{D(z)} = \frac{Z}{z} - \lambda^2 \int \frac{d\mathbf{q} f^2(\omega_q)}{|D^+(\omega_q)|^2 (\omega_q - z)}, \tag{2.39}$$

yields exactly

$$H = \int d\mathbf{k} \omega_k \theta_{\text{in}}^+(k)\theta_{\text{in}}(k) + m_N N_{\text{in}}^+ N_{\text{in}} + M_v V_{\text{in}}^+ V_{\text{in}} + (m_N + m_v)N^+NV^+V, \tag{2.40}$$

where the last term is the energy of the N^+V^+ field. The scattering states of this Hamiltonian are obvious.

Consider, now, the algebra of the in-fields. Because these fields are not known to all orders, the commutation (or anticommutation) relations are not expected to be known exactly. Nor, in general, will they be known up to even the N–θ sector. This can be seen, for example, with the aid of the commutator

$$[\theta_{\text{in}}(k), \theta_{\text{in}}^+(p)]_- = \theta_{\text{in}}(k)\theta_{\text{in}}^+(p) - \theta_{\text{in}}^+(p)\theta_{\text{in}}(k). \tag{2.41}$$

We may insert complete sets of states

$$[\theta_{\text{in}}(k), \theta_{\text{in}}^+(p)]_- = \sum_m \theta_{\text{in}}(k)\,|m\rangle\langle m|\,\theta_{\text{in}}^+(p) - \sum_m \theta_{\text{in}}^+(p)\,|m\rangle\langle m|\,\theta_{\text{in}}(k). \tag{2.42}$$

Considering a matrix element of the commutator, then, is equivalent to considering a matrix element of $\theta_{\text{in}}^+(p)$ [or $\theta_{\text{in}}(k)$] between $|m\rangle$ and some other state; if $|m\rangle$ belongs to a higher sector than N–θ, the original matrix element of the commutator cannot be determined from the solutions given in Eqs. (2.31), (2.32), and (2.33). Thus, the choice of commutator matrix elements which should be correct is narrowed extremely: only those elements which do not connect to higher sectors may be considered.

For the commutator of Eq. (2.42) this restriction allows only the $\langle 0|\,|0\rangle$ and $\langle N|\,|N\rangle$ elements to be known exactly. Using Eqs. (2.31) and (2.39) results in

$$[\theta_{\text{in}}(k), \theta_{\text{in}}^+(p)]_- = \delta(\mathbf{p} - \mathbf{k}) + (\text{incomplete higher sector terms}). \tag{2.43}$$

Therefore, to the lowest sector, the $\theta_{\text{in}}^+(p)$ field obeys boson commutation relations.

The commutator of $\theta_{\text{in}}^+(p)$ with other in-fields may be investigated. Consider

$$[N^+V^+, \theta_{\text{in}}^+(p)]_- = -[F_4^+(p) + F_5^+(p) + F_6^+(p)]N^+V^+ = 0 \tag{2.44}$$

to all orders in the θ-particle operators; this follows from Eqs. (2.7) and (2.16). And similarly,

$$[VN, \theta_{\text{in}}^+(p)]_- = [F_4^+(p) + F_5^+(p) + F_6^+(p)]VN = 0 \tag{2.45}$$

to all orders.

Considering the lowest-order terms for which it is possible to know the matrix elements, it is found that

$$[N_{\text{in}}^{+}, \theta_{\text{in}}^{+}(p)]_{-} = 0, \tag{2.46}$$

$$[N_{\text{in}}, \theta_{\text{in}}^{+}(p)]_{-} = 0, \tag{2.47}$$

and

$$[V_{\text{in}}, \theta_{\text{in}}^{+}(p)]_{-} = 0. \tag{2.48}$$

$[V_{\text{in}}^{+}, \theta_{\text{in}}^{+}(p)]_{-}$ has no matrix elements which lie in a low enough sector to be known.

The algebra for the fermion fields is found to be

$$\{V_{\text{in}}, V_{\text{in}}^{+}\}_{+} = 1 - N_{\text{in}}^{+}N_{\text{in}} - N^{+}NV^{+}V, \tag{2.49}$$

$$\{N_{\text{in}}, N_{\text{in}}^{+}\}_{+} = 1 - V_{\text{in}}^{+}V_{\text{in}} - N^{+}NV^{+}V, \tag{2.50}$$

and

$$\{VN, N^{+}V^{+}\}_{+} = 1 - N_{\text{in}}^{+}N_{\text{in}} - V_{\text{in}}^{+}V_{\text{in}} \tag{2.51}$$

to the lowest correct orders. These anticommutation relations are not those of free-fermion fields. The physical and formal reasons for this algebra will be discussed in Sec. 3.

3. ASYMPTOTIC FIELD ALGEBRA

In this section we shall consider more closely the unusual anticommutation relations exhibited by the fermion asymptotic fields of the Lee model [Eqs. (2.49), (2.50), and (2.51)]. In order to do so, the general properties of these fields will first be examined.

The fermion in-fields may be said to obey a *general Pauli exclusion principle* (GPEP) to all sectors. In the present context this statement means that any two creation operators, for the same or different in-fields, when multiplied together yield zero:

$$\left.\begin{array}{llll} \text{(a)} & N_{\text{in}}^{+}N_{\text{in}}^{+} = 0, & \text{(f)} & N^{+}V^{+}V_{\text{in}}^{+} = 0, \\ \text{(b)} & V_{\text{in}}^{+}V_{\text{in}}^{+} = 0, & \text{(g)} & V_{\text{in}}^{+}N^{+}V^{+} = 0, \\ \text{(c)} & N^{+}V^{+}N^{+}V^{+} = 0, & \text{(h)} & N^{+}V^{+}N_{\text{in}}^{+} = 0, \\ \text{(d)} & N_{\text{in}}^{+}V_{\text{in}}^{+} = 0, & \text{(i)} & N_{\text{in}}^{+}N^{+}V^{+} = 0. \\ \text{(e)} & V_{\text{in}}^{+}N_{\text{in}}^{+} = 0, & & \end{array}\right\} \tag{3.1}$$

This result follows from Eqs. (2.2) and (2.34).

The fermion in-fields are also subject to an *orthogonality* condition to all sectors. This will mean that the state created by any fermion in-field of the Lee model is orthogonal to a state created by any other fermion in-field. This condition is expressed by the equations

$$\left.\begin{array}{llll} \text{(a)} & V_{\text{in}}N_{\text{in}}^{+} = 0, & \text{(d)} & N_{\text{in}}V_{\text{in}}^{+} = 0, \\ \text{(b)} & VNV_{\text{in}}^{+} = 0, & \text{(e)} & V_{\text{in}}N^{+}V^{+} = 0, \\ \text{(c)} & VNN_{\text{in}}^{+} = 0, & \text{(f)} & N_{\text{in}}N^{+}V^{+} = 0, \end{array}\right\} \tag{3.2}$$

where Eqs. (3.2d)–(3.2f) are the Hermitian conjugates of Eqs. (3.2a)–(3.2c) and have been written here for completeness. The proof of Eqs. (3.2) is given in the Appendix to all sectors; it will later be shown (in the proof of Theorem I in this section) that the GPEP follows as a consequence of orthogonality alone and is actually independent of Eqs. (2.34).

The further consequences of the GPEP and orthogonality are

$$\begin{array}{llll} \text{(a)} & \{V_{\text{in}}^{+}, V_{\text{in}}^{+}\}_{+} = 0, & \text{(d)} & \{V_{\text{in}}^{+}, N_{\text{in}}^{+}\}_{+} = 0, \\ \text{(b)} & \{N_{\text{in}}^{+}, N_{\text{in}}^{+}\}_{+} = 0, & \text{(e)} & \{V_{\text{in}}^{+}, N^{+}V^{+}\}_{+} = 0, \\ \text{(c)} & \{N^{+}V^{+}, N^{+}V^{+}\}_{+} = 0, & \text{(f)} & \{N_{\text{in}}^{+}, N^{+}V^{+}\}_{+} = 0, \end{array} \tag{3.3}$$

and

$$\{V_{\text{in}}, N_{\text{in}}^{+}\}_{+} = N_{\text{in}}^{+}V_{\text{in}} \neq 0, \tag{3.4a}$$

$$\{V_{\text{in}}, N^{+}V^{+}\}_{+} = N^{+}V^{+}V_{\text{in}} \neq 0, \tag{3.4b}$$

$$\{N_{\text{in}}, N^{+}V^{+}\}_{+} = N^{+}V^{+}N_{\text{in}} \neq 0. \tag{3.4c}$$

Using the orthogonality conditions, the anticommutation relations, Eqs. (2.49), (2.50), and (2.51), may be written, respectively, as

$$\{V_{\text{in}}, V_{\text{in}}^{+}\}_{+} = (1 - N_{\text{in}}^{+}N_{\text{in}})(1 - N^{+}NV^{+}V), \tag{3.5}$$

$$\{N_{\text{in}}, N_{\text{in}}^{+}\}_{+} = (1 - V_{\text{in}}^{+}V_{\text{in}})(1 - N^{+}NV^{+}V), \tag{3.6}$$

$$\{VN, N^{+}V^{+}\}_{+} = (1 - N_{\text{in}}^{+}N_{\text{in}})(1 - V_{\text{in}}^{+}V_{\text{in}}). \tag{3.7}$$

And these equations yield

$$V_{\text{in}}V_{\text{in}}^{+} = N_{\text{in}}N_{\text{in}}^{+} = VNN^{+}V^{+}. \tag{3.8}$$

The fermion-asymptotic fields of the Lee model, then, because of the GPEP, are fields which create individual states; no two of them may be combined to form a third state. These fields, furthermore, can be used to express the original Hamiltonian in a diagonal form, Eq. (2.40). And they do not have a free-fermion field algebra.

Some insight into the origin of these properties may be gained by considering the general diagonal representation for any Hermitian Hamiltonian:

$$H = \sum_{m} E_m |m\rangle\langle m|, \tag{3.9}$$

where the sum is over all orthonormal eigenstates of the Hamiltonian, E_m being their energies.

It is clear that a solution to the eigenvalue equation

$$[H, A_i^{+}]_{-} = E_i A_i^{+}, \quad i = 1, 2, 3, \cdots, \tag{3.10}$$

is the operator

$$A_i^{+} \equiv |i\rangle\langle 0|, \quad A_i \equiv |0\rangle\langle i|, \tag{3.11}$$

where $|0\rangle$ is the ground state (zero-energy state) of H.

Other operators exist which satisfy Eq. (3.10) (e.g., $|i\rangle\langle m|$, if we take $E_i' \equiv E_i - E_m$), but the A_i^{+} of Eq. (3.11) is one which will not destroy $|0\rangle$. A_i^{+} can thus be

called an eigenfield. The set of all A^+ is complete in that it can create a complete set of states from the ground state $|0\rangle$. In what follows $i \neq 0$; the zero-state projection operator $|0\rangle\langle 0|$ will be discussed in detail later.

Because the eigenstates are orthonormal, the eigenfields obey orthogonality as well as a general Pauli exclusion principle:

$$A_i A_j^+ = \delta_{ij} |0\rangle\langle 0| \tag{3.12}$$

and

$$A_i^+ A_j^+ = 0, \quad \text{for every } i, j. \tag{3.13}$$

Furthermore, it is apparent that

$$H = \sum_m E_m A_m^+ A_m. \tag{3.14}$$

There is a connection between orthogonality and the general Pauli exclusion principle for all operators, such as A_i^+, regardless of their representations. As will later become apparent, the in-fields V_{in}^+, N_{in}^+, and N^+V^+ correspond to A_i^+; the connection between Eqs. (3.2) and (3.1) will be made clear in the proof of the following theorem.

Theorem 1: If a set of creation operators B^+ exists and is complete in that it can create a complete set of states of different quantum numbers (from the ground state $|0\rangle$), then the existence of a GPEP between the operators implies orthogonality, and conversely.

Proof: (a) *GPEP implies orthogonality.* This statement may be proved by a consideration of the operator $B_i B_j^+$ $(i \neq j)$. It is obvious, because of the GPEP, that this operator has no nonzero matrix elements between any states $\langle k|\ |m\rangle$, where $|m\rangle \equiv B_m^+ |0\rangle$. The element $\langle 0|\ B_i B_j^+\ |0\rangle$ is zero because all operators create states of different quantum numbers. The set of states $|m\rangle$ is complete; therefore

$$B_i B_j^+ = 0 \quad \text{for all} \quad i, j, i \neq j.$$

(b) *Orthogonality implies a GPEP.* The proof is similar to that above. Because of orthogonality and the fact that $\langle 0|\ B_i^+ = 0$, the operator $B_i^+ B_j^+$ has no nonzero matrix elements except possibly $\langle i|\ B_i^+ B_j^+\ |m\rangle$. This last matrix element is zero from a consideration of quantum numbers. Therefore

$$B_i^+ B_j^+ = 0$$

for every i, j. Q.E.D.

Note that the presence of $\theta_{\text{in}}^+(p)$ makes the Lee-model fermion-asymptotic fields an incomplete set. The first part of this theorem, then, does not apply, but the second part can be used, and the statement made previously is proved: orthogonality does imply the GPEP for the fermion-asymptotic fields of the Lee model.

It is possible to show the connection between the GPEP (and orthogonality) and the algebra of the operators B^+.

Theorem 2: If a set of creation operators B^+ exists and is complete in that it can create a complete set of orthonormal states of different quantum numbers (from $|0\rangle$) and obeys a GPEP (or, equivalently, orthogonality), then it has the algebra

$$\{B_n, B_n^+\}_+ = 1 - \sum_{m \neq n} B_m^+ B_m = \prod_{m \neq n} (1 - B_m^+ B_m). \tag{3.15}$$

(No summation is intended on left-hand side.)

Proof: This theorem may be proved by expanding the operator $B_n B_n^+$ in a complete set of states. A shorter method consists of simply noticing that, because of the GPEP, the only nonzero element in the matrix representation of $B_n B_n^+$ is

$$\langle 0|\ B_n B_n^+\ |0\rangle.$$

If the set B^+ creates orthonormal states, then

(a) $\langle 0|\ B_n B_n^+\ |0\rangle = 1$ for any n.

There is another operator with this matrix representation and that is $|0\rangle\langle 0|$. Thus the two operators are identical. Now

$$\text{(b)}\ |0\rangle\langle 0| = 1 - \sum_m |m\rangle\langle m| = 1 - \sum_m B_m^+ |0\rangle\langle 0|\ B_m,$$

where the sum is over all states but $|0\rangle$. The equation above is an integral equation for $|0\rangle\langle 0|$, the first iteration of which yields a solution:

$$\text{(c)}\ |0\rangle\langle 0| = 1 - \sum_m B_m^+ \left(1 - \sum_n |n\rangle\langle n|\right) B_m = 1 - \sum_m B_m^+ B_m$$

by the GPEP. Therefore,

$$\text{(d)}\ B_n B_n^+ = 1 - \sum_m B_m^+ B_m = \prod_m (1 - B_m^+ B_m),$$

where the second equality results from orthogonality. Thus

$$\text{(e)}\ \{B_n, B_n^+\}_+ = 1 - \sum_{m \neq n} B_m^+ B_m = \prod_{m \neq n} (1 - B_m^+ B_m).$$

Q.E.D.

It is now possible to show that any set of operators B^+ which creates a complete set of states $(B_i^+ |0\rangle = |i\rangle)$ and which obeys a GPEP $(B_i^+ B_n^+ = 0)$ is identical

to the set A^+. Consider

$$B_i^+ = B_i^+ \sum_{n=0} |n\rangle\langle n| = B_i^+ |0\rangle\langle 0| = |i\rangle\langle 0| = A_i^+,$$

where the sum is over all states including the ground state.

Similarly, one can show that any set of operators B^+ which obeys orthogonality and which creates orthonormal states is identical to the set A^+. Consider

$$B_i^+ = \sum_{n=0} |n\rangle\langle n| B_i^+ = \sum_{n=0} |n\rangle\langle 0| B_n B_i^+ = |i\rangle\langle 0| B_i B_i^+,$$

where the sum is over all states including the ground state. Therefore, by Theorem 2,

$$B_i^+ = |i\rangle\langle 0| \left(1 - \sum_m B_m^+ B_m\right) = |i\rangle\langle 0| = A_i^+.$$

Thus the operators A_i^+ obey the field algebra of Theorem 2 (and of the Lee model). This, of course, can be proved by direct computation using Eq. (3.11). Note that everything which has been said about A_i^+ applies for any Hermitian Hamiltonian, irrespective of the question of second quantization (it can apply to the hydrogen atom, for example). Operators have been found which have a definite asymptotic field algebra and which diagonalize the Hamiltonian.

To relate the eigenfields to field-theoretical concepts, consider the harmonic oscillator

$$H = \hbar\omega a^+ a, \tag{3.16}$$

where

$$[a, a^+]_- = 1. \tag{3.17}$$

Now

$$A_i^+ = \frac{(a^+)^i}{(i!)^{\frac{1}{2}}} |0\rangle\langle 0| = |i\rangle\langle 0|. \tag{3.18}$$

The operator $|0\rangle\langle 0|$, which destroys every state but $|0\rangle$, may be expressed in terms of a^+ and a. The result, which is proved in the Appendix, is

$$|0\rangle\langle 0| = 1 + \sum_{j=1}^{\infty} (-1)^j (a^+)^j (a)^j / j!. \tag{3.19}$$

And, therefore,

$$A_i^+ = \frac{(a^+)^i}{(i!)^{\frac{1}{2}}} \left(1 + \sum_{j=1}^{\infty} (-1)^j (a^+)^j (a)^j / j!\right). \tag{3.20}$$

Conversely, the operator a^+ may be expressed in terms of A_i^+. Since

$$a^+ = \sum_{n=0}^{\infty} |n+1\rangle\langle n| (n+1)^{\frac{1}{2}}, \tag{3.21}$$

then

$$a^+ = A_1^+ + \sum_{n=1}^{\infty} A_{n+1}^+ A_n (n+1)^{\frac{1}{2}}. \tag{3.22}$$

The two expressions, Eqs. (3.20) and (3.22), can be shown to each obey the correct field algebra [Eqs. (3.15) and (3.17)]. It is apparent, then, that boson operators as well as fermion can have the anticommutator algebra of Eq. (3.15). The Hamiltonian of Eq. (3.16) can be recast into either of the two diagonal forms:

$$H = \sum_{n=1} n\hbar\omega A_n^+ A_n = \hbar\omega a^+ a. \tag{3.23}$$

Thus there are clearly two types of eigenfields:

$$[H, (a^+)^n]_- = n\hbar\omega (a^+)^n \tag{3.24}$$

and

$$[H, A_n^+]_- = n\hbar\omega A_n^+. \tag{3.25}$$

Before presenting the analogous problems in the Lee model, a discussion of the simpler considerations arising for a model Hamiltonian with two bosons seems appropriate. Consider the Hamiltonian

$$H = m_1 a^+ a + m_2 b^+ b, \tag{3.26}$$

where

$$[a, a^+]_- = 1 = [b, b^+]_-,$$
$$[a, b^+]_- = [b, a^+]_- = 0. \tag{3.27}$$

The results are

$$|0\rangle\langle 0| = \left(1 + \sum_{j=1}^{\infty} \frac{(-1)^j (a^+)^j (a)^j}{j!}\right) \times \left(1 + \sum_{m=1}^{\infty} \frac{(-1)^m (b^+)^m (b)^m}{m!}\right) \tag{3.28}$$

and

$$A_{r,s}^+ = \frac{(a^+)^r (b^+)^s}{(r!)^{\frac{1}{2}} (s!)^{\frac{1}{2}}} |0\rangle\langle 0|, \tag{3.29}$$

where r and s are not both zero. Also

$$a^+ = \sum_{\substack{r=0 \\ s=0}}^{\infty} A_{r+1,s}^+ A_{r,s} (r+1)^{\frac{1}{2}}, \tag{3.30}$$

where, for the $r = s = 0$ term, $A_{1,0}^+$ is meant, and

$$b^+ = \sum_{\substack{r=0 \\ s=0}}^{\infty} A_{r,s+1}^+ A_{r,s} (s+1)^{\frac{1}{2}}, \tag{3.31}$$

where, for the $r = s = 0$ term, $A_{0,1}^+$ is meant. The Hamiltonian may be written in the diagonal form

$$H = \sum_{\substack{r=0, s=0 \\ (r=s\neq 0)}} (r m_1 + s m_2) A_{r,s}^+ A_{r,s}. \tag{3.32}$$

Carrying out this program for the Lee-model Hamiltonian in the absence of interactions and for a static boson

$$H = m_N N^+ N + m_V V^+ V + \hbar\omega a^+ a, \tag{3.33}$$

with the usual commutation and anticommutation

rules, results in

$$|0\rangle\langle 0| = (1 - N^+N)(1 - V^+V)\left(1 + \sum_{j=1}^{\infty} \frac{(-1)^j (a^+)^j (a)^j}{j!}\right). \tag{3.34}$$

Labeling the eigenfields A_i^+ in accordance with their quantum numbers and indicating the numbers of bosons in a field by the subscript n, we have

$$\begin{aligned}
V_{\text{in},n}^+ &\equiv \frac{(a^+)^n}{(n!)^{\frac{1}{2}}} V^+ |0\rangle\langle 0| \\
&= \frac{(a^+)^n}{(n!)^{\frac{1}{2}}} (V^+ - V^+N^+N) \\
&\quad \times \left(1 + \sum_{j=1} \frac{(-1)^j (a^+)^j (a)^j}{j!}\right), \\
N_{\text{in},n} &\equiv \frac{(a^+)^n}{(n!)^{\frac{1}{2}}} N^+ |0\rangle\langle 0| \\
&= \frac{(a^+)^n}{(n!)^{\frac{1}{2}}} (N^+ - N^+V^+V) \\
&\quad \times \left(1 + \sum_{j=1} \frac{(-1)^j (a^+)^j (a)^j}{j!}\right), \\
(N^+V^+)_{\text{in},n} &\equiv \frac{(a^+)^n}{(n!)^{\frac{1}{2}}} N^+V^+ |0\rangle\langle 0| \\
&= \frac{(a^+)^n}{(n!)^{\frac{1}{2}}} N^+V^+ \left(1 + \sum_{j=1} \frac{(-1)^j (a^+)^j (a)^j}{j!}\right), \\
\theta_{\text{in},n}^+ &\equiv \frac{(a^+)}{(n!)^{\frac{1}{2}}} |0\rangle\langle 0| \\
&= \frac{(a^+)^n}{(n!)^{\frac{1}{2}}} \left(1 + \sum_{j=1} \frac{(-1)^j (a^+)^j (a)^j}{j!}\right) \\
&\quad \times (1 - N^+N)(1 - V^+V),
\end{aligned} \tag{3.35}$$

from Eqs. (3.34) and (2.2). We may express the bare particle operators in terms of the above eigenfields; for example, similar to Eq. (3.21),

$$N^+ = \sum_{n=0} N_{\text{in},n}^+ \theta_{\text{in},n} + \sum_{n=0} (N^+V^+)_{\text{in},n} V_{\text{in},n} \tag{3.36}$$

or, rewriting the term at $n = 0$,

$$N^+ = N_{\text{in}}^+ + (N^+V^+)_{\text{in}} V_{\text{in}} + \sum_{n=1} N_{\text{in},n}^+ \theta_{\text{in},n} + \sum_{n=1} (N^+V^+)_{\text{in},n} V_{\text{in},n}, \tag{3.37}$$

etc.

It is possible to consider the case of nonstatic bosons for the Lee model without interaction. In that case the result, which is proved in the Appendix, is

$$\begin{aligned}
|0\rangle\langle 0| &= (1 - N^+N)(1 - V^+V) \\
&\quad \times \left[1 + \sum_{j=1}^{\infty} \frac{(-1)^j}{j!} \int dp_1 \cdots dp_j \right. \\
&\quad \left. \times \theta^+(p_1) \cdots \theta^+(p_j) \theta(p_1) \cdots \theta(p_j)\right].
\end{aligned} \tag{3.38}$$

Note that this equation is expected to hold even in the presence of interactions, for the state $|0\rangle$ is both the physical and the bare vacuum state.

We now have the explanation for the algebra of the fermion-asymptotic fields of the Lee model. In the absence of the interaction there are two eigenfields: the elementary particle field operators and the fields corresponding to A_i^+. If either orthogonality or a general Pauli exclusion principle is to be satisfied by the in-fields, only A_i^+ will do; then an algebra such as the one found for the full Lee model results.

For the full Lee model with interaction, the double choice is not present: the fermion in-fields can correspond only to A_i^+ and not to the bare particle operators. To see this better, consider $\lambda \to 0$ in Eqs. (2.32), (2.33), and N^+V^+:

$$\lim_{\lambda \to 0} V_{\text{in}}^+ = V^+ - V^+N^+N = V^+(1 - N^+N)(1 - V^+V), \tag{3.39}$$

$$\lim_{\lambda \to 0} N_{\text{in}}^+ = N^+ - N^+V^+V = N^+(1 - N^+N)(1 - V^+V), \tag{3.40}$$

$$\lim_{\lambda \to 0} N^+V^+ = N^+V^+ = N^+V^+(1 - N^+N)(1 - V^+V). \tag{3.41}$$

The right-hand sides of these equations contain the fermion part of the $|0\rangle\langle 0|$ operator; thus, aside from boson terms, the in-fields above are the A_i^+ of Eq. (3.11).

The fact that the asymptotic fields do not go to the free-particle operators in the limit of $\lambda \to 0$ is perhaps surprising, but it is not a condition unique to the Lee model. Consider, for instance, the model Hamiltonian

$$H = H_0 + H_I, \tag{3.42}$$

$$H_0 \equiv mU^+U + mW^+W, \tag{3.43}$$

$$H_I \equiv \lambda_1 U^+W + \lambda_1 W^+U + \lambda_2 W^+WU^+U, \tag{3.44}$$

where U^+ and W^+ are fermion fields of the same quantum number, obeying

$$\{U, U^+\}_+ = \{W, W^+\}_+ = +1 \tag{3.45}$$

and

$$\begin{aligned}
\{U^+, U^+\}_+ &= \{W^+, W^+\}_+ = 0 \\
&= \{U^+, W^+\}_+ = \{U, W^+\}_+.
\end{aligned} \tag{3.46}$$

The eigenfields for this model, Ω_n^+, may be found by considering

$$[H, \Omega_n^+]_- = M_n \Omega_n^+, \quad n = 1, 2, 3. \tag{3.47}$$

If the Ω_n^+ are constructed in the same manner as the

Lee-model in-fields,

$$\Omega_n^+ \equiv R_{1,n}^+ U^+ + R_{2,n}^+ U^+W^+W + R_{3,n}^+ W^+U^+U + R_{4,n}^+ W^+, \quad n = 1, 2 \ (R_{i,n}^+ \text{ are } c \text{ numbers}), \tag{3.48a}$$

$$\Omega_3^+ = U^+W^+, \tag{3.48b}$$

then

$$\Omega_1^+ = (2)^{-\frac{1}{2}}(U^+ - U^+W^+W - W^+U^+U + W^+), \tag{3.49a}$$

$$\Omega_2^+ = (2)^{-\frac{1}{2}}(U^+ - U^+W^+W + W^+U^+U - W^+) \tag{3.49b}$$

results. The eigenfields of Eq. (3.48a) have single-fermion quantum numbers, while Ω_3^+ can create the two-fermion state. The eigenstates of H are $(2)^{-\frac{1}{2}} \times (U^+ \pm W^+)|0\rangle$ and $U^+W^+|0\rangle$. It is apparent, since $|0\rangle\langle 0| = (1 - U^+U)(1 - W^+W)$, that

$$\Omega_1^+ = (2)^{-\frac{1}{2}}(U^+ + W^+)|0\rangle\langle 0|, \tag{3.50a}$$

$$\Omega_2^+ = (2)^{-\frac{1}{2}}(U^+ - W^+)|0\rangle\langle 0|, \tag{3.40b}$$

$$\Omega_3^+ = U^+W^+|0\rangle\langle 0|, \tag{3.50c}$$

where the $(2)^{\frac{1}{2}}$ is from normalization. The eigenvalues are

$$M_1 = m + \lambda_1, \tag{3.51a}$$

$$M_2 = m - \lambda_1, \tag{3.51b}$$

$$M_3 = 2m + \lambda_2. \tag{3.51c}$$

The Ω_n^+ obey orthogonality and a GPEP. They have the algebra of the Lee-model asymptotic fermion fields. And the Hamiltonian of Eq. (3.42) can be written in the diagonal form

$$H = \sum_{n=1}^{n=3} M_n \Omega_n^+ \Omega_n. \tag{3.52}$$

In the limit of $\lambda_1 \to 0$, $\lambda_2 \to 0$ the free-particle fields are not retrieved; the fields Ω_n^+ remain the same. Only the eigenvalues change; all the algebraic conditions are unaltered. Therefore, this simple model, Eq. (3.42), repeats what had been found for the Lee model; the eigenfields Ω_n^+ are the A_i^+ of Eq. (3.11).

It must be noted that the fields N_{in}^+, V_{in}^+, and N^+V^+ of Sec. 2 are not exactly equivalent to A_i^+, except in the fermion subspace. There are several reasons for this. One is that, in solving the equations of motion, Eqs. (2.11), for L_n^+ or P_n^+, only those solutions which had no δ functions of energy were sought [see the discussion of Eqs. (2.27) and (2.28)]. Since the boson terms of Eq. (3.38) correspond to such δ function terms, this would remove the boson subspace from the $|0\rangle\langle 0|$ part of V_{in}^+ and N_{in}^+. Another reason is that boson expressions were definitely disregarded in setting the initial conditions for the N^+V^+ asymptotic field [see the discussion after Eq. (2.38)]. Thus, the fermion in-fields are independent of the boson terms of $|0\rangle\langle 0|$.

Similar reasons account for the failure of $\theta_{\text{in}}^+(p)$ to exhibit orthogonality or a GPEP with the other asymptotic fields. The initial conditions for $\theta_{\text{in}}^+(p)$ [Eqs. (2.14) and (2.16)] were specified in such a way as to leave off the boson summation term in the expression for $|0\rangle\langle 0|$. This automatically removes the boson subspace. The fermion subspace was deleted by dropping δ functions of energy in the solution for $f_5(p; p_1)$, the coefficient of N^+N [see the discussion concerning Eqs. (2.18) and (2.20)]. Hence, that fermion part of $|0\rangle\langle 0|$ which lies within the N–θ sector is missing from $\theta_{\text{in}}^+(p)$. The conditions chosen were, therefore, just those necessary to make the boson field operator behave as an elementary-particle operator.

4. CONCLUSIONS

The Lee-model asymptotic fields were constructed by an expansion in terms of bare particle operators. This expansion led to equations of motion [Eqs. (2.10) and (2.11)] which have certain novel features: the eigenvalue equation, Eq. (2.21), appears immediately as do the integral equations for scattering processes. These result without considering the Lippmann–Schwinger or Low equations and therefore represent new derivations of the scattering amplitudes. Also, the V_{in}^+ and N_{in}^+ fields contain terms (V^+N^+N and N^+V^+V) which have no matrix elements in the N–θ sector, but which nevertheless can be solved for. The V_{in}^+ field thus agrees with that found by Ezawa[2] (except for normalization) in the N–θ sector; Ezawa did not find the V^+N^+N or N^+V^+V, terms which are vital.

Perhaps the most interesting result of this study was the discovery that the fermion in-fields of the Lee model do not obey a free-fermion algebra. These in-fields represent operators which create states, not particles, and they obey a general Pauli exclusion principle and orthogonality. As such, they correspond to the operator A_i^+ of Eq. (3.11). Therefore, they have the algebra of Eq. (3.15) and, in the limit of $\lambda \to 0$, they do not go into the free-particle operators. These eigenfields do diagonalize the Lee-model Hamiltonian, however, as shown by Eqs. (2.40) and (3.14). For the Lee model this diagonalization is complete, and Eq. (2.40) is exactly equal to the Hamiltonian of Eq. (2.1), even though the in-fields are known only up to the N–θ sector. The discussion and examples of Sec. 3 shows that these results are not peculiar to the Lee model.

It was found that the $\theta_{\text{in}}^{+}(p)$ fields of the Lee model exhibit a freedom which the fermion fields do not have. It was possible to choose initial conditions such that the boson in-field has neither the fermion nor the boson terms of $|0\rangle\langle 0|$. The $\theta_{\text{in}}^{+}(p)$ is thus not the A_i^{+} of Eq. (3.11); it obeys canonical commutation rules and is not subject to a general Pauli exclusion principle. The states

$$|\theta_{\text{in}}^{+}(p_1), \theta_{\text{in}}^{+}(p_2)\rangle, |\theta_{\text{in}}^{+}(p_1), \theta_{\text{in}}^{+}(p_2), \theta_{\text{in}}^{+}(p_3)\rangle, \cdots,$$
$$|\theta_{\text{in}}^{+}(p_1), \theta_{\text{in}}^{+}(p_2), V_{\text{in}}^{+}\rangle,$$
$$|\theta_{\text{in}}^{+}(p_1), \theta_{\text{in}}^{+}(p_2), \theta_{\text{in}}^{+}(p_3), V_{\text{in}}^{+}\rangle, \cdots,$$

etc., may therefore be created by successive application of the one asymptotic field. A similar freedom was found for the boson terms of the $|0\rangle\langle 0|$ part of the fermion in-fields: they may be set equal to the c number 1.

The equations of motion demand that the fermion part of the $|0\rangle\langle 0|$ operator be present for V_{in}^{+}, N_{in}^{+}, and $N^{+}V^{+}$, however. The fundamental cause for this restriction on the form of the fermion in-fields for both the Lee model and the Hamiltonian of Eq. (3.42) remains unknown. It is certainly an object for future study. However, the mere presence of such a restriction indicates that additional conditions must be satisfied when constructing asymptotic fields by the use of the Yang–Feldman equation. Furthermore, a field-theoretic model may have to obey special conditions in order to escape having asymptotic fields of the form found here, for the dyadic $|i\rangle\langle 0|$ is one that can exist for any model which has both eigenstates and a ground state.

ACKNOWLEDGMENTS

One of us (S. J.) would like to thank Dr. Gordon Fleming, Dr. Clifford Schumacher, Dr. George Payne, Mr. Donald Fahnline, and Mr. Michael Fitelson for many helpful discussions. Special appreciation goes to Dr. David Selengut for discussions concerning the material in the Appendix.

APPENDIX: PROOF OF SYMMETRY TO ALL SECTORS, EQ. (2.30)

A proof Eq. (2.30) for the lowest sectors for L_n^{+} can be obtained by considering those equations which result from adding Eq. (2.11a) to Eq. (2.11b) and Eq. (2.11c) to Eq. (2.11d):

$$(M_v - m_v)(L_1^{+} + L_2^{+}) = [H_\theta, L_1^{+} + L_2^{+}]_- + \lambda(L_3^{+} + L_4^{+})G, \quad \text{(A1a)}$$

$$(M_v - m_N)(L_3^{+} + L_4^{+}) = [H_\theta, L_3^{+} + L_4^{+}]_- + \lambda(L_1^{+} + L_2^{+})G^{+}. \quad \text{(A1b)}$$

Expanding the operators $A_v^{+} \equiv L_1^{+} + L_2^{+}$ and $B_v^{+} \equiv L_3^{+} + L_4^{+}$ in the most general combination of boson operators gives

$$A_v^{+} = \alpha_1 + \int \alpha_2(p, k)\theta^{+}(p)\theta(k) + \cdots,$$
$$B_v^{+} = \int \beta_1(p)\theta^{+}(p) + \int \beta_2(p_1, p_2; k_1)\theta^{+}(p_1)\theta^{+}(p_2)\theta(k_1) + \cdots. \quad \text{(A2)}$$

Substituting Eq. (A2) into Eq. (A1) and taking appropriate matrix elements yields

$$\alpha_1 = 0, \quad \alpha_2(p, k) = 0, \quad \beta_1(p) = 0. \quad \text{(A3)}$$

Therefore, to the lowest sector, Eq. (2.30) is true.

In order to prove symmetry to all sectors, write Eqs. (2.11) in the following form:

$$L_1^{+} = \frac{1}{(M_v - m_v)}[H_\theta, L_1^{+}]_- + \frac{\lambda}{(M_v - m_v)} GL_4^{+}, \quad \text{(A4a)}$$

$$L_2^{+} = \frac{1}{(M_v - m_v)}[H_\theta, L_2^{+}]_- + \frac{1}{(M_v - m_v)}(-\lambda[G, L_4^{+}]_- + \lambda L_3^{+}G), \quad \text{(A4b)}$$

$$L_3^{+} = \frac{1}{(M_v - m_N)}[H_\theta, L_3^{+}]_- + \frac{1}{(M_v - m_N)}(-\lambda[G^{+}, L_1^{+}]_- + \lambda L_2^{+}G^{+}), \quad \text{(A4c)}$$

$$L_4^{+} = \frac{1}{(M_v - m_N)}[H_\theta, L_4^{+}]_- + \frac{\lambda}{(M_v - m_N)} G^{+}L_1^{+}. \quad \text{(A4d)}$$

Each of these equations is characterized by having the commutator of the unknown boson term with H_θ on the right-hand side plus an inhomogeneous term which is an algebraic combination of other boson operators.

We may iterate Eqs. (A4), using the inhomogeneous terms which appear. Define, for any operator Q,

$${}_N(H_\theta^{\infty}, Q) \equiv \sum_{n=0}^{\infty} \frac{1}{(M_v - m_N)^{n+1}} \times [H_\theta, \cdots, [H_\theta, Q]_- \cdots]_n, \quad \text{(A5)}$$

where the symbol N on the left-hand side stands for

the bare mass of the N particle, m_N. Thus we get

$$L_1^+ = \lambda_V(H_\theta^\infty, GL_4^+), \tag{A6a}$$

$$L_2^+ = -\lambda_V(H_\theta^\infty, [G, L_4^+]_-) + \lambda_V(H_\theta^\infty, L_3^+G), \tag{A6b}$$

$$L_3^+ = -\lambda_N(H_\theta^\infty, [G^+, L_1^+]_-) + \lambda_N(H_\theta^\infty, L_2^+G^+), \tag{A6c}$$

$$L_4^+ = \lambda_N(H_\theta^\infty, G^+L_1^+). \tag{A6d}$$

It is possible to substitute Eq. (A6d) into Eq. (A6a) for L_4^+ and to get

$$L_1^+ = \lambda^2{}_V(H_\theta^\infty, G_N(H_\theta^\infty, G^+L_1^+)). \tag{A7}$$

Substitution of Eq. (A6d) into Eq. (A6b) for L_4^+ and of the resulting equation into (A6c) for L_2^+ leads to

$$\begin{aligned} L_3^+ = &-\lambda_N(H_\theta^\infty, [G^+, L_1^+]_-) \\ &+\lambda^2{}_N(H_\theta^\infty, {}_V(H_\theta^\infty, L_3^+G)G^+) \\ &-\lambda^3{}_N(H_\theta^\infty, {}_V(H_\theta^\infty, [G, {}_N(H_\theta^\infty, G^+L_1^+)]_-)G^+). \end{aligned} \tag{A8}$$

Now use of Eq. (A7) to reexpress the last term in Eq. (A8) as

$$-\lambda_N(H_\theta^\infty, L_1^+G^+) + \lambda^3{}_N(H_\theta^\infty, {}_V(H_\theta^\infty, {}_N(H_\theta^\infty, G^+L_1^+)G)G^+)$$

and rearrangement of terms results in

$$\begin{aligned} L_3^+ = &\lambda^2{}_N(H_\theta^\infty, {}_V(H_\theta^\infty, L_3^+G)G^+) \\ &+ {}_N(H_\theta^\infty, \{-\lambda G^+L_1^+ + \lambda^3{}_V(H_\theta^\infty, {}_N(H_\theta^\infty, G^+L_1^+)G)G^+\}). \end{aligned} \tag{A9}$$

The second term on the right-hand side of Eq. (A9) may be defined as Ω, and thus

$$L_3^+ = {}_N(H_\theta^\infty, \lambda_V(H_\theta^\infty, \lambda L_3^+G)G^+) + \Omega. \tag{A10}$$

It is now possible to iterate this equation using Ω; this procedure results in

$$L_3^+ = \sum_{n=1}^{\infty} \left\{ {}_N(H_\theta^\infty, \lambda\, {}_V(H_\theta^\infty, \lambda\Omega G)G^+) \right\}^n + \Omega, \tag{A11}$$

where an obvious notation has been used for the iterative series. Rewriting Ω in terms of its original definition and realizing that the iterative process is distributive allows one to cancel terms. This yields

$$L_3^+ = -\lambda\, {}_N(H_\theta^\infty, G^+L_1^+). \tag{A12}$$

A comparison with Eq. (A6d) shows

$$L_3^+ = -L_4^+, \tag{A13}$$

to all sectors. If Eq. (A13) is substituted into Eq. (A6b) for L_3^+,

$$L_2^+ = -\lambda\, {}_V(H_\theta^\infty, GL_4^+)$$

results. A comparison with Eq. (A6a) then shows that

$$L_1^+ = -L_2^+.$$

Thus Eq. (2.30) is proved to all sectors for the L_n^+.

The equations for the boson terms of the N_{in}^+ field, P_n^+, have the same structure as those for L_n^+, and therefore can be treated in exactly the same manner. The fact that the physical N_{in}^+ mass is equal to the bare N^+ mass m_N does not lead to difficulty; we may consider a Lee-model interaction Hamiltonian with an extra term δN^+N. The physical mass would then be $m_N + \delta$ and the structure of the equations of motion would remain unchanged. It would then be possible to carry out the proof of $P_1^+ = -P_2^+$, $P_3^+ = -P_4^+$ with the limit $\delta \to 0$ being taken in the end.

Proof of Orthogonality to All Sectors, Eq. (3.2)

Most of the equations of orthogonality [Eqs. (3.2b), (3.2c), (3.2e), and (3.2f)] can be shown to all sectors merely by substituting the proper Eq. (2.34) into the corresponding field and using the bare fermion algebra of Eq. (2.2). Equations (3.2a) and (3.2d) may not be proved this way, however, and must be looked at in greater detail.

Consider the term in Eq. (3.2a), $V_{\text{in}}N_{\text{in}}^+$; substituting Eq. (2.34) and the proper Hermitian conjugate equation into it leads to

$$V_{\text{in}}N_{\text{in}}^+ = \Phi - \Phi N^+N - \Phi V^+V + \Phi N^+NV^+V, \tag{A14}$$

where

$$\Phi \equiv L_1P_1^+ + L_4P_4^+. \tag{A15}$$

The boson operator Φ may be investigated by proper manipulation of the equations of motion, Eqs. (2.11). One obtains

$$\begin{aligned} &(m_N - M_v)L_1P_1^+ \\ &\quad = [H_\theta, L_1P_1^+] - \lambda L_4G^+P_1^+ + \lambda L_1GP_4^+, \\ &(m_N - M_v)L_4P_4^+ \\ &\quad = [H_\theta, L_4P_4^+] + \lambda L_4G^+P_1^+ - \lambda L_1GP_4^+. \end{aligned} \tag{A16}$$

Referring to the manner in which Eqs. (A4) were treated shows that an iterative solution of Eqs. (A16) leads to

$$L_1P_1^+ = -L_4P_4^+,$$

and thus

$$V_{\text{in}}N_{\text{in}}^+ = 0$$

is proved to all sectors. The final equation of orthogonality, Eq. (3.2d), is just the Hermitian conjugate of the above.

Proof of Eq. (3.19)

In order to prove this equation, realize that the desired operator $|0\rangle\langle 0|$ obeys

$$|0\rangle\langle 0|\, |n\rangle = 0, \tag{A17}$$

$$|0\rangle\langle 0|\, |0\rangle = |0\rangle, \tag{A18}$$

where $|n\rangle$ is the n-boson state ($n > 0$). Thus $|0\rangle\langle 0|$ must be of the form

$$|0\rangle\langle 0| = 1 + \sum_{j=1}^{\infty} \beta_j (a^+)^j (a)^j, \tag{A19}$$

where

$$\sum_{j=1}^{\infty} \beta j (a^+)^j (a)^j |n\rangle = -|n\rangle. \tag{A20}$$

The coefficient β_j must now be determined.

If we consider

$$\sum_{j=1}^{\infty} \beta_j (a^+)^j (a)^j \left(\frac{a^+}{(1!)^{\frac{1}{2}}} |0\rangle \right) = -\left(\frac{a^+}{(1!)^{\frac{1}{2}}} |0\rangle \right), \tag{A21}$$

where the quantity in parenthesis is the $|1\rangle$ state, then all terms in the summation, except the first, yield zero. This follows from

$$\begin{aligned} (a^+)^j (a)^j (a^+)^n |0\rangle &= 0, && j > n, \\ &= \frac{n!}{(n-j)!} (a^+)^n |0\rangle, && j \leq n, \end{aligned} \tag{A22}$$

which is a result of Eq. (3.17).

Thus we see that $\beta_1 = -1$. It is now possible to reason by induction and to assume

$$\beta_{N-1} = \frac{(-1)^{N-1}}{(N-1)!}.$$

Demanding that Eq. (A20) be true for $n = N$, using Eq. (A22), and taking matrix elements yields

$$\beta_N = -\frac{1}{N!} - \sum_{j=1}^{j=N-1} \frac{(-1)^j}{j!\,(N-j)!}. \tag{A23}$$

The summation on the right-hand side of this equation may be evaluated by using a well-known series for binomial coefficients[13]:

$$1 - \frac{N!}{1!\,(N-1)!} + \frac{N!}{2!\,(N-2)!} - \cdots + \frac{(-1)^N N!}{N!\,(N-N)!} = 0. \tag{A24}$$

Thus

$$\beta_N = (-1)^N / N!, \tag{A25}$$

and Eq. (3.19) is proved.

[13] L. B. W. Jolley, *Summation of Series* (Dover Publications, Inc., New York, 1961), p. 210 (no. 1102).

One should also note that Eq. (3.19) is an explicit solution of the integral equation

$$|0\rangle\langle 0| = 1 - \sum_{m=1}^{\infty} B_m^+ |0\rangle\langle 0| B_m, \tag{A26}$$

where

$$B_m^+ |0\rangle = [(a^+)^m / (m!)^{\frac{1}{2}}] |0\rangle. \tag{A27}$$

This fact is readily seen by the substitution of Eq. (3.19) into Eq. (A26) and the use of Eq. (A24).

Proof of Eq. (3.38)

Here we are interested only in the boson terms of Eq. (3.38). The fermion terms are obvious and shall be suppressed in the work following. The proof is similar to that of Eq. (3.19); we may construct $|0\rangle\langle 0|$ of the form

$$|0\rangle\langle 0| = 1 + \sum_{j=1}^{\infty} \alpha_j \times \int dp_1 \cdots dp_j \theta^+(p_1) \cdots \theta^+(p_j) \theta(p_1) \cdots \theta(p_j). \tag{A28}$$

The series on the right-hand side of this equation must be an operator which has the eigenvalue -1 when applied to a state containing any number (not zero) of θ particles of any momenta. We realize that, because

$$[\theta(k), \theta^+(p)]_- = \delta(\mathbf{p} - \mathbf{k}), \tag{A29}$$

the only terms in the series which will survive when applied to a state containing $n\theta$ particles are those for which $j \leq n$. The term with $j = n$ can be seen to obey the eigenvalue equation

$$\begin{aligned} \alpha_n \int dp_1 \cdots dp_n \theta^+(p_1) \cdots \theta^+(p_n) & \\ \times \theta(p_1) \cdots \theta(p_n) |\theta^+(q_1) \cdots \theta^+(q_n)\rangle & \\ = n!\, \alpha_n |\theta^+(q_1) \cdots \theta^+(q_n)\rangle. & \end{aligned} \tag{A30}$$

Using Eq. (A29), each term with $j < n$ can be shown to obey an equation similar to that above, but with eigenvalue $[n!/(n-j)!]\alpha_n$. Thus we wish the α_j to obey

$$\sum_{j=1}^{n} \alpha_j \frac{n!}{(n-j)!} = -1. \tag{A31}$$

We may use Eq. (A24) to determine that $\alpha_j = (-1)^j / j$. Thus Eq. (3.38) is proved.

Diffraction of Electromagnetic Waves by a Right-Angle Dielectric Wedge

EDGAR A. KRAUT AND GUY W. LEHMAN
Science Center of North American Rockwell Corporation, Thousand Oaks, California

(Received 15 January 1969)

The problem of diffraction of a plane-polarized electromagnetic wave incident on a right-angle dielectric wedge is formulated as a singular integral equation in **k** space. A solution of the singular integral equation is constructed as a power series in the index of refraction. This series converges when the index of refraction is near unity. Using this solution, the electric-field amplitude at the tip of the wedge is examined. We also prove as incorrect a closed-form analytic expression claimed in the literature to be a global solution of the problem considered here.

1. FORMULATION

As shown in Fig. 1, we consider a piecewise-homogeneous, isotropic, conducting, dielectric medium of infinite extent referred to rectangular Cartesian coordinates (x_1, x_2, x_3). The 90° wedge occupying the region $x_1 \geq 0$, $x_2 \geq 0$, $|x_3| \leq \infty$ is characterized by a constant electrical conductivity σ_d, a constant magnetic inductive capacity μ_d, and a constant electric inductive capacity ϵ_d. The medium external to the 90° wedge is similarly characterized by different constant parameters σ_v, μ_v, ϵ_v. It will be further assumed that both regions have the magnetic inductive capacity of free space, that is, $\mu_d = \mu_v = \mu_0$, and that the relative dielectric constant of the 90° wedge is unity or greater so that $\epsilon_d/\epsilon_v \geq 1$.

Our starting point is to derive an integral equation from Maxwell's equations relating the electric- and magnetic-field intensities $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{H}(\mathbf{r}, t)$ with the electric displacement $\mathbf{D}(\mathbf{r}, t)$ and magnetic induction $\mathbf{B}(\mathbf{r}, t)$ at the point $\mathbf{r}$ and time t. Maxwell's equations in the MKS system are

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0, \quad \nabla \cdot \mathbf{B} = 0,$$

$$\nabla \times \mathbf{H} - \frac{\partial \mathbf{D}}{\partial t} = \mathbf{J}, \quad \nabla \cdot \mathbf{D} = \rho, \tag{1.1}$$

with a continuity equation relating the charge density ρ and the current density $\mathbf{J}$ given by

$$\nabla \cdot \mathbf{J} + \frac{\partial \rho}{\partial t} = 0. \tag{1.2}$$

We define piecewise-constant electrical inductive capacity and conductivity functions by

$$\epsilon(\mathbf{r}) = \epsilon_d, x_1 \geq 0 \cap x_2 \geq 0 \cap |x_3| \leq \infty,$$
$$= \epsilon_v, x_1 \leq 0 \cup x_2 \leq 0 \cap |x_3| \leq \infty, \tag{1.3}$$

$$\sigma(\mathbf{r}) = \sigma_d, x_1 \geq 0 \cap x_2 \geq 0 \cap |x_3| \leq \infty,$$
$$= \sigma_v, x_1 \leq 0 \cup x_2 \leq 0 \cap |x_3| \leq \infty. \tag{1.4}$$

In terms of (1.4) the constitutive equations of the medium are

$$\mathbf{D}(\mathbf{r}, t) = \epsilon(\mathbf{r})\mathbf{E}(\mathbf{r}, t), \tag{1.5}$$

$$\mathbf{J}(\mathbf{r}, t) = \sigma(\mathbf{r})\mathbf{E}(\mathbf{r}, t), \tag{1.6}$$

$$\mathbf{B}(\mathbf{r}, t) = \mu_0\mathbf{H}(\mathbf{r}, t). \tag{1.7}$$

For a harmonic time dependence of the form

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r}) \exp(-i\omega t)$$

and

$$\mathbf{H}(\mathbf{r}, t) = \mathbf{H}(\mathbf{r}) \exp(-i\omega t),$$

Maxwell's equations yield

$$\nabla \times \mathbf{H} + i\omega\hat{\epsilon}(\mathbf{r})\mathbf{E} = 0, \tag{1.8}$$

$$\nabla \times \mathbf{E} - i\omega\mu_0\mathbf{H} = 0, \tag{1.9}$$

where

$$\hat{\epsilon}(\mathbf{r}) = \epsilon(\mathbf{r}) + i\sigma(\mathbf{r})\omega^{-1}. \tag{1.10}$$

Now taking the curl of (1.9) and using (1.8) yields

$$\{\nabla^2 + \omega^2\mu_0\hat{\epsilon}_v\}\mathbf{E} = \nabla\nabla \cdot \mathbf{E} - \omega^2\mu_0\hat{\epsilon}_v[(\hat{\epsilon}/\hat{\epsilon}_v) - 1]\mathbf{E}, \tag{1.11}$$

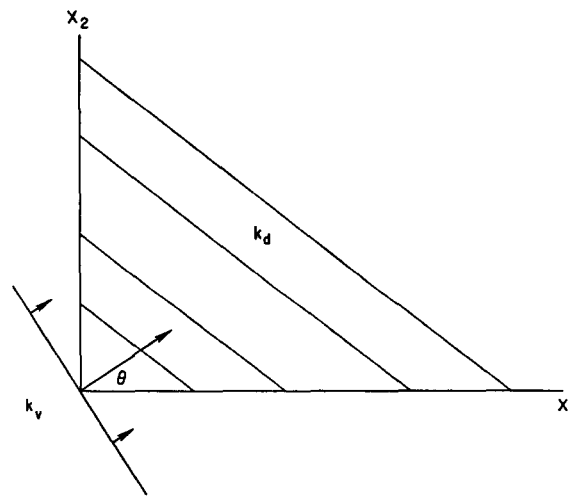

FIG. 1. Planewave incident on a dielectric wedge.

where

$$\hat{\epsilon}_v = \epsilon_v + i\sigma_v\omega^{-1}. \tag{1.12}$$

With the aid of the divergence of (1.8) we can express $\nabla\nabla \cdot \mathbf{E}$ as

$$\nabla\nabla \cdot \mathbf{E} = -\nabla\nabla \cdot [(\hat{\epsilon}(\mathbf{r})/\hat{\epsilon}_v) - 1]\mathbf{E}, \tag{1.13}$$

and using this in (1.11) gives

$$\{\nabla^2 + \omega^2\mu_0\hat{\epsilon}_v\}\mathbf{E} = -\{\nabla\nabla \cdot + \omega^2\mu_0\hat{\epsilon}_v\}[(\hat{\epsilon}(\mathbf{r})/\hat{\epsilon}_v) - 1]\mathbf{E}. \tag{1.14}$$

We shall restrict our considerations to electric fields which are polarized parallel to the vertex of the wedge, i.e., along the x_3 axis, and which do not depend on the x_3 coordinate. For such fields given by

$$\mathbf{E}(x_1, x_2) = [0, 0, E(x_1, x_2)], \tag{1.15}$$

Eq. (1.14) reduces to the scalar two-dimensional form

$$\{\nabla^2 + \omega^2\mu_0\hat{\epsilon}_v\}E(x_1, x_2) = -\omega^2\mu_0\hat{\epsilon}_v[(\hat{\epsilon}(\mathbf{r})/\hat{\epsilon}_v) - 1]E(x_1, x_2). \tag{1.16}$$

It is convenient to define complex wave vectors $k(\mathbf{r})$, k_d, and k_v by

$$\begin{aligned} k^2(\mathbf{r}) &= \omega^2\mu_0\hat{\epsilon}(\mathbf{r}) \\ &= k_d^2,\ x_1 \geq 0 \cap x_2 \geq 0 \cap |x_3| \leq \infty, \\ &= k_v^2,\ x_1 \leq 0 \cup x_2 \leq 0 \cap |x_3| \leq \infty, \end{aligned} \tag{1.17}$$

in terms of which (1.16) becomes

$$(\nabla^2 + k_v^2)E(x_1, x_2) = -\{k^2(\mathbf{r}) - k_v^2\}E(x_1, x_2). \tag{1.18}$$

Equation (1.18) immediately yields the basic integral equation for $E(x_1, x_2)$:

$$E(x_1, x_2) = E_0(x_1, x_2) + \frac{i}{4}(k_d^2 - k_v^2) \times \int_0^\infty \int_0^\infty H_0^{(1)}(k_v[(x_1 - x_1')^2 + (x_2 - x_2')^2]^{\frac{1}{2}}) \times E(x_1', x_2')\, dx_1'\, dx_2'. \tag{1.19}$$

Here $E_0(x_1, x_2)$ represents the electric field incident on the wedge $x_1 \geq 0 \cap x_2 \geq 0 \cap |x_3| \leq \infty$ and E_0 satisfies the equation

$$(\nabla^2 + k_v^2)E_0(x_1, x_2) = 0. \tag{1.20}$$

The kernel in (1.19) is a Hankel function of the first kind of order zero, and k_v and k_d are chosen to have positive imaginary parts

$$\text{Im}\,(k_v) > 0, \quad \text{Im}\,(k_d) > 0. \tag{1.21}$$

Our problem now is to solve the singular integral equation (1.19). In order to do this we shall first convert (1.19) into a singular integral equation for the Fourier transform of $E(x_1, x_2)$.

2. SINGULAR INTEGRAL EQUATION FOR THE FOURIER TRANSFORM OF THE ELECTRIC FIELD

Let (1.19) be multiplied by $\exp(ik_1x_1)$ and the integral $\int_{-\infty}^{\infty} \cdots dx_1$ taken. With the aid of the relation

$$\frac{i}{4}\int_{-\infty}^{\infty} H_0^{(1)}(k_v[(x_1 - x_1')^2 + (x_2 - x_2')^2]^{\frac{1}{2}}) \times \exp[ik_1(x_1 - x_1')]\, dx_1 = \frac{\exp[-\gamma|x_2 - x_2'|]}{2\gamma}, \tag{2.1}$$

where

$$\gamma = (k_1^2 - k_v^2)^{\frac{1}{2}}, \quad \text{Re}\,(\gamma) \geq 0, \tag{2.2}$$

Eq. (1.19) yields, on interchanging orders of integration, the result

$$E_+(k_1, x_2) + R_-(k_1, x_2) = E_+^{(0)}(k_1, x_2) + (k_d^2 - k_v^2) \times \int_0^\infty dx_2' \frac{\exp[-\gamma|x_2 - x_2'|]}{2\gamma} E_+(k_1, x_2'). \tag{2.3}$$

The functions $E_+(k_1, x_2)$, $R_-(k_1, x_2)$, and $E_+^{(0)}(k_1, x_2)$ are defined by

$$E_+(k_1, x_2) = \int_0^\infty E(x_1, x_2)\exp(ik_1x_1)\, dx_1, \tag{2.4}$$

$$R_-(k_1, x_2) = \int_{-\infty}^0 \exp(ik_1x_1) \times \{E(x_1, x_2) - E_0(x_1, x_2)\}\, dx_1, \tag{2.5}$$

$$E_+^{(0)}(k_1, x_2) = \int_0^\infty E_0(x_1, x_2)\exp(ik_1x_1)\, dx_1. \tag{2.6}$$

Each of the integrals (2.4)–(2.6) will converge uniformly for k_1 restricted to a certain half plane determined by asymptotic behavior assumed for $E(x_1, x_2)$ and $E_0(x_1, x_2)$ when $|x_1| \to \infty$. Recalling that the time dependence is $\exp(-i\omega t)$, the reflected field $\{E(x_1, x_2) - E_0(x_1, x_2)\}$ is assumed to have the asymptotic form

$$\{E(x_1, x_2) - E_0(x_1, x_2)\} = O\{\exp(-ib_1x_1 - ib_2x_2)\}, \tag{2.7}$$

for $x_1 \to -\infty$ or $x_2 \to -\infty$. Similarly, it is assumed that the incident field has the asymptotic form

$$E_0(x_1, x_2) = O\{\exp(ia_1x_1 + ia_2x_2)\}, \tag{2.8}$$

for $x_1 \to +\infty$ or $x_2 \to +\infty$, and that the transmitted field decays for large positive x_1 or x_2 more rapidly than does the incident field, i.e.,

$$E(x_1, x_2) = o\{\exp(ia_1x_1 + ia_2x_2)\}, \tag{2.9}$$

for $x_1 \to +\infty$ or $x_2 = +\infty$. It now follows from (2.4)–(2.9) that $E_+(k_1, x_2)$ and $E_+^{(0)}(k_1, x_2)$ are analytic in the half plane $\mathrm{Im}\,(k_1) > -\mathrm{Im}\,(a_1)$ and that $R_-(k_1, x_2)$ is analytic in the half plane $\mathrm{Im}\,(k_1) < \mathrm{Im}\,(b_1)$. These three functions (2.4)–(2.6) share the common strip of analyticity

$$-\mathrm{Im}\,(a_1) < \mathrm{Im}\,(k_1) < \mathrm{Im}\,(b_1). \tag{2.10}$$

Multiplying (2.3) by $\exp(ik_2x_2)$ and calculating $\int_{-\infty}^{\infty} \cdots dx_2$ with the aid of (2.11),

$$\int_{-\infty}^{\infty} \frac{\exp[ik_2(x_2 - x_2') - \gamma|x_2 - x_2'|]}{2\gamma}\, dx_2 = \frac{1}{k_1^2 + k_2^2 - k_v^2}, \tag{2.11}$$

yields

$$E_{++}(k_1, k_2) + R_{-+}(k_1, k_2) + R_{--}(k_1, k_2) + R_{+-}(k_1, k_2) = E_{++}^{(0)}(k_1, k_2) + \frac{(k_d^2 - k_v^2)}{(k_1^2 + k_2^2 - k_v^2)} E_{++}(k_1, k_2), \tag{2.12}$$

where

$$E_{++}(k_1, k_2) = \int_0^{\infty}\int_0^{\infty} E(x_1, x_2) \times \exp[ik_1x_1 + ik_2x_2]\, dx_1\, dx_2, \tag{2.13}$$

$$R_{-+}(k_1, k_2) = \int_0^{\infty}\int_{-\infty}^{0} \{E(x_1, x_2) - E_0(x_1, x_2)\} \times \exp[ik_1x_1 + ik_2x_2]\, dx_1\, dx_2, \tag{2.14}$$

$$R_{--}(k_1, k_2) = \int_{-\infty}^{0}\int_{-\infty}^{0} \{E(x_1, x_2) - E_0(x_1, x_2)\} \times \exp[ik_1x_1 + ik_2x_2]\, dx_1\, dx_2, \tag{2.15}$$

$$R_{+-}(k_1, k_2) = \int_{-\infty}^{0}\int_0^{\infty} \{E(x_1, x_2) - E_0(x_1, x_2)\} \times \exp[ik_1x_1 + ik_2x_2]\, dx_1\, dx_2, \tag{2.16}$$

$$E_{++}^{(0)}(k_1, k_2) = \int_0^{\infty}\int_0^{\infty} E_0(x_1, x_2) \times \exp[ik_1x_1 + ik_2x_2]\, dx_1\, dx_2. \tag{2.17}$$

Arguments based on the assumed asymptotic forms (2.7)–(2.9) show that the four functions (2.13)–(2.16) are analytic, respectively, in the four pairs of half planes

$$(++)\colon \mathrm{Im}\,(k_1) > -\mathrm{Im}\,(a_1), \quad \mathrm{Im}\,(k_2) > -\mathrm{Im}\,(a_2), \tag{2.18}$$

$$(-+)\colon \mathrm{Im}\,(k_1) < \mathrm{Im}\,(b_1), \quad \mathrm{Im}\,(k_2) > -\mathrm{Im}\,(a_2), \tag{2.19}$$

$$(--)\colon \mathrm{Im}\,(k_1) < \mathrm{Im}\,(b_1), \quad \mathrm{Im}\,(k_2) < \mathrm{Im}\,(b_2), \tag{2.20}$$

$$(+-)\colon \mathrm{Im}\,(k_1) > -\mathrm{Im}\,(a_1), \quad \mathrm{Im}\,(k_2) < \mathrm{Im}\,(b_2). \tag{2.21}$$

The intersection of these four pairs of half planes is the tube or pair of strips $T(A)$ specified by

$$-\mathrm{Im}\,(a_J) < \mathrm{Im}\,(k_J) < \mathrm{Im}\,(b_J), \quad J = 1, 2. \tag{2.22}$$

Now consider the function

$$H(k_1, k_2) = 1/(k_1^2 + k_2^2 - k_v^2), \tag{2.23}$$

multiplying $E_{++}(k_1, k_2)$ on the right side of (2.12). This function is analytic and uniformly bounded in the tube (pair of strips) defined by

$$|\mathrm{Im}\,(k_J)| < B_J, \quad J = 1, 2, \tag{2.24}$$

$$(B_1^2 + B_2^2)^{\frac{1}{2}} < |\mathrm{Im}\,(k_v)|, \tag{2.25}$$

where

$$|\mathrm{Re}\,(k_J)| \leq \infty. \tag{2.26}$$

As a result of (1.20) and (2.8), the numbers a_1 and a_2 appearing in (2.8) and (2.22) must satisfy

$$a_1 = k_v \cos\theta, \quad a_2 = k_v \sin\theta; \tag{2.27}$$

consequently

$$|\mathrm{Im}\,(a_1)| < \mathrm{Im}\,(k_v), \tag{2.28}$$

$$|\mathrm{Im}\,(a_2)| < \mathrm{Im}\,(k_v), \tag{2.29}$$

when

$$0 < \theta < \pi/2.$$

Since the incident field $E_{++}^{(0)}(k_1, k_2)$ specified by (2.17) is also analytic in the pair of half planes (2.18), it is possible to make (2.22) a common tube of analyticity for each term in (2.12) by choosing $b_J = B_J$ where B_J satisfies (2.25).

We see that (2.12) is an equation of the Wiener–Hopf type in two complex variables, valid in the common pair of strips (2.22) and requiring for its solution the determination of four unknown analytic functions E_{++}, E_{-+}, R_{--}, R_{+-}, in terms of the kernel $H(k_1, k_2)$ and the given incident field $E_{++}^{(0)}(k_1, k_2)$. Our approach to solving (2.12) will be to convert it to a singular integral equation in two complex variables which can be solved by a Neumann series. To do this, we shall assume and justify *a posteriori* that $E_{++}(k_1, k_2)$ has a bounded L_2 norm in the tube (2.22); in other words,

$$\|E_{++}\|_2 = \left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |E_{++}(k_1, k_2)|^2\, dk_1\, dk_2\right\}^{\frac{1}{2}} < \infty \tag{2.30}$$

when $\mathrm{Im}\,(k_1)$ and $\mathrm{Im}\,(k_2)$ lie in the tube (2.22). It follows from (2.26)–(2.29) that $H(k_1, k_2)$ is both analytic and *bounded* in the tube (2.22) provided

$$b_1 = B_1, \quad b_2 = B_2. \tag{2.31}$$

Therefore, the product $H(k_1, k_2)E_{++}(k_1, k_2)$ must also be in L_2 because of (2.30) and we have

$$\|HE_{++}\|_2 < \infty. \tag{2.32}$$

Now, by a slight modification[1] of a remark due to Bochner,[2] we can assert that the function $H(k_1, k_2) \times E_{++}(k_1, k_2)$, which is analytic and of bounded L_2 norm in the tube (2.22), has a *unique* additive decomposition into the sum of four functions analytic and bounded, respectively, in the four pairs of half planes (2.18)–(2.21) intersecting to form the tube (2.22). These four functions are uniquely determined and are representable by Cauchy integrals.

On decomposing the last term in (2.12) by Bochner's theorem and equating the (++) parts on both sides of (2.12) (because of uniqueness), we immediately obtain the required singular integral equation for E_{++}:

$$E_{++}(k_1, k_2) = E_{++}^{(0)}(k_1, k_2) + \frac{(k_d^2 - k_v^2)}{(2\pi i)^2} \times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{E_{++}(z_1, z_2)\, dz_1\, dz_2}{(z_1^2 + z_2^2 - k_v^2)(z_1 - k_1)(z_2 - k_2)}, \tag{2.33}$$

where $\text{Im}\,(k_1) \geq 0$ and $\text{Im}\,(k_2) \geq 0$. If we can obtain a solution of (2.33) which belongs to L_2, then additively decomposing $H(k_1, k_2)E_{++}(k_1, k_2)$ by means of Cauchy integrals yields unique expressions for the remaining unknowns $R_{-+}(k_1, k_2)$, $R_{--}(k_1, k_2)$, and $R_{+-}(k_1, k_2)$ by virtue of Bochner's theorem. In other words, we need only compute $E_{++}(k_1, k_2)$ to completely solve the two variable Wiener–Hopf problem (2.12).

3. SOLUTION OF THE SINGULAR INTEGRAL EQUATION

We begin our discussion of (2.33) by allowing $\text{Im}\,(k_1) \to 0^+$ and $\text{Im}\,(k_2) \to 0^+$ to obtain

$$E_{++}(k_1, k_2) = E_{++}^{(0)}(k_1, k_2) + \frac{(k_d^2 - k_v^2)}{4}\Bigg\{H(k_1, k_2)E_{++}(k_1, k_2) + \frac{P}{\pi i}\int_{-\infty}^{\infty}\frac{H(z_1, k_2)E_{++}(z_1, k_2)}{(z_1 - k_1)}\,dz_1 + \frac{P}{\pi i}\int_{-\infty}^{\infty}\frac{H(k_1, z_2)E_{++}(k_1, z_2)}{(z_2 - k_2)}\,dz_2 + \frac{P}{\pi i}\int_{-\infty}^{\infty}\frac{dz_1}{(z_1 - k_1)}\frac{P}{\pi i} \times \int_{-\infty}^{\infty}\frac{H(z_1, z_2)E_{++}(z_1, z_2)}{(z_2 - k_2)}\,dz_2\Bigg\}, \tag{3.1}$$

which is the singular integral equation we shall solve. Equation (3.1) can be written in a more compact operator form by introducing the singular commuting operators[3] S_i defined by

$$S_i = \frac{P}{\pi i}\int_{-\infty}^{\infty}\frac{dz_i}{(z_i - k_i)}, \quad i = 1, 2, \tag{3.2}$$

in terms of which (3.1) becomes

$$E_{++} = E_{++}^{(0)} + \tfrac{1}{4}\lambda(I + S_1)(I + S_2)(HE_{++}), \tag{3.3}$$

where I is the identity operator and $\lambda = (k_d^2 - k_v^2)$.

In deriving (2.33) it was assumed in (2.30) that the L_2 norm of E_{++} is bounded. It will be shown next that when the L_2 norm of the incident field $E_{++}^{(0)}$ is bounded then, because $H(k_1, k_2)$ is bounded in the tube (2.22) it is possible to choose $(k_d^2 - k_v^2)$ such that (2.30) is satisfied.

Theorem 1: Suppose that

$$\|E_{++}^{(0)}\|_2 < \infty, \tag{3.4}$$

and that

$$0 < \max_{\text{Im}\,(\mathbf{k})=\mathbf{0}} |H(k_1, k_2)| \equiv \max |H| < \infty; \tag{3.5}$$

then there exist values of λ with $0 < |\lambda| < \infty$, such that the function $E_{++}(k_1, k_2)$ satisfying (3.3) has a bounded L_2 norm

$$\|E_{++}\|_2 < \infty. \tag{3.6}$$

Proof: Application of Minkowski's inequality to (3.3) yields

$$\|E_{++}\|_2 \leq \|E_{++}^{(0)}\|_2 + \tfrac{1}{4}|\lambda|\{\|HE_{++}\|_2 + \|S_1HE_{++}\|_2 + \|S_2HE_{++}\|_2 + \|S_1S_2HE_{++}\|_2\}. \tag{3.7}$$

Observe now that the principal-value operator S_i gives the Hilbert transform of the function on which it operates and that the Hilbert transform is a bounded linear operator in L_2 satisfying[4]

$$\|S_iHE_{++}\|_2 = \|HE_{++}\|_2, \quad i = 1, 2. \tag{3.8}$$

The double principal-value integral appearing in the last term in (3.7) is taken with respect to two different variables; hence

$$\|S_1S_2HE_{++}\|_2 = \|S_2HE_{++}\|_2 = \|HE_{++}\|_2. \tag{3.9}$$

Because of (3.5),

$$\|HE_{++}\|_2 < \max|H|\,\|E_{++}\|_2, \tag{3.10}$$

[1] E. Kraut, S. Busenberg, and W. Hall, Bull. Am. Math. Soc. **74**, 372 (1968).

[2] S. Bochner, Am. J. Math. **59**, 732 (1937).

[3] F. D. Gakhov, *Boundary Value Problems* (Pergamon Press, Inc., New York, 1966), pp. 70–72.

[4] E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals* (Clarendon Press, Oxford, England, 1948), 2nd ed., p. 122.

and combining (3.8), (3.9), and (3.10) with (3.7) yields

$$\|E_{++}\|_2 \leq \frac{\|E_{++}^{(0)}\|_2}{(1 - |\lambda| \max |H|)}. \tag{3.11}$$

The choice

$$|\lambda| < \{\max |H|\}^{-1} \tag{3.12}$$

establishes (3.6) and proves the theorem.

Arguments similar to those used to prove Theorem 1 will now be employed to show that λ may be chosen small enough so that the integral equation (3.3) is a contraction mapping with respect to the L_2 norm. Consequently, (3.3) has a unique solution E_{++} in L_2 *which can be obtained as a limit of successive approximations* converging in the L_2 norm to E_{++}.

Theorem 2: For $0 < |\lambda| < \{\max |H|\}^{-1}$ and E_{++} in the complete normed linear space L_2,

$$T(E_{++}) = E_{++}^{(0)} + \tfrac{1}{4}\lambda(I + S_1)(I + S_2)(HE_{++}) \tag{3.13}$$

is a contraction mapping with respect to the L_2 norm. Hence, the integral equation

$$T(E_{++}) = E_{++} \tag{3.14}$$

has one and only one fixed point belonging to L_2. This fixed point is the limit of a sequence of successive approximations converging in L_2 norm to E_{++}.

Proof: Let $E_{++}^{(2)}$ and $E_{++}^{(1)}$ be members of the L_2 function space; we have

$$\|T(E_{++}^{(2)}) - T(E_{++}^{(1)})\|_2 = \tfrac{1}{4}|\lambda| \, \|(I + S_1)(I + S_2)H(E_{++}^{(2)} - E_{++}^{(1)})\|_2. \tag{3.15}$$

Expanding the operators $(I + S_1)(I + S_2)$ on the right side of (3.15) and using Minkowski's inequality and (3.8)–(3.10) yields

$$\|T(E_{++}^{(2)}) - T(E_{++}^{(1)})\|_2 \leq |\lambda| \max |H| \, \|E_{++}^{(2)} - E_{++}^{(1)}\|_2. \tag{3.16}$$

This shows that $T(E_{++})$ is a contraction mapping with respect to the L_2 norm, provided $|\lambda|$ is restricted as in (3.12). The remainder of the theorem is then a consequence of Banach's fixed-point theorem[5] and the observation that the normed function space L_2 is complete.

The results we have just obtained apply immediately to (3.1) with

$$\lambda = (k_d^2 - k_v^2) \tag{3.17}$$

and

$$\max |H| = \left| \frac{1}{2 \operatorname{Im}(k_v) \operatorname{Re}(k_v)} \right|, \tag{3.18}$$

from (2.23). In other words, the convergence criterion (3.12) shows that the iterative solution of (3.1) converges for

$$|k_d^2 - k_v^2| < 2 \, |\operatorname{Im}(k_v) \operatorname{Re}(k_v)|. \tag{3.19}$$

Let p be a complex relative index of refraction defined by

$$p = k_d/k_v. \tag{3.20}$$

In terms of p, Eq. (3.19) becomes

$$|p^2 - 1| < 2 \left|\operatorname{Im}(k_v/|k_v|) \operatorname{Re}(k_v/|k_v|)\right|. \tag{3.21}$$

A maximum value on the right side of (3.21) is taken on when $k_v = |k_v| \exp(i\pi/4)$ and the iterative solution of (3.1) will converge in this case for wedges whose complex relative dielectric constant satisfies

$$0 \leq |p^2 - 1| < 1 \tag{3.22}$$

or

$$1 \leq |p| < \sqrt{2}. \tag{3.23}$$

The successive approximations to the solution

$$E_{++}^{(0)}(k_1, k_2), E_{++}^{(1)}(k_1, k_2), \cdots, E_{++}^{(n)}(k_1, k_2)$$

take the form

$$E_{++}^{(n+1)}(k_1, k_2) = E_{++}^{(0)}(k_1, k_2) + \tfrac{1}{4}(k_d^2 - k_v^2)(I + S_1)(I + S_2)(HE_{++}^{(n)}), \tag{3.24}$$

where $n = 0, 1, 2, \cdots$, $\operatorname{Im}(k_1) = 0$, $\operatorname{Im}(k_2) = 0$, and $(k_d^2 - k_v^2)$ satisfies (3.19). The solution of (3.1) for real k_1 and k_2 is the limit of the sequence (3.24):

$$E_{++}(k_1, k_2) = \lim_{n \to \infty} E_{++}^{(n)}(k_1, k_2). \tag{3.25}$$

4. ELECTRIC FIELD AT THE VERTEX OF THE WEDGE

The electric-field amplitude at the vertex of the wedge is given by

$$E(0, 0) = \lim_{n \to \infty} E^{(n)}(0, 0), \tag{4.1}$$

where

$$E^{(n)}(0, 0) = \lim_{\substack{x_1 \to 0^+ \\ x_2 \to 0^+}} \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E_{++}^{(n)}(k_1, k_2) \times \exp(-ik_1x_1 - ik_2x_2) \, dk_1 \, dk_2. \tag{4.2}$$

[5] W. Pogorzelski, *Integral Equations and Their Applications* (Pergamon Press, Inc., New York, 1966), Vol. I, p. 197.

In performing the Fourier inversion (4.2), it is convenient to use the relation

$$\lim_{\substack{x_1\to 0^+\\ x_2\to 0^+}} \frac{1}{(2\pi i)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E_{++}^{(n)}(k_1,k_2) \times \exp(-ik_1x_1 - ik_2x_2)\,dk_1\,dk_2 = \lim_{\substack{\mathrm{Im}(k_1)\to+\infty\\ \mathrm{Im}(k_2)\to+\infty}} \frac{1}{(2\pi i)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{k_1k_2E_{++}^{(n)}(z_1,z_2)\,dz_1\,dz_2}{(z_1-k_1)(z_2-k_2)}, \tag{4.3}$$

obtained with the aid of

$$\frac{k_1k_2}{(z_1-k_1)(z_2-k_2)} = \left(-1+\frac{z_1}{z_1-k_1}\right)\left(-1+\frac{z_2}{z_2-k_2}\right). \tag{4.4}$$

Equations (4.2) and (4.3) now give

$$E^{(n)}(0,0) = -\lim_{\substack{\mathrm{Im}(k_1)\to\infty\\ \mathrm{Im}(k_2)\to\infty}} k_1k_2E_{++}^{(n)}(k_1,k_2), \tag{4.5}$$

for the nth-order approximation to the electric field at the vertex of the wedge. To first order in $(k_d - k_v)$, using (3.24) and (2.33),

$$E_{++}^{(1)}(k_1,k_2) = E_{++}^{(0)}(k_1,k_2) + \frac{(k_d^2-k_v^2)}{(2\pi i)^2} \times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{E_{++}^{(0)}(z_1,z_2)\,dz_1\,dz_2}{(z_1^2+z_2^2-k_v^2)(z_1-k_1)(z_2-k_2)} \tag{4.6}$$

and, from (4.5),

$$E^{(1)}(0,0) = E_0(0,0) - \frac{(k_d^2-k_v^2)}{(2\pi i)^2} \times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{E_{++}^{(0)}(z_1,z_2)\,dz_1\,dz_2}{(z_1^2+z_2^2-k_v^2)}. \tag{4.7}$$

When the incident field $E_0(x_1, x_2)$ is a plane wave of the form

$$E_0(x_1,x_2) = E_0\exp[i(a_1x_1 + a_2x_2)], \tag{4.8}$$

where

$$a_1 = k_v\cos\theta, \quad a_2 = k_v\sin\theta, \tag{4.9}$$

then

$$E_{++}^{(0)}(z_1,z_2) = -E_0/(z_1+a_1)(z_2+a_2) \tag{4.10}$$

and (4.7) becomes

$$E^{(1)}(0,0) = E_0\left\{1 + \frac{(k_d^2-k_v^2)}{(2\pi i)^2} \times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{dz_1\,dz_2}{(z_1^2+z_2^2-k_v^2)(z_1+a_1)(z_2+a_2)}\right\}. \tag{4.11}$$

A short calculation yields

$$E^{(1)}(0,0) = E_0\left\{1 - \frac{(k_d^2-k_v^2)}{2\pi k_v^2}\int_0^{\infty}\frac{\xi\,d\xi}{(1+\xi^2)^2} \times \left(\frac{\sin\theta}{(\xi^2+\cos^2\theta)^{\frac{1}{2}}} + \frac{\cos\theta}{(\xi^2+\sin^2\theta)^{\frac{1}{2}}}\right)\right\}, \tag{4.12}$$

for the first-order field at the wedge tip. For the case of grazing incidence where $\theta = 0$ or $\theta = \pi/2$, Eq. (4.12) becomes

$$E^{(1)}(0,0) = E_0\{1 - \tfrac{1}{8}(p^2-1)\}, \tag{4.13}$$

where $p = k_d/k_v$. To second order in $(k_d^2 - k_v^2)$, the field at the wedge tip is given by

$$E^{(2)}(0,0) = E^{(1)}(0,0) + \frac{(k_d^2-k_v^2)^2}{(2\pi i)^4}E_0\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{dz_1\,dz_2}{(z_1^2+z_2^2-k_v^2)} \times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{d\zeta_1\,d\zeta_2}{(\zeta_1^2+\zeta_2^2-k_v^2)(\zeta_1+a_1)(\zeta_2+a_2)(\zeta_1-z_1)(\zeta_2-z_2)}. \tag{4.14}$$

5. COMPARISON WITH OTHER INVESTIGATIONS

The dielectric-wedge problem has also been investigated by Radlow[6] and by Kuo and Plonus.[7] Radlow[6] gives what he claims is a closed-form global analytic solution of the problem. His solution, given in Eqs. (5.1) and (5.3) of his paper,[6] becomes in our notation

$$E_{++}(k_1,k_2) = \frac{-E_0}{(k_1+a_1)(k_2+a_2)}\frac{1}{K_{++}(k_1,k_2)K_{-+}(-a_1,k_2)K_{--}(-a_1,-a_2)K_{+-}(k_1,-a_2)}, \tag{5.1}$$

where[8]

$$K_{++}(k_1,k_2) = \exp\frac{1}{(2\pi i)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{dz_1\,dz_2}{(z_1-k_1)(z_2-k_2)}\log\left[\frac{k_d^2-(z_1^2+z_2^2)}{k_v^2-(z_1^2+z_2^2)}\right], \tag{5.2}$$

[6] J. Radlow, Intern. J. Eng. Sci. 2, 275 (1964).
[7] N. H. Kuo and M. A. Plonus, J. Math. & Phys. 46, 394 (1967).
[8] E. A. Kraut, J. Math. Phys. 9, 1481 (1968).

with

$$\text{Im}\,(k_1) > 0, \quad \text{Im}\,(k_2) > 0. \tag{5.3}$$

The other three factors K_{-+}, K_{--}, and K_{+-} are defined exactly as in (5.2) except that, instead of (5.3), one has Im $(k_1) < 0$, Im $(k_2) > 0$; Im $(k_1) < 0$, Im $(k_2) < 0$; and Im $(k_1) > 0$, Im $(k_2) < 0$, respectively. It is asserted in Ref. 6 that (5.1) is the unique solution of the singular integral equation (2.33). One of us has pointed out[9] that an expression of the form (5.1) satisfies (2.33) *only if* the branch line integrals associated with the branch cuts in K_{++}, K_{-+}, K_{+-} cancel one another out when the double integral in (2.33) is evaluated. We shall show, by directly substituting (5.1) into (2.33) and explicitly evaluating the double integral, that (5.1) *does not* satisfy (2.33) and is therefore not a solution of the problem. Contrary to our results, Kuo and Plonus indicate[7] that they believe that (5.1) is the correct solution of (2.33). However, they claim that (5.1) is too complicated to be inverse Fourier-transformed and therefore they suggest the use of approximate methods to solve (2.33). In Eqs. (6.4) and (6.5) of Ref. 7, the determination of approximate expressions for $E_{++}(k_1, k_2)$ and $E(x_1, x_2)$ (when the dielectric constant is near unity) is reduced to single quadratures. These quadratures, however, were not carried out, nor was the validity of the approximate decoupling scheme used to obtain them critically examined.

After showing that (5.1) does not satisfy (2.33), we shall compute the inverse Fourier transform of (5.1) at the wedge tip and compare the result with (4.12)–(4.14).

We begin by observing that (2.33) can be written in the form

$$\frac{1}{2\pi i)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{z_1^2+z_2^2-k_d^2}{z_1^2+z_2^2-k_v^2}\right)\frac{E_{++}(z_1, z_2)\,dz_1\,dz_2}{(z_1-k_1)(z_2-k_2)} = E_{++}^{(0)}(k_1, k_2), \tag{5.4}$$

where the kernel

$$K(z_1, z_2) = (z_1^2 + z_2^2 - k_d^2)/(z_1^2 + z_2^2 - k_v^2) \tag{5.5}$$

can be factorized in the form[8]

$$K(z_1, z_2) = K_{++}(z_1, z_2)K_{-+}(z_1, z_2)K_{--}(z_1, z_2)K_{+-}(z_1, z_2), \tag{5.6}$$

with the K's determined as mentioned in connection with (5.2). It is possible, for example, to write down pairwise products of factors in (5.6) algebraically by inspection:

$$K_{++}(z_1, z_2)K_{+-}(z_1, z_2) = \frac{z_1 + i(z_2^2 - k_d^2)^{\frac{1}{2}}}{z_1 + i(z_2^2 - k_v^2)^{\frac{1}{2}}}. \tag{5.7}$$

This observation permits us to write KE_{++} in (5.4) with E_{++} given by (5.1) as the product of a function analytic in Im $(k_1) < 0$, Im $(k_2) < 0$, and a function having explicit algebraic singularities in Im $(k_1) < 0$, Im $(k_2) < 0$. Thus (5.4) becomes

$$\frac{1}{(2\pi i)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{[z_1-(k_d^2-z_2^2)^{\frac{1}{2}}][z_2-(k_d^2-z_1^2)^{\frac{1}{2}}][a_1+(k_v^2-z_2^2)^{\frac{1}{2}}][a_2+(k_v^2-z_1^2)^{\frac{1}{2}}]}{[z_1-(k_v^2-z_2^2)^{\frac{1}{2}}][z_2-(k_v^2-z_1^2)^{\frac{1}{2}}][a_1+(k_d^2-z_2^2)^{\frac{1}{2}}][a_2+(k_d^2-z_1^2)^{\frac{1}{2}}]} \times \frac{\Psi(z_1, z_2)\,dz_1\,dz_2}{(z_1+a_1)(z_2+a_2)(z_1-k_1)(z_2-k_2)} = \frac{1}{(k_1+a_1)(k_2+a_2)}, \tag{5.8}$$

where

$$\Psi(z_1, z_2) = \frac{K_{--}(-a_1, z_2)K_{--}(z_1, -a_2)}{K_{--}(-a_1, -a_2)K_{--}(z_1, z_2)}. \tag{5.9}$$

It is convenient to rewrite (5.8) in the form

$$I(k_d, k_v) = \frac{1}{(2\pi i)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\{f(z_1, z_2)\Psi(z_1, z_2) - 1\}}{(z_1+a_1)(z_2+a_2)(z_1-k_1)(z_2-k_2)} \equiv 0, \tag{5.10}$$

where

$$f(z_1, z_2) = \frac{[z_1-(k_d^2-z_2^2)^{\frac{1}{2}}][z_2-(k_d^2-z_1^2)^{\frac{1}{2}}][a_1+(k_v^2-z_2^2)^{\frac{1}{2}}][a_2+(k_v^2-z_1^2)^{\frac{1}{2}}]}{[z_1-(k_v^2-z_2^2)^{\frac{1}{2}}][z_2-(k_v^2-z_1^2)^{\frac{1}{2}}][a_1+(k_d^2-z_2^2)^{\frac{1}{2}}][a_2+(k_d^2-z_1^2)^{\frac{1}{2}}]}. \tag{5.11}$$

Our procedure now is to expand $I(k_d, k_v)$ in a Taylor series of the form

$$I(k_d, k_v) = I(k_v, k_v) + (k_d - k_v)\left(\frac{\partial I}{\partial k_d}\right)_{k_d=k_v} + \tfrac{1}{2}(k_d - k_v)^2\left(\frac{\partial^2 I}{\partial k_d^2}\right)_{k_d=k_v} + \cdots \equiv 0. \tag{5.12}$$

If (5.12) vanishes identically for arbitrary k_d and k_v, then each and every coefficient in the Taylor series

[9] E. A. Kraut, Bull. Seism. Soc. Am. **58**, 1083 (1968).

expansion (5.12) must vanish. We shall show that

$$I(k_v, k_v) \equiv 0, \tag{5.13}$$

$$\left(\frac{\partial I}{\partial k_d}\right)_{k_d=k_v} \equiv 0, \tag{5.14}$$

but that

$$\left(\frac{\partial^2 I}{\partial k_d^2}\right)_{k_d=k_v} \not\equiv 0, \tag{5.15}$$

and *consequently* (5.1) *does not satisfy* (5.4) *or* (2.33). Equation (5.13) follows immediately on noting that $f(z_1, z_2) \equiv 1$ and $\Psi(z_1, z_2) \equiv 1$ when $k_d = k_v$. To obtain (5.14) we observe that

$$\left(\frac{\partial I}{\partial k_d}\right)_{k_d=k_v} = \frac{1}{(2\pi i)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{\left(\frac{\partial f}{\partial k_d}\right)_{k_d=k_v} + \left(\frac{\partial \Psi}{\partial k_d}\right)_{k_d=k_v}\right\} \times \frac{dz_1\, dz_2}{(z_1 + a_1)(z_2 + a_2)(z_1 - k_1)(z_2 - k_2)}. \tag{5.16}$$

Since $(\partial\Psi/\partial k_d)_{k_d=k_v}$ is analytic in $\mathrm{Im}\,(k_1) < 0$, $\mathrm{Im}\,(k_2) < 0$, the contribution from this term arises from the residues at the poles $z_1 = -a_1$, $z_2 = -a_2$, provided $\mathrm{Im}\,(a_1) > 0$ and $\mathrm{Im}\,(a_2) > 0$. However, since the constant term in the Taylor series expansion of $\Psi(z_1, z_2, k_d, k_v)$ about $z_1 = -a_1$, $z_2 = -a_2$ is unity, independent of the choice of k_d and k_v,

$$\left(\frac{\partial \Psi}{\partial k_d}\right)_{\substack{z_1=-a_1 \\ z_2=-a_2}} = 0, \tag{5.17}$$

and therefore

$$\left(\frac{\partial I}{\partial k_d}\right)_{k_d=k_v} = \frac{1}{(2\pi i)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{\partial f}{\partial k_d}\right)_{k_d=k_v} \times \frac{dz_1\, dz_2}{(z_1 + a_1)(z_2 + a_2)(z_1 - k_1)(z_2 - k_2)}. \tag{5.18}$$

From (5.11),

$$\left(\frac{\partial f}{\partial k_d}\right)_{k_d=k_v} = \frac{-k_v}{(k_v^2 - z_1^2)^{\frac{1}{2}}}\left\{\frac{1}{z_2 - (k_v^2 - z_1^2)^{\frac{1}{2}}} + \frac{1}{a_2 + (k_v^2 - z_1^2)^{\frac{1}{2}}}\right\} - \frac{k_v}{(k_v^2 - z_2^2)^{\frac{1}{2}}}\left\{\frac{1}{z_1 - (k_v^2 - z_2^2)^{\frac{1}{2}}} + \frac{1}{a_1 + (k_v^2 - z_2^2)^{\frac{1}{2}}}\right\}, \tag{5.19}$$

where the first term in brackets in (5.19) is analytic for $\mathrm{Im}\,(z_2) < 0$ when $\mathrm{Im}\,[(k_v^2 - z_1^2)^{\frac{1}{2}}] > 0$ and the second term in brackets is analytic for $\mathrm{Im}\,(z_1) < 0$ when $\mathrm{Im}\,[(k_v^2 - z_2^2)^{\frac{1}{2}}] > 0$. The integral (5.18) may now be performed by closing the first term of (5.19) in the lower z_2 half plane and the second term of (5.19) in the lower z_1 half plane. The residues at $z_2 = -a_2$ and $z_1 = -a_1$ are zero and (5.14) follows.

To establish (5.15), we note that

$$\left(\frac{\partial^2 I}{\partial k_d^2}\right)_{k_d=k_v} = \frac{1}{(2\pi i)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{\frac{\partial^2 f}{\partial k_d^2} + \frac{\partial^2 \Psi}{\partial k_d^2} + 2\frac{\partial f}{\partial k_d}\frac{\partial \Psi}{\partial k_d}\right\}_{k_d=k_v} \frac{dz_1\, dz_2}{(z_1 + a_1)(z_2 + a_2)(z_1 - k_1)(z_2 - k_2)}. \tag{5.20}$$

The last two terms in (5.20) separately integrate to zero because of (5.17) and (5.19), and one has

$$\left(\frac{\partial^2 I}{\partial k_d^2}\right)_{k_d=k_v} = \frac{1}{(2\pi i)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{\partial^2 f}{\partial k_d^2}\right)_{k_d=k_v} \frac{dz_1\, dz_2}{(z_1 + a_1)(z_2 + a_2)(z_1 - k_1)(z_2 - k_2)}. \tag{5.21}$$

Using (5.11) gives for (5.21), after some algebra and analytic continuation to the imaginary axes,

$$\left(\frac{\partial^2 I}{\partial k_d^2}\right)_{k_d=k_v} = \frac{1}{(2\pi i)^2}\int_{-i\infty}^{i\infty}\int_{-i\infty}^{i\infty} dz_1\, dz_2 2k_v^2 F(z_1, z_2),$$

$$\{F(z_1, z_2) = \{(z_1 - k_1)(z_2 - k_2)(k_v^2 - z_1^2)^{\frac{1}{2}}(k_v^2 - z_2^2)^{\frac{1}{2}}[z_1 - (k_v^2 - z_2^2)^{\frac{1}{2}}][z_2 - (k_v^2 - z_1^2)^{\frac{1}{2}}][a_1 + (k_v^2 - z_2^2)^{\frac{1}{2}}] \times [a_2 + (k_v^2 - z_1^2)^{\frac{1}{2}}]\}^{-1}$$

where $0 < \arg(k_i) < \pi/2$ and $0 < \arg(a_i) < \pi/2$, $i = 1, 2$. (5.22)

Next, we compute that

$$\lim_{\substack{|k_1|\to\infty \\ |k_2|\to\infty}} k_1 k_2\left(\frac{\partial^2 I}{\partial k_d^2}\right)_{k_d=k_v} = \frac{2k_v^2}{(2\pi i)^2}\int_{-i\infty}^{i\infty}\int_{-i\infty}^{i\infty} \frac{dz_1\, dz_2}{(z_1^2 + z_2^2 - k_v^2)^2[a_1 + (k_v^2 - z_2^2)^{\frac{1}{2}}][a_2 + (k_v^2 - z_1^2)^{\frac{1}{2}}]}. \tag{5.23}$$

Now let $z_1 = i\tau_1$ and $z_2 = i\tau_2$ in (5.23) to obtain

$$\lim_{\substack{|k_1|\to\infty \\ |k_2|\to\infty}} k_1 k_2 \left(\frac{\partial^2 I}{\partial k_d^2}\right)_{k_d=k_v} = \frac{k_v^2}{2\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{d\tau_1\, d\tau_2}{(\tau_1^2+\tau_2^2+k_v^2)^2[a_1+(1+\tau_2^2)^{\frac{1}{2}}][a_2+(1+\tau_1^2)^{\frac{1}{2}}]} \not\equiv 0. \tag{5.24}$$

It now follows, from the fact that (5.24) does not vanish identically, that

$$\left(\frac{\partial^2 I}{\partial k_d^2}\right)_{k_d=k_v} \not\equiv 0. \tag{5.25}$$

This proves (5.15) and establishes that (5.1) is not a solution of (5.4) or (2.33).

We conclude our paper with the computation of the inverse Fourier transform of (5.1) at the vertex of the wedge. It follows from (4.2)–(4.5) and (5.1) and (5.2) that

$$E(0,0) = \frac{E_0}{K_{--}(-a_1,-a_2)} = E_0 \exp\left\{\frac{-1}{(2\pi i)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{dz_1\, dz_2}{(z_1+a_1)(z_2+a_2)} \log\left[\frac{k_d^2-(z_1^2+z_2^2)}{k_v^2-(z_1^2+z_2^2)}\right]\right\}. \tag{5.26}$$

Analytically continuing (5.26) to the imaginary axes and transforming to polar coordinates noting (4.9) reduces (5.26) to

$$E(0,0) = E_0 \exp\left\{\frac{-1}{2\pi}\int_0^{\infty} \frac{\xi\, d\xi}{(1+\xi^2)} \log\left(\frac{\xi^2+p^2}{\xi^2+1}\right)\left(\frac{\sin\theta}{(\xi^2+\cos^2\theta)^{\frac{1}{2}}} + \frac{\cos\theta}{(\xi^2+\sin^2\theta)^{\frac{1}{2}}}\right)\right\}, \tag{5.27}$$

where $p = k_d/k_v$.

We now wish to compare the field at the vertex of wedge as computed correctly to second order in $(k_d^2 - k_v^2)$ in (4.14) with the corresponding terms obtained from (5.26). Power-series expanding the exponential in (5.26) yields

$$\begin{aligned} E(0,0) = E_0\Bigg\{1 &- \frac{1}{(2\pi i)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{dz_1\, dz_2}{(z_1+a_1)(z_2+a_2)} \log\left(\frac{k_d^2-(z_1^2+z_2^2)}{k_v^2-(z_1^2+z_2^2)}\right) \\ &+ \frac{1}{2}\left[\frac{1}{(2\pi i)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{dz_1\, dz_2}{(z_1+a_1)(z_2+a_2)} \log\left(\frac{k_d^2-(z_1^2+z_2^2)}{k_v^2-(z_1^2+z_2^2)}\right)\right]^2 - \cdots\Bigg\}, \end{aligned} \tag{5.28}$$

and power-series expanding the logarithm for k_d sufficiently near k_v gives

$$\log\left(\frac{k_d^2-(z_1^2+z_2^2)}{k_v^2-(z_1^2+z_2^2)}\right) = \frac{(k_d^2-k_v^2)}{k_v^2-(z_1^2+z_2^2)} - \frac{1}{2}\left(\frac{(k_d^2-k_v^2)}{k_v^2-(z_1^2+z_2^2)}\right)^2 + \cdots. \tag{5.29}$$

When (5.29) and (5.28) are combined and compared with (4.11) and (4.14), the first-order terms in $(k_d^2 - k_v^2)$ are seen to be identical, but the second-order terms in $(k_d^2 - k_v^2)$ differ. In particular, to second order, (5.29) and (5.28) yield

$$\begin{aligned} E^{(2)}(0,0) = E^{(1)}(0,0) + \frac{(k_d^2-k_v^2)^2}{2(2\pi i)^2} E_0 &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{dz_1\, dz_2}{(z_1+a_1)(z_2+a_2)[k_v^2-(z_1^2+z_2^2)]^2} + \frac{(k_d^2-k_v^2)^2}{2(2\pi i)^2} E_0 \\ &\times \left\{\frac{1}{2\pi i}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{dz_1 dz_2}{(z_1+a_1)(z_2+a_2)[k_v^2-(z_1^2+z_2^2)]}\right\}^2, \end{aligned} \tag{5.30}$$

which should be compared with (4.14). This constitutes an independent proof that Radlow's result (5.1) is incorrect. Finally, we remark that for grazing incidence, $\theta = 0$ or $\theta = \pi/2$, Eq. (5.27) can be evaluated exactly and yields

$$E(0,0) = \{2k_v/(k_d+k_v)\}^{\frac{1}{2}}. \tag{5.31}$$

This the reader will recognize as the square root of the transmission coefficient for a planewave normally incident on a dielectric half space.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Instantaneous Action-at-a-Distance Formulation of Classical Electrodynamics*

FREDERICK JAMES KENNEDY†
Department of Physics, University of Delaware, Newark, Delaware

(Received 25 July 1968)

The possibility of formally translating the interaction of charges from charge ↔ field ↔ charge to charge ↔ charge, where the orbits satisfy Newtonian (second order in t), yet covariant, equations of motion, is exploited for the Wheeler–Feynman interaction. A method for computing the forces on the charges correct to second order in the coupling constant e^2 is presented, and ten constants of the motion correct to e^2 are found. The integration is effected via the Noether theorem with the inhomogeneous Lorentz group as symmetry transformations. An important result is that a well-known correction to the Coulomb interaction which accounts for the uniform motion of charges is revealed to be, to first order in e^2, a frame-invariant expression. The consequent corrected Coulomb dynamics admits first-order integrals identical to those of the Wheeler–Feynman dynamics.

I. INTRODUCTION

Recent interest in the two-body problem of electrodynamics, which after almost eighty years[1] continues to defy solution, has prompted a critical reexamination of the objections against a purely Newtonian description of the interaction, that is, an instantaneous action-at-a-distance formulation. The objections have been found to be invalid.

Close inspection discloses that the relativistic proscription of instantaneous causal connections in nature does not logically debar equations correlating, with respect to some Lorentz frame, the state of a system of particles and the coincident acceleration of a member. Further, though Lorentz transformations properly describe the transformation of world points, the same set of world points would not be used in phrasing such correlations in different Lorentz frames. If, in addition to the Lorentz transformations, invariance of world lines of particles is demanded, the result is a transformation, not of particle variables at a fixed set of world points, but between the simultaneous particle variables of one frame and those of another. If one then enforces the requirement that dynamical equations be form invariant under such transformations, an explicit condition on the functional form of the forces results. For the case of two particles, the condition has been deduced independently by both Hill[2] and Currie.[3]

Special relativity, in other words, far from legislating against instantaneous action-at-a-distance formulations, rather provides the guidelines within which such formulations are viable.

For two particles interacting in any way, electrodynamically, mesonically, etc., the import of this revelation is that it is not, in principle, vain to seek a description of the interaction via Newtonian-order equations of motion (in the single independent variable t) whose solutions are twelve-parameter families of world lines. A Hamiltonian formulation, canonical in all Lorentz frames, may then be sought. Previously, the existence of such a Hamiltonian was thought to be denied by the zero-interaction theorem,[4] but it has since been found[5] that, if one does not insist that physical position variables be canonical, an essentially unique Hamiltonian formulation ensues with the inhomogeneous Lorentz group[6] canonically represented. Finally, quantization could perhaps be approached in the usual manner. The problems here, not yet fully appreciated as this is only a prospectus,

* This work forms part of a dissertation presented for the Ph.D. at the University of Delaware in 1967.

† Present address: Theoretical Physics Institute, The University of Alberta, Edmonton, Alberta, Canada.

[1] The problem's importance goes back to Heaviside's introduction in 1889 of the Lorentz force law, subsequently appended to Maxwell's equations by Lorentz to provide a theory in which electromagnetic effects derived from the dynamics of moving charges. For a compendious history of action-at-a-distance theories in physics, see M. B. Hesse, *Forces and Fields* (Philosophical Library, New York, 1962). For the immediate background to this present work, plus further documentation, see Footnotes 2–8 below.

[2] R. N. Hill, J. Math. Phys. **8**, 201 (1967).

[3] D. G. Currie, Phys. Rev. **142**, 817 (1966).

[4] D. G. Currie, T. F. Jordan, and E. C. G. Sudarshan, Rev. Mod. Phys. **35**, 350 (1963); D. G. Currie, J. Math. Phys. **4**, 1470 (1963); J. T. Cannon and T. F. Jordan, J. Math. Phys. **5**, 299 (1964); H. Leutwyler, Nuovo Cimento **37**, 556 (1965).

[5] R. N. Hill and E. H. Kerner, Phys. Rev. Letters **17**, 1156 (1966). See also E. H. Kerner, J. Math. Phys. **6**, 1218 (1965). Note that one is not prevented from finding Hamiltonian formulations with physical positions canonical; it is necessary only to relinquish the demand that the formulation be (canonically) equivalent in different Lorentz frames in order to do so. The difficulty is that the resulting quantum theory is thereby vitiated. The nonrelativistic version of this situation is graphically illustrated by considering a single particle moving in one dimension in the potential $\phi(x)$. Multitudes of canonically inequivalent Hamiltonians can be found, but the only one admitting a consistent quantum theory is that for which the Galilei transformations are canonical. For remarks about such classical-valid, quantum-invalid Hamiltonians, see P. Havas, Bull. Am. Phys. Soc. **1**, 337 (1956); F. J. Kennedy and E. H. Kerner, Am. J. Phys. **33**, 463 (1965); **34**, 271 (1966).

[6] "Inhomogeneous Lorentz group" throughout this paper means the usual group of space rotations and space and time translations, plus pure Lorentz transformations in the sense of Hill and Currie—ordinary Lorentz transformations plus world-point shifts along invariant world lines (just compensating for the noncovariance of simultaneity). See Sec. III, Eqs. (17d).

would probably entail methodological difficulties, as, for instance, how to Hermitize uniquely the Hamiltonian. But the conceptual framework would remain Newtonian; the procedure, presumably, would be basically on the same ground as that of quantizing a nonrelativistic two-body problem, but the issue would be the relativistic quantum theory of the interaction. In short, the different behavior of relativistic and nonrelativistic particles at both classical and quantal levels might possibly be accounted for at the classical outset, not by different formalisms, but by only one—the Newtonian.

This present work, confined to the classical rudiments of the program just outlined, focuses on the force law and constants of the motion for a particular two-body problem. Because of its utility and mathematical tractability, the Wheeler–Feynman interaction[7] of electrodynamics has been chosen for investigation. This half-advanced + half-retarded interaction, the time-symmetric character of which readily admits an action principle, is equivalent, under certain well-defined conditions, to the ordinary damped and retarded interaction of charges. Its investigation is therefore bound to reveal important characteristics of the forces, energies, etc., of high-energy charges interacting in the customary purely retarded way.

The first stage of this problem has already been attacked. Kerner[8] has invented a formal scheme for educing Newtonian equations of motion comprehending the full Wheeler–Feynman orbits from the field-particle dynamics. The method, briefly, is this:

A Lagrangian for one charge is assembled in the usual way, the potentials being expressed in terms of the other charge's kinematical behavior at times advanced and retarded with respect to time t reckoned at the first charge. Expansion of the second charge's motion about t yields a single-time Lagrangian which, however, because of the expansion, involves all orders of time derivatives of that charge's motion. When the steps are repeated with the charges exchanged, the final result is a pair of equations of motion which are themselves of infinite order. Further, it turns out that the interaction term of each of the one-charge Lagrangians can be symmetrized, thus providing a single two-charge Lagrangian and casting the infinite-order dynamics into formal canonical form.

The hypothesis is then made that all solutions of the dynamical equations are spurious, save those continuous with free-particle motions in the zero-interaction limit. There results an algorithm for extracting from the infinite-order equations of motion a second-order Newtonian set. The equations of motion finally emerge as infinite series in the coupling constant e^2 or, alternatively, in c^{-2}, the first of which is the more interesting because of the appearance of that parameter in the perturbative calculations of quantum electrodynamics.

In the first part of this paper the mathematical details of Kerner's method are recalled and departed from slightly. The departure is only one of procedure: If the basic (infinite-order) equations of motion are rewritten so that the forces appear as complex integrations about certain time singularities, some of the combinatorial frustrations, which otherwise haunt the method, disappear and terms beyond the opening (simple) e^2 term can be found in closed form. The computation is alternative, but equivalent, to the iteration of Hill's integrodifferential equation for the determination of the forces.[2]

The principal part of this discussion begins in the third section. Here the Noether theorem[9] for an infinite-degree-of-freedom dynamics is developed. After establishing that the inhomogeneous Lorentz group is canonically represented with respect to the two-charge Lagrangian assembled in Sec. II, application of the theorem produces three three-vector and one three-scalar constants of the motion. These are, of course, the momentum and energy—corresponding to the form invariance of the dynamical equations under space and time translation, the angular momentum—under space rotation, and the center-of-mass constant—under pure Lorentz transformation. Order reduction of these integrals by the method applied to the forces gives the integrals to first order in e^2 of the Newtonian equations of motion. It is then shown that these constants have the expected transformation character in shifting from one Lorentz frame to another.

Following this, in the fourth section, the connection between the results of the present investigation and the classical dynamics of charges interacting in the conventional retarded manner is examined. It develops that, since the equations of motion are identical to first order in the coupling constant for both the time-symmetric and purely retarded interactions, all first-order results derived for the former are just as valid for the latter. As a consequence, some new light is shed on an old result of classical electrodynamics.

II. EQUATIONS OF MOTION

Order Reduction

These opening remarks are a brief recapitulation of Kerner's scheme.[8]

[7] J. A. Wheeler and R. P. Feynman, Rev. Mod. Phys. **17**, 157 (1945); **21**, 425 (1949).

[8] E. H. Kerner, J. Math. Phys. **3**, 35 (1962); **6**, 1218 (1965).

[9] E. L. Hill, Rev. Mod. Phys. **23**, 253 (1951).

Imagine that two charges interact by half the sum of advanced and retarded fields. The Lagrangian for one of these, e_1 say, can be written as

$$L_1 = -m_1c^2(1 - v_1^2/c^2)^{\frac{1}{2}} - e_1e_2\sum_{p=0}^{\infty}\frac{D_2^{2p}}{(2p)!\,c^{2p}}\left(1 - \frac{\mathbf{v}_1\cdot\mathbf{v}_2}{c^2}\right)r^{2p-1} \quad (1)$$

by expanding about t the advanced and retarded Lienard–Wiechert potentials at the location of e_1 due to the motion of e_2. D_2 is the time derivative of e_2's variables only and r is the distance between the charges.

The equations of motion of e_1 are given by the Euler–Lagrange derivative of (1):

$$D\frac{m_1\mathbf{v}_1}{(1 - v_1^2)^{\frac{1}{2}}} = -e^2\left(\frac{\partial}{\partial\mathbf{r}_1} - D\frac{\partial}{\partial\mathbf{v}_1}\right)\times\sum_{p=0}^{\infty}\frac{D_2^{2p}}{(2p)!}(1 - \mathbf{v}_1\cdot\mathbf{v}_2)r^{2p-1},$$

where, of course, D means total time derivative $D_1 + D_2$; e^2 is e_1e_2, and the speed of light has been set equal to unity. The left-hand side may be written

$$(1 - v_1^2)^{-\frac{3}{2}}[(1 - v_1^2)\mathbf{I} + \mathbf{v}_1\mathbf{v}_1]\cdot m_1\dot{\mathbf{v}}_1,$$

and the dyadic inverted to yield the system

$$m_1\dot{\mathbf{v}}_1 = e^2(1 - v_1^2)^{\frac{1}{2}}(\mathbf{I} - \mathbf{v}_1\mathbf{v}_1)\cdot\left(D\frac{\partial}{\partial\mathbf{v}_1} - \frac{\partial}{\partial\mathbf{r}_1}\right)\sum_{p=0}^{\infty}\frac{D_2^{2p}}{(2p)!}(1 - \mathbf{v}_1\cdot\mathbf{v}_2)r^{2p-1}, \quad (2a)$$

$$m_2\dot{\mathbf{v}}_2 = e^2(1 - v_2^2)^{\frac{1}{2}}(\mathbf{I} - \mathbf{v}_2\mathbf{v}_2)\cdot\left(D\frac{\partial}{\partial\mathbf{v}_2} - \frac{\partial}{\partial\mathbf{r}_2}\right)\sum_{p=0}^{\infty}\frac{D_1^{2p}}{(2p)!}(1 - \mathbf{v}_1\cdot\mathbf{v}_2)r^{2p-1}. \quad (2b)$$

This result is merely a rewrite of the field-particle dynamics. In the same way one derives the corresponding system of equations for the completely retarded interaction. The distinction between the two is marked by the appearance of even-order derivatives only in the Wheeler–Feynman case, a consequence of its time-symmetric nature.

A further consequence of the time symmetry is that the individual (one-charge) Lagrangians may give way to a joint (two-charge) Lagrangian. This is easily seen by writing $D_2^{2p} = D_2^p(D - D_1)^p$ in L_1, for instance. Then $D_2^{2p} = (-D_1D_2)^p$ + exact derivatives. As usual, the latter may be neglected, thereby symmetrizing the interaction term. The joint Lagrangian is

$$L = -m_1(1 - v_1^2)^{\frac{1}{2}} - m_2(1 - v_2^2)^{\frac{1}{2}} - e^2\sum_{p=0}^{\infty}\frac{(-D_1D_2)^p}{(2p)!}(1 - \mathbf{v}_1\cdot\mathbf{v}_2)r^{2p-1}. \quad (3)$$

It is easily established that the equations of motion (2) will now follow by application of the generalized Euler–Lagrange, or "Ostrogradsky," operator to L.

The ambiguities associated with the infinite degrees of freedom of the interaction become quite manifest in its mechanical formulation. It would seem, for example, that the initial data $\mathbf{r}_i(0)$, $\mathbf{v}_i(0)$, $\dot{\mathbf{v}}_i(0)\cdots$ for the problem amount to nothing less than the solution itself, so that the system (2), rather than being equations of motion, is merely a *constraint* on the motion. To deal with this extravagance it is hypothesized that admissible orbits continue to be determined by the specification of twelve parameters and tend smoothly towards straight lines as the interaction is shut off. [Or, in other words, the admissible solutions of (2) satisfy a pair of Newtonian-order equations with the forces expansible as a power series in e^2.] Such a system may be extracted from (2) by arranging its right-hand side as a multinomial expansion in time derivatives of the velocities as, for instance,

$$m_1\ddot{x}_1 = e^2\left(f + \sum_{i,k} f_{ik}v_k^{(i)} + \sum_{i,j,k,l} f_{ijkl}v_k^{(i)}v_l^{(j)} + \cdots\right). \quad (4)$$

The $v_k^{(i)}$ (the ith derivative of v_k) are then computed in terms of all the other derivatives via (2) itself, substituted in (4), and the result again arranged in the form of (4). The upshot of this process, when continued ad infinitum, is that the right-hand side of (2) is replaced with an infinite series in e^2 whose coefficients are functions of positions and velocities; the solutions of this reduced system are among those of (2) but analytic at $e^2 = 0$.

Although criteria of convergence, if any, have yet to be elaborated, it will be seen that the validity of the calculations to come do not hang upon the convergence of the e^2 expansions; they are meaningful even if these expansions are asymptotic.

The Forces to Order e^4

One of the difficulties in effecting the reduction has been the summation of the infinite number of pieces going into a given multinomial term. Thus, the only easily obtained term has been the opening one, giving the equations of motion

$$m_1\dot{\mathbf{v}}_1 = e^2(1 - v_1^2)^{\frac{1}{2}}(\mathbf{I} - \mathbf{v}_1\mathbf{v}_1)\cdot(1 - v_2^2)\frac{\hat{\mathbf{r}} - \mathbf{v}_1\times(\hat{\mathbf{r}}\times\mathbf{v}_2)}{r^2[1 - (\hat{\mathbf{r}}\times\mathbf{v}_2)^2]^{\frac{3}{2}}} + e^4\cdots, \quad (5a)$$

$$m_2\dot{\mathbf{v}}_2 = -e^2(1 - v_2^2)^{\frac{1}{2}}(\mathbf{I} - \mathbf{v}_2\mathbf{v}_2)\cdot(1 - v_1^2)\frac{\hat{\mathbf{r}} - \mathbf{v}_2\times(\hat{\mathbf{r}}\times\mathbf{v}_1)}{r^2[1 - (\hat{\mathbf{r}}\times\mathbf{v}_1)^2]^{\frac{3}{2}}} + e^4\cdots, \quad (5b)$$

a familiar result (to be discussed in Sec. IV). These first terms are just what the forces on each charge

would be if the other were constrained to move uniformly, whence the name "straight-line approximation" for the opening terms alone.

Here we develop a formal summation of the forces from which the individual e^2 terms can be recovered systematically, at least in low orders. The calculations, which will be illustrated to e^4 order, are equivalent to the iteration of Hill's integrodifferential equation.[2] This latter equation, satisfied by the forces, derives from the fact that the entire force can be induced from a knowledge of the *opening* e^2 (or c^{-2}) term by the strictures of Lorentz covariance alone, whereas the idea to be presented now comes straight from the Lienard–Wiechert potentials.

The primitive dynamics (2) is first re-expressed (working with just the first charge) as

$$\begin{aligned} m_1\dot{\mathbf{v}}_1 = e^2(1-v_1^2)^{\frac{1}{2}}(\mathbf{I}-\mathbf{v}_1\mathbf{v}_1) \\ \cdot\sum_{p=0}^{\infty}\frac{1}{2\pi i}\oint\frac{1}{(z-t)^{2p+1}} \\ \times\{[\mathbf{v}_2\mathbf{r}\cdot\mathbf{v}_2+\mathbf{v}_1\times(\mathbf{r}\times\mathbf{v}_2)-\mathbf{r}] \\ \times(2p-1)r^{2p-3}-\dot{\mathbf{v}}_2 r^{2p-1}\}\,dz, \end{aligned} \tag{6}$$

where charge-2 variables in the integrand are functions of z. This comes about by carrying out the Euler–Lagrange operation in (2), and representing the D_2's by the Cauchy integral formula

$$D_2^{2p}f(\mathbf{r},\mathbf{v}_1,\mathbf{v}_2)=\frac{(2p)!}{2\pi i}\oint\frac{f(\mathbf{r}_1-\mathbf{r}_2(z),\mathbf{v}_1,\mathbf{v}_2(z))}{(z-t)^{2p+1}}\,dz.$$

If it is assumed that the order of integration and summation may be switched, the series in (6) may be summed with the result

$$\begin{aligned} m_1\dot{\mathbf{v}}_1 = e^2(1-v_1^2)^{\frac{1}{2}}(\mathbf{I}-\mathbf{v}_1\mathbf{v}_1) \\ \cdot\frac{1}{2\pi i}\Bigg[\oint\frac{(z-t)(3r^2-(z-t)^2)}{r^3[(z-t)^2-r^2]^2} \\ \times(\mathbf{v}_2\mathbf{r}\cdot\mathbf{v}_2+\mathbf{v}_1\times(\mathbf{r}\times\mathbf{v}_2)-\mathbf{r})\,dz \\ -\oint\frac{(z-t)\dot{\mathbf{v}}_2}{r[(z-t)^2-r^2]}\,dz\Bigg]. \end{aligned}$$

Next, with the change of variable $s=z-t$, charge-2's argument is shifted; the finite-shift operator is then introduced, and the equations of motion of the first charge become finally

$$\begin{aligned} m_1\dot{\mathbf{v}}_1 = e^2(1-v_1^2)^{\frac{1}{2}}(\mathbf{I}-\mathbf{v}_1\mathbf{v}_1)\frac{1}{2\pi i} \\ \cdot\Bigg\{\oint\exp(sD_2)\Bigg[\frac{s(3r^2-s^2)}{r^3(s^2-r^2)^2} \\ \times(\mathbf{v}_2\mathbf{r}\cdot\mathbf{v}_2+\mathbf{v}_1\times(\mathbf{r}\times\mathbf{v}_2)-\mathbf{r})\Bigg]\,ds \\ -\oint\exp(sD_2)\frac{s\dot{\mathbf{v}}_2}{r(s^2-r^2)}\,ds\Bigg\}. \end{aligned} \tag{7}$$

Thus, once again, all charge variables are functions of t.

The problem of summing the individual multinomial terms of (2) is now replaced with the more tractable one of summing such terms in the finite-shift operator. The ordering, or "disentangling," of such operators has already been investigated by Feynman,[10] whose operator calculus may be applied here with the result that

$$\begin{aligned} &\exp(sD_2) \\ &=\exp\left[s\left(\mathbf{v}_2\cdot\frac{\partial}{\partial\mathbf{r}_2}+\dot{\mathbf{v}}_2\cdot\frac{\partial}{\partial\mathbf{v}_2}+\mathbf{v}_2^{(2)}\cdot\frac{\partial}{\partial\dot{\mathbf{v}}_2}+\cdots\right)\right] \\ &=\left[\exp\left(s\mathbf{v}_2\cdot\frac{\partial}{\partial\mathbf{r}_2}\right)\right]\left[1+\int_{\eta=0}^{s}\int_{\xi=0}^{\eta}\dot{\mathbf{v}}_2(t+\xi)\right. \\ &\quad\left.\cdot\frac{\partial}{\partial\mathbf{r}_2}\,d\xi\,d\eta+\int_{\xi=0}^{s}\dot{\mathbf{v}}_2(t+\xi)\cdot\frac{\partial}{\partial\mathbf{v}_2}\,d\xi\right. \\ &\quad\left.+\dot{\mathbf{v}}_2(t+s)\cdot\frac{\partial}{\partial\dot{\mathbf{v}}_2}-\dot{\mathbf{v}}_2(t)\cdot\frac{\partial}{\partial\dot{\mathbf{v}}_2}\right] \\ &\quad+\dot{v}\dot{v}\text{ type terms}\ldots\text{etc.} \end{aligned} \tag{8}$$

The details of this expansion are found in Appendix A. Here, linear terms giving no contribution in (7) have already been discarded.

The recovery of the e^2 forces of (5) from (7) provides a simple illustration of the computations. To this order the accelerations in (7) and (8) are put to zero and (7) becomes simply

$$\begin{aligned} m_1\dot{\mathbf{v}}_1 = e^2(1-v_1^2)^{\frac{1}{2}}(\mathbf{I}-\mathbf{v}_1\mathbf{v}_1)\frac{1}{2\pi i} \\ \cdot\oint\frac{s[3(\mathbf{r}-\mathbf{v}_2 s)^2-s^2]}{|\mathbf{r}-\mathbf{v}_2 s|^3\,[s^2-(\mathbf{r}-\mathbf{v}_2 s)^2]^2} \\ \times[\mathbf{v}_2\mathbf{r}\cdot\mathbf{v}_2+\mathbf{v}_1\times(\mathbf{r}\times\mathbf{v}_2)-\mathbf{r}+\mathbf{v}_2 s(1-v_2^2)]\,ds. \end{aligned}$$

The integration encircles both roots of $[s^2-(\mathbf{r}-\mathbf{v}_2 s)^2]$, that is, s_1 and s_2 where

$$s_{\frac{1}{2}}=-\frac{\mathbf{r}\cdot\mathbf{v}_2}{1-v_2^2}\pm\frac{r}{1-v_2^2}[1-(\hat{\mathbf{r}}\times\mathbf{v}_2)^2]^{\frac{1}{2}}. \tag{9}$$

Then the integral may be written

$$\begin{aligned} \frac{1}{(1-v_2^2)}\frac{\partial^2}{\partial s_1\partial s_2}\oint\frac{s[3(\mathbf{r}-\mathbf{v}_2 s)^2-s^2]}{|\mathbf{r}-\mathbf{v}_2 s|^3\,(s-s_1)(s-s_2)} \\ \times[\mathbf{v}_2\mathbf{r}\cdot\mathbf{v}_2+\mathbf{v}_1\times(\mathbf{r}\times\mathbf{v}_2)-\mathbf{r}+\mathbf{v}_2 s(1-v_2^2)]\,ds \end{aligned}$$

and rapidly evaluated. (Note that $s_{1,2}=\pm|\mathbf{r}-\mathbf{v}_2 s_{1,2}|$, and that these relations are used after the differentiation with respect to $s_{1,2}$.) The equations of motion of

[10] R. P. Feynman, Phys. Rev. **84**, 108 (1951); W. L. Miranker and B. Weiss, SIAM Review **8**, 224 (1966).

the first charge emerge as

$$m_1\dot{\mathbf{v}}_1 = -e^2\frac{4(1-v_1^2)^{\frac{1}{2}}}{(1-v_2^2)(s_1-s_2)^3}(\mathbf{I}-\mathbf{v}_1\mathbf{v}_1)\cdot\{2[\mathbf{v}_2\mathbf{r}\cdot\mathbf{v}_2+\mathbf{v}_1\times(\mathbf{r}\times\mathbf{v}_2)-\mathbf{r}]+\mathbf{v}_2(1-v_2^2)(s_1+s_2)\},$$

which upon substitution of the roots (9) is found to be identically (5a). This is the gist of the method.

The second-order term is given by

$$e^2(1-v_1^2)(\mathbf{I}-\mathbf{v}_1\mathbf{v}_1)\frac{1}{2\pi i}\cdot\left\{\oint\left[\exp\left(s\mathbf{v}_2\cdot\frac{\partial}{\partial\mathbf{r}_2}\right)\right]\times\left[\int_{\eta=0}^{s}\int_{\xi=0}^{\eta}\dot{\mathbf{v}}_2(t+\xi)\cdot\frac{\partial}{\partial\mathbf{r}_2}\,d\xi\,d\eta+\int_{\xi=0}^{s}\dot{\mathbf{v}}_2(t+\xi)\cdot\frac{\partial}{\partial\mathbf{v}_2}\,d\xi\right]\times\frac{s(3r^2-s^2)}{r^3(s^2-r^2)^2}[\mathbf{v}_2\mathbf{r}\cdot\mathbf{v}_2+\mathbf{v}_1\times(\mathbf{r}\times\mathbf{v}_2)-\mathbf{r}]\,ds-\oint\exp\left(s\mathbf{v}_2\cdot\frac{\partial}{\partial\mathbf{r}_2}\right)\frac{s\dot{\mathbf{v}}_2(t+s)}{r(s^2-r^2)}\,ds\right\}.$$

The calculation is not essentially different from that of the first order, but there are a few cautionary points that it would be well to emphasize. The sequence of operations is: (a) derivatives with respect to $\mathbf{r}_2$ and $\mathbf{v}_2$, excepting the finite shift; (b) then the finite shift; (c) $\dot{\mathbf{v}}_2$ is evaluated at $t+\xi$, but since it is of order e^2 itself and the calculation is to be of order e^4, this amounts to replacing $\mathbf{r}$ with $\mathbf{r}+\mathbf{v}\xi$ in the first-order equations (5). Note that $\dot{\mathbf{v}}_2(t+\xi)$ is not an operand in steps (a) or (b). At this point there remains no ambiguity; there is just the indicated single and double integrals of $\dot{\mathbf{v}}_2(t+\xi)$ and evaluation of the complex integrals. The poles turn out to be the same as those in the example above but of higher order.

III. CONSTANTS OF THE MOTION

The infinite-order equations of motion (2) can be integrated to yield constants interpretable as energy, momentum, and so on, even though such a step seems purely formal. The order of the system is not reduced as with a pair of finite-order equations; one merely has a new relation among, again, an infinitude of time derivatives. Nevertheless the integrals of (2) are important, for they contain the integrals of the order-reduced Newtonian equations determining the admissible solutions of (2). The latter integrals are just those of (2) delimited by the relations among the time derivatives implied by the Newtonian equations themselves.

Two ways thus appear for integrating the Newtonian dynamics: Work out the forces term by term, as in Sec. II, and integrate directly; or integrate the antecedent infinite-order dynamics and extract the Newtonian constants as special cases. The second way seems the more circuitous at first sight, but, since it happens that the delimitation of the infinite-order integrals may be effected by order reduction, the Newtonian constants may be secured without prior determination of the force law. Moreover, unlike the case for the second-order equations of motion, a Lagrangian (3) is already in hand for the infinite-order system, and this alone makes available the most efficacious of integration schemes, namely Noether's theorem.[9]

In what follows, Noether's theorem is developed for an infinite-degree-of-freedom dynamics and is applied to the system (2). There results ten constants of the motion, corresponding to the ten parameters of the inhomogeneous Lorentz group, which order reduction then renders Newtonian. This last step is carried out only to terms of order e^2 so that the final products are constants (to e^2) of the equations of motion in straight-line approximation.

Noether's Theorem

Let there be equations of motion

$$\sum_{n=0}^{\infty}(-D')^n\frac{\partial L(D'^l\mathbf{r}_j')}{\partial(D'^n\mathbf{r}_i')}=0,\quad i=1,2,\qquad(9')$$

following from the action principle

$$\delta\int_a^b L(D'^l\mathbf{r}_j')\,dt'=0\qquad(10)$$

in which the Lagrangian is explicitly independent of the time. If the equations of motion are form invariant under the infinitesimal transformations

$$\mathbf{r}_i'=\mathbf{r}_i+\epsilon\mathbf{f}_i(\mathbf{r}_j,\mathbf{v}_j,t),\qquad t'=t+\epsilon g(\mathbf{r}_j,\mathbf{v}_j,t),\qquad(11)$$

then the equations of motion in unprimed coordinates are just (9′) without primes, which means that (10) written without primes necessarily holds.

Now, the transformation rule for time derivatives is quickly induced from (11):

$$D'^n\mathbf{r}_i'=D^n\mathbf{r}_i+\epsilon[D^n(\mathbf{f}_i-\mathbf{v}_ig)+gD^n\mathbf{v}_i].$$

Introducing this in (10) and expanding to first order in ϵ, one finds

$$\delta\int_c^d\left(L(D^n\mathbf{r}_i)+\epsilon\left\{L(D^n\mathbf{r}_i)Dg+\sum_{n=0}^{\infty}\sum_{i=1}^{2}\frac{\partial L}{\partial(D^n\mathbf{r}_i)}\cdot[D^n(\mathbf{f}_i-\mathbf{v}_ig)+gD^n\mathbf{v}_i]\right\}\right)dt=0,\qquad(12)$$

where c and d are the transformed, but still fixed, limits. As pointed out above, the lowest-order piece vanishes owing to the form invariance of (9). Furthermore, the first and last terms in the ϵ piece go together to make the exact time derivative $D(Lg)$, the variation of which is identically zero. Result: A necessary and sufficient condition that (11) are symmetry transformations of the equations of motion is

$$\delta\int_c^d \sum_{n,i} \frac{\partial L}{\partial(D^n\mathbf{r}_i)} \cdot D^n(\mathbf{f}_i - \mathbf{v}_i g)\, dt = 0. \tag{13}$$

The Lagrangian and the transformations cooperate in deciding how this happens. When the integrand of (13) is an exact derivative, (13) is an identity, and the transformations are, by definition, canonical. In any more general case the vanishing is conditional; for example, the integrand could itself be any Lagrangian for the system, with the equations of motion being the condition. But for the present purpose the important and relevant case is the former, wherefore

$$\sum_{n,i} \frac{\partial L}{\partial(D^n\mathbf{r}_i)} \cdot D^n(\mathbf{f}_i - \mathbf{v}_i g) = D\Omega, \tag{14}$$

which may be called the "canonicity condition" for the infinitesimal symmetry transformations (11).

Assuming that (14) is satisfied, the equations of motion (9′), without primes, are dotted with $(\mathbf{f}_i - \mathbf{v}_i g)$, summed on i from 1 to 2, and the result subtracted from the left-hand side of (14), giving

$$\sum_{n,i}\left\{\frac{\partial L}{\partial(D^n\mathbf{r}_i)} \cdot D^n(\mathbf{f}_i - \mathbf{v}_i g) - \left[(-D)^n \frac{\partial L}{\partial(D^n\mathbf{r}_i)}\right] \cdot (\mathbf{f}_i - \mathbf{v}_i g)\right\} = D\Omega. \tag{15}$$

By the rule

$$D\sum_{l=0}^{n} (D^{n-l}\mathbf{B}) \cdot (-D)^l \mathbf{A} = \mathbf{A} \cdot D^{n+1}\mathbf{B} - \mathbf{B} \cdot (-D)^{n+1}\mathbf{A},$$

the curly braces on the left-hand side of (15) contain an exact derivative. The integral

$$\sum_{n,i}\sum_{l=0}^{n} D^{n-l}(\mathbf{f}_i - \mathbf{v}_i g) \cdot (-D)^l \frac{\partial L}{\partial(D^n\mathbf{v}_i)} - \Omega = \text{const} \tag{16}$$

immediately results. This constitutes Noether's theorem for a system containing all orders of time derivatives—one constant for each canonically represented symmetry transformation of the equations of motion. The association is not one-to-one, however, as the replacing of $\mathbf{f}_i$ and g with $\mathbf{f}_i + \mathbf{v}_i h$ and $g + h$, with h arbitrary, leaves the constant unchanged.

Infinite-Order Integrals

The dynamical equations (2) implied by the Lagrangian (3) have no explicit time dependence, have *relative* coordinate dependence only, and are three-vector equations. They are thus form invariant under the following:

$$\text{time translation (TT): } \mathbf{r}_i' = \mathbf{r}_i, \quad t' = t + \epsilon, \tag{17a}$$

$$\text{space translation (ST): } \mathbf{r}_i' = \mathbf{r}_i + \epsilon\hat{\mathbf{n}}, \quad t' = t, \tag{17b}$$

$$\text{space rotation (SR): } \mathbf{r}_i' = \mathbf{r}_i + \epsilon\hat{\boldsymbol{\theta}} \times \mathbf{r}_i, \quad t' = t, \tag{17c}$$

respectively, where the unit vectors are fixed in the directions of translation and rotation. An Ω exists for each of these cases and can be found from (14), although for these particular symmetries, it is easier to refer back to (12). The Lagrangian is manifestly invariant under the above transformations (being a time-independent, relative-coordinate-dependent scalar), so that the ϵ piece in the integrand of (12)—not only its variation—vanishes; since that piece is just $D(Lg + \Omega)$, then $\Omega = -Lg$, and, by comparing (11) with (17), the three Ω's follow as

$$\Omega_{\mathrm{TT}} = -L, \tag{18a}$$

$$\Omega_{\mathrm{ST}} = 0, \tag{18b}$$

$$\Omega_{\mathrm{SR}} = 0. \tag{18c}$$

Finally, the transformation between infinitesimally different Lorentz frames is[2,3,6]

$$\begin{aligned} \mathbf{r}_1' &= \mathbf{r}_1 + \epsilon[\mathbf{v}_1\hat{\mathbf{n}} \cdot (\mathbf{r}_1 - \mathbf{r}_2) - \hat{\mathbf{n}}t], \\ \mathbf{r}_2' &= \mathbf{r}_2 - \epsilon\hat{\mathbf{n}}t, \\ t' &= t - \epsilon\hat{\mathbf{n}} \cdot \mathbf{r}_2, \end{aligned} \tag{17d}$$

where $\epsilon\hat{\mathbf{n}}$ is the velocity of the primed with respect to the unprimed frame. Though it is not so obvious that an Ω exists, the transformations (17d) are canonical with respect to the Lagrangian (3), and the result of the computation of Ω from (14), carried out in Appendix B, is

$$\Omega_{\mathrm{LT}} = -m_1\hat{\mathbf{n}} \cdot \mathbf{r}_1(1 - v_1^2)^{\frac{1}{2}} - m_2\hat{\mathbf{n}} \cdot \mathbf{r}_2(1 - v_2^2)^{\frac{1}{2}} - \frac{e^2}{2}\sum_{p=0}^{\infty}\frac{(-D_1D_2)^p}{(2p)!}(1 - \mathbf{v}_1 \cdot \mathbf{v}_2)r^{2p-1}\hat{\mathbf{n}} \cdot (\mathbf{r}_1 + \mathbf{r}_2). \tag{18d}$$

The corresponding integrals are

$$C_{\mathrm{TT}} = L - \sum_{n,i}\sum_{l=0}^{n}(D^{n-l}\mathbf{v}_i)\cdot(-D)^l\frac{\partial L}{\partial(D^n\mathbf{v}_i)}, \tag{19a}$$

$$\mathbf{C}_{\mathrm{ST}} = \sum_{n,i}(-D)^n\frac{\partial L}{\partial(D^n\mathbf{v}_i)}, \tag{19b}$$

$$\mathbf{C}_{\mathrm{SR}} = \sum_{n,i}\sum_{l=0}^{n}(D^{n-l}\mathbf{r}_i)\times(-D)^l\frac{\partial L}{\partial(D^n\mathbf{v}_i)}, \tag{19c}$$

$$\mathbf{C}_{\mathrm{LT}} = \sum_{n,i}\sum_{l=0}^{n}[D^{n-l}(\mathbf{r}_i\mathbf{v}_i - \mathbf{I}t)]\cdot(-D)^l\frac{\partial L}{\partial(D^n\mathbf{v}_i)} - \mathbf{\Omega}_{\mathrm{LT}}, \tag{19d}$$

where $\Omega_{\mathrm{LT}} = \mathbf{\Omega}_{\mathrm{LT}}\cdot\hat{\mathbf{n}}$.

The first of these but for the sign can be recognized as the extension to the infinite-degree-of-freedom case of the usual energy. The last three vector integrals are the linear momentum, angular momentum, and the so-called "center-of-mass" constant, the generalization of $M\mathbf{R}_{cm} - \mathbf{P}t$ going with the Galilei transformation in the nonrelativistic case.

To complete the computation the integrals must be rendered functions of the primitive variables. The expression for the derivative of the Lagrangian with respect to $D^n\mathbf{v}_i$ [Eq. (B3) in Appendix B—just shift n to $n+1$] is substituted in each of the constants and the results simplified. For example, for the linear momentum,

$$\begin{aligned}\mathbf{P} = \sum_{n,i}(-D)^n\Big[&m_i\mathbf{v}_i(1-v_i^2)^{-\frac{1}{2}}\delta_{0n}\\ &- e^2\sum_{p=n}^{\infty}\frac{(-D_j)^{p+1}D_i^{p-n}}{(2p+2)!}(2p+1)\\ &\times\binom{p+1}{n+1}(1-\mathbf{v}_1\cdot\mathbf{v}_2)r^{2p-1}(\mathbf{r}_i-\mathbf{r}_j)\\ &+ e^2\sum_{p=n}^{\infty}\frac{(-D_j)^pD_i^{p-n}}{(2p)!}\binom{p}{n}\mathbf{v}_jr^{2p-1}\Big],\end{aligned}$$

where $j = 2$ when $i = 1$, and vice-versa. The counting in the second term may start with $n = -1$ since the term vanishes at that number. Then shifting the index, $n \to n+1$, and using the counting rule

$$\sum_{n=0}^{\infty}\sum_{p=n}^{\infty} = \sum_{p=0}^{\infty}\sum_{n=0}^{p}$$

leads to

$$\begin{aligned}\mathbf{P} = &\sum_i m_i\mathbf{v}_i(1-v_i^2)^{-\frac{1}{2}}\\ &+\frac{e^2}{D}\sum_{p,i}\frac{(D_j)^{2(p+1)}}{(2p+2)!}(2p+1)\\ &\times(1-\mathbf{v}_1\cdot\mathbf{v}_2)r^{2p-1}(\mathbf{r}_i-\mathbf{r}_j)\\ &+e^2\sum_{p,i}\frac{(D_j)^{2p}}{(2p)!}\mathbf{v}_jr^{2p-1},\end{aligned}$$

where, in addition, the binomial expansion

$$(-D_j)^k = (D_i - D)^k = \sum_{n=0}^{k}\binom{k}{n}(-D)^nD_i^{k-n}$$

has been used in the last two terms. In the middle term, however, this last step can only be effected if the term is first differentiated ("multiplied" by $-D$); it must then be integrated, here represented as division by D. The representation is unambiguous since D can be factored out of the middle term and the previous expression recovered, a fact more visible when the above is summed on i:

$$\begin{aligned}\mathbf{P} = &m_1\mathbf{v}_1(1-v_1^2)^{-\frac{1}{2}} + m_2\mathbf{v}_2(1-v_2^2)^{-\frac{1}{2}}\\ &+e^2\sum_{p=0}^{\infty}\frac{(2p-1)}{(2p)!}\frac{1}{D}(D_2^{2p}-D_1^{2p})(1-\mathbf{v}_1\cdot\mathbf{v}_2)r^{2p-3}\mathbf{r}\\ &+e^2\sum_{p=0}^{\infty}\frac{1}{(2p)!}(D_2^{2p}\mathbf{v}_2r^{2p-1}+D_1^{2p}\mathbf{v}_1r^{2p-1}).\end{aligned} \tag{20a}$$

The other integrals may be similarly expressed, and are found to be

$$E = m_1(1-v_1^2)^{-\frac{1}{2}} + m_2(1-v_2^2)^{-\frac{1}{2}} + e^2\sum_{p=0}^{\infty}\frac{1}{(2p)!}(D_1^{2p}+D_2^{2p})r^{2p-1} - e^2\sum_{p=0}^{\infty}\frac{1}{(2p)!}\frac{1}{D}(D_1^{2p+1}+D_2^{2p+1})(1-\mathbf{v}_1\cdot\mathbf{v}_2)r^{2p-1}, \tag{20b}$$

$$\mathbf{L} = m_1(\mathbf{r}_1\times\mathbf{v}_1)(1-v_1^2)^{-\frac{1}{2}} + m_2(\mathbf{r}_2\times\mathbf{v}_2)(1-v_2^2)^{-\frac{1}{2}} + e^2\sum_{p=0}^{\infty}\frac{1}{(2p)!}[D_2^{2p}(\mathbf{r}_1\times\mathbf{v}_2)r^{2p-1}+D_1^{2p}(\mathbf{r}_2\times\mathbf{v}_1)r^{2p-1}] + e^2\sum_{p=0}^{\infty}\frac{1}{(2p)!}\frac{1}{D}(D_1^{2p}-D_2^{2p})[(\mathbf{v}_1\times\mathbf{v}_2)r^{2p-1} + (2p-1)(1-\mathbf{v}_1\cdot\mathbf{v}_2)(\mathbf{r}_1\times\mathbf{r}_2)r^{2p-3}], \tag{20c}$$

$$\begin{aligned}\mathbf{K} = &m_1\mathbf{r}_1(1-v_1^2)^{-\frac{1}{2}} + m_2\mathbf{r}_2(1-v_2^2)^{-\frac{1}{2}} - \mathbf{P}t\\ &+e^2\sum_{p=0}^{\infty}\frac{1}{(2p)!}\frac{1}{D}\{\mathbf{r}_1D_2^{2p}[D_1r^{2p-1}+D_2(\mathbf{v}_1\cdot\mathbf{v}_2)r^{2p-1}] + \mathbf{r}_2D_1^{2p}[D_2r^{2p-1}+D_1(\mathbf{v}_1\cdot\mathbf{v}_2)r^{2p-1}]\}\\ &+e^2\sum_{p=0}^{\infty}\frac{1}{(2p)!}\frac{1}{D}(D_1^{2p}\mathbf{v}_1r^{2p-1}+D_2^{2p}\mathbf{v}_2r^{2p-1}) + e^2\sum_{p=0}^{\infty}\frac{1}{(2p)!}\frac{1}{D^2}(D_2^{2p}-D_1^{2p})(2p-1)(1-\mathbf{v}_1\cdot\mathbf{v}_2)r^{2p-3}\mathbf{r}.\end{aligned} \tag{20d}$$

As in the linear momentum, the inverse D's cancel the numerator D's and are therefore integrations. It will presently be seen, however, that actual integration, attempted order by order in e^2, is ambiguous due to the possibility of adding any term whose derivative is of order e^{2n+2} to the integral computed to e^{2n} accuracy. A test of the admissibility of such additive pieces is to reintroduce the speed of light into (20) wherever it dimensionally belongs [in (20a), for example, c^{-2p} follows each summation sign, c^2 divides each quadratic velocity term], and then to compare the expansion in c^{-2} powers of the direct integration result with the c^{-2} expansion of (20). The latter is unique owing to D's explicit and remainderless division into the operand following it.

Newtonian Integrals To Order e^2

The order depression proceeds exactly as before, with the first step the expression of the constants in complex integral form. Just as was the force law, the linear momentum (again singled out for illustration) may be written

$$\mathbf{P} = m_1\mathbf{v}_1(1-v_1^2)^{-\frac{1}{2}} + m_2\mathbf{v}_2(1-v_2^2)^{-\frac{1}{2}} + e^2\frac{1}{2\pi i}\frac{1}{D}\frac{\partial}{\partial \mathbf{r}}\oint[\exp(sD_2)-\exp(sD_1)]\frac{(1-\mathbf{v}_1\cdot\mathbf{v}_2)}{r(s^2-r^2)}\,ds$$
$$+ e^2\frac{1}{2\pi i}\oint\left[\exp(sD_2)\frac{\mathbf{v}_2}{r(s^2-r^2)}+\exp(sD_1)\frac{\mathbf{v}_1}{r(s^2-r^2)}\right]s\,ds. \quad (21)$$

To find $\mathbf{P}$ correctly to order e^2, the finite-shift operators are adequately represented by the first term in the expansion (8). Thus, the only complex integrals that appear are of the type

$$\oint \exp\left(s\mathbf{v}_j\cdot\frac{\partial}{\partial \mathbf{r}_j}\right)\frac{s}{r(s^2-r^2)}\,ds, \quad (22)$$

which, for $j=2$ say, is

$$\frac{2\pi i}{(1-v_2^2)(s_1-s_2)}\left[\frac{s_1}{|\mathbf{r}-\mathbf{v}_2 s_1|}-\frac{s_2}{|\mathbf{r}-\mathbf{v}_2 s_2|}\right],$$

just matching the analogous force computation. The roots s_1 and s_2 are as before (9) and, again, the denominators in the square bracket are, respectively, s_1 and $-s_2$. The integral is therefore

$$\frac{2\pi i}{r[1-(\hat{\mathbf{r}}\times\mathbf{v}_2)^2]^{\frac{1}{2}}}. \quad (23)$$

The gradient of this quantity is

$$-2\pi i\frac{\hat{\mathbf{r}}-\mathbf{v}_2\times(\hat{\mathbf{r}}\times\mathbf{v}_2)}{r^2[1-(\hat{\mathbf{r}}\times\mathbf{v}_2)^2]^{\frac{3}{2}}},$$

and the time integral of the gradient, needed in the second term of $\mathbf{P}$, may be written

$$-2\pi i\int_{t_0}^{t}\frac{\mathbf{r}_0+\mathbf{v}t-\mathbf{v}_2\times[(\mathbf{r}_0+\mathbf{v}t)\times\mathbf{v}_2]}{\{(\mathbf{r}_0+\mathbf{v}t)^2-[(\mathbf{r}_0+\mathbf{v}t)\times\mathbf{v}_2]^2\}^{\frac{3}{2}}}\,dt, \quad (24)$$

that is, $\mathbf{r}$ is considered to change uniformly. A more accurate calculation would contribute terms at least of order e^2 in the integral or of order e^4 in the momentum. The integration is straightforward. It is guessed, and confirmed later by the final form of $\mathbf{P}$, that the result vanishes at t_0, so that upon eliminating $\mathbf{r}_0$ (by $\mathbf{r}_0=\mathbf{r}-\mathbf{v}t$), one has for (24)

$$-2\pi i\mathbf{N}r^{-1}\{(\hat{\mathbf{r}}\times\mathbf{v})^2-[(\hat{\mathbf{r}}\times\mathbf{v}_1)\times(\hat{\mathbf{r}}\times\mathbf{v}_2)]^2\}^{-1}$$
$$\times[1-(\hat{\mathbf{r}}\times\mathbf{v}_2)^2]^{-\frac{1}{2}}, \quad (25)$$

where

$$\mathbf{N}=\hat{\mathbf{r}}[\hat{\mathbf{r}}\cdot\mathbf{v}-(\hat{\mathbf{r}}\times\mathbf{v}_2)\cdot(\mathbf{v}_1\times\mathbf{v}_2)]$$
$$+\mathbf{v}_1[1-(\hat{\mathbf{r}}\times\mathbf{v}_2)^2]$$
$$+\mathbf{v}_2[1-(\hat{\mathbf{r}}\times\mathbf{v}_1)\cdot(\hat{\mathbf{r}}\times\mathbf{v}_2)].$$

Finally, these fragments, (23) and (25) for $j=2$, along with their exchanged-subscripts counterparts, are assembled according as they appear in (21). The result is to order e^2 the linear momentum associated with the admissible solutions of (2):

$$\mathbf{P}=\frac{m_1\mathbf{v}_1}{(1-v_1^2)^{\frac{1}{2}}}+\frac{m_2\mathbf{v}_2}{(1-v_2^2)^{\frac{1}{2}}}+\frac{e^2}{r}\left[\frac{\mathbf{v}_1}{(1-\mu_1^2)^{\frac{1}{2}}}+\frac{\mathbf{v}_2}{(1-\mu_2^2)^{\frac{1}{2}}}\right]+\frac{e^2(1-\mathbf{v}_1\cdot\mathbf{v}_2)}{r[(\mu_1-\mu_2)^2-(\mu_1\times\mu_2)^2]}$$
$$\times\left(\frac{1}{(1-\mu_2^2)^{\frac{1}{2}}}\{\mathbf{v}_1(1-\mu_2^2)-\mathbf{v}_2(1-\mu_1\cdot\mu_2)-\hat{\mathbf{r}}[\hat{\mathbf{r}}\cdot\mathbf{v}-\mu_2\cdot(\mathbf{v}_1\times\mathbf{v}_2)]\}\right.$$
$$\left.+\frac{1}{(1-\mu_1^2)^{\frac{1}{2}}}\{\mathbf{v}_2(1-\mu_1^2)-\mathbf{v}_1(1-\mu_1\cdot\mu_2)+\hat{\mathbf{r}}[\hat{\mathbf{r}}\cdot\mathbf{v}-\mu_1\cdot(\mathbf{v}_1\times\mathbf{v}_2)]\}\right), \quad (26a)$$

in which $\hat{\mathbf{r}} \times \mathbf{v}_1 = \boldsymbol{\mu}_1$, $\hat{\mathbf{r}} \times \mathbf{v}_2 = \boldsymbol{\mu}_2$ for short, and $\mathbf{v}$ is the relative velocity $\mathbf{v}_1 - \mathbf{v}_2$.

An expansion of $\mathbf{P}$ (after putting c back in) in powers of c^{-2} is identical with the same expansion of (20a) with velocities held constant, thereby ratifying the choice of lower limit.

The other constants of the motion are handled in the same way. The only little puzzle in each case concerns the proper choice of t_0 in the time-integration step, there apparently being no systematic way for knowing what it should be ahead of time. The simplest guess, that the time integral should vanish at the lower limit, works for the linear momentum and also, as it turns out, for the energy; but it fails for the angular momentum and for the center-of-mass constant. These latter two are each distinguished by requiring among their time integrations a type significantly different from the form of (24). For integrations in which the curly-braced expression appears just as in (24), in the denominator and to the $\frac{3}{2}$ power, the above rule for t_0 applies—the integral vanishes at the lower limit. In all other cases that curly-braced expression occurs also in the denominators, but to the $\frac{1}{2}$ power. Then, although the vanishing rule fails, a factor of the form

$$\ln\left[\{(\mathbf{r}_0 + \mathbf{v}t)^2 - [(\mathbf{r}_0 + \mathbf{v}t) \times \mathbf{v}_2]^2\}^{\frac{1}{2}} + t[v^2 - (\mathbf{v}_1 \times \mathbf{v}_2)^2]^{\frac{1}{2}} + \frac{\mathbf{r}_0 \cdot \mathbf{v} - (\mathbf{r}_0 \times \mathbf{v}_2) \cdot (\mathbf{v}_1 \times \mathbf{v}_2)}{[v^2 - (\mathbf{v}_1 \times \mathbf{v}_2)^2]^{\frac{1}{2}}}\right],$$

arising when each of these remaining time integrations are evaluated, hints at another simple t_0, namely, that which gives zero for the curly brace itself. The conjecture is reinforced when elimination of $\mathbf{r}_0$, as in (25), results in the logarithm's argument being a function constant to e^2 with no explicit time dependence, and is established by a c^{-2} expansion test of the final $\mathbf{L}$ and $\mathbf{K}$.

The energy, angular momentum, and center-of-mass constant are, to e^2 order, respectively,

$$E = \frac{m_1}{(1 - v_1^2)^{\frac{1}{2}}} + \frac{m_2}{(1 - v_2^2)^{\frac{1}{2}}} + \frac{e^2}{r}\left[\frac{1}{(1 - \mu_1^2)^{\frac{1}{2}}} + \frac{1}{(1 - \mu_2^2)^{\frac{1}{2}}}\right] + \frac{e^2(1 - \mathbf{v}_1 \cdot \mathbf{v}_2)}{r[(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^2 - (\boldsymbol{\mu}_1 \times \boldsymbol{\mu}_2)^2]} \times \left[\frac{\boldsymbol{\mu}_2 \cdot (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}{(1 - \mu_2^2)^{\frac{1}{2}}} - \frac{\boldsymbol{\mu}_1 \cdot (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}{(1 - \mu_1^2)^{\frac{1}{2}}}\right], \tag{26b}$$

$$\begin{aligned}\mathbf{L} = {} & \frac{m_1(\mathbf{r}_1 \times \mathbf{v}_1)}{(1 - v_1^2)^{\frac{1}{2}}} + \frac{m_2(\mathbf{r}_2 \times \mathbf{v}_2)}{(1 - v_2^2)^{\frac{1}{2}}} + \frac{e^2}{r}\left[\frac{\mathbf{r}_1 \times \mathbf{v}_2}{(1 - \mu_2^2)^{\frac{1}{2}}} + \frac{\mathbf{r}_2 \times \mathbf{v}_1}{(1 - \mu_1^2)^{\frac{1}{2}}}\right] + \frac{e^2(1 - \mathbf{v}_1 \cdot \mathbf{v}_2)}{r[(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^2 - (\boldsymbol{\mu}_1 \times \boldsymbol{\mu}_2)^2]} \\ & \times \left(\frac{1}{(1 - \mu_2^2)^{\frac{1}{2}}}\{(\mathbf{r}_2 \times \mathbf{v}_1)(1 - \mu_2^2) - (\mathbf{r}_1 \times \mathbf{v}_2)(1 - \boldsymbol{\mu}_1 \cdot \boldsymbol{\mu}_2) + (1/r)(\mathbf{r}_1 \times \mathbf{r}_2)[\hat{\mathbf{r}} \cdot \mathbf{v} - \boldsymbol{\mu}_2 \cdot (\mathbf{v}_1 \times \mathbf{v}_2)]\}\right. \\ & \left. + \frac{1}{(1 - \mu_1^2)^{\frac{1}{2}}}\{(\mathbf{r}_1 \times \mathbf{v}_2)(1 - \mu_1^2) - (\mathbf{r}_2 \times \mathbf{v}_1)(1 - \boldsymbol{\mu}_1 \cdot \boldsymbol{\mu}_2) - (1/r)(\mathbf{r}_1 \times \mathbf{r}_2)[\hat{\mathbf{r}} \cdot \mathbf{v} - \boldsymbol{\mu}_1 \cdot (\mathbf{v}_1 \times \mathbf{v}_2)]\}\right) \\ & + \frac{e^2(\mathbf{v}_1 \times \mathbf{v}_2)(1 - \mathbf{v}_1 \cdot \mathbf{v}_2)}{[(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^2 - (\boldsymbol{\mu}_1 \times \boldsymbol{\mu}_2)^2][v^2 - (\mathbf{v}_1 \times \mathbf{v}_2)^2]}\{[\hat{\mathbf{r}} \cdot \mathbf{v} - \boldsymbol{\mu}_1 \cdot (\mathbf{v}_1 \times \mathbf{v}_2)](1 - \mu_2^2)^{\frac{1}{2}} \\ & - [\hat{\mathbf{r}} \cdot \mathbf{v} - \boldsymbol{\mu}_2 \cdot (\mathbf{v}_1 \times \mathbf{v}_2)](1 - \mu_1^2)^{\frac{1}{2}}\} + e^2\frac{(1 - v_1^2)(1 - v_2^2)(\mathbf{v}_1 \times \mathbf{v}_2)}{[v^2 - (\mathbf{v}_1 \times \mathbf{v}_2)^2]^{\frac{3}{2}}} \\ & \times \ln\left\{\left(\frac{1 - v_1^2}{1 - v_2^2}\right)^{\frac{1}{2}}\left[\frac{(1 - \mu_2^2)^{\frac{1}{2}}[v^2 - (\mathbf{v}_1 \times \mathbf{v}_2)^2]^{\frac{1}{2}} + \hat{\mathbf{r}} \cdot \mathbf{v} - \boldsymbol{\mu}_2 \cdot (\mathbf{v}_1 \times \mathbf{v}_2)}{(1 - \mu_1^2)^{\frac{1}{2}}[v^2 - (\mathbf{v}_1 \times \mathbf{v}_2)^2]^{\frac{1}{2}} + \hat{\mathbf{r}} \cdot \mathbf{v} - \boldsymbol{\mu}_1 \cdot (\mathbf{v}_1 \times \mathbf{v}_2)}\right]\right\},\end{aligned} \tag{26c}$$

$$\begin{aligned}\mathbf{K} = {} & \frac{m_1\mathbf{r}_1}{(1 - v_1^2)^{\frac{1}{2}}} + \frac{m_2\mathbf{r}_2}{(1 - v_2^2)^{\frac{1}{2}}} - \mathbf{P}t + \frac{e^2}{r}\left[\frac{\mathbf{r}_1}{(1 - \mu_2^2)^{\frac{1}{2}}} + \frac{\mathbf{r}_2}{(1 - \mu_1^2)^{\frac{1}{2}}}\right] + \frac{e^2(1 - \mathbf{v}_1 \cdot \mathbf{v}_2)}{r[(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^2 - (\boldsymbol{\mu}_1 \times \boldsymbol{\mu}_2)^2]} \\ & \times \left\{\mathbf{r}_1(1 - \mu_1^2)^{\frac{1}{2}} + \mathbf{r}_2(1 - \mu_2^2)^{\frac{1}{2}} - (1 - \boldsymbol{\mu}_1 \cdot \boldsymbol{\mu}_2)\left[\frac{\mathbf{r}_1}{(1 - \mu_2^2)^{\frac{1}{2}}} + \frac{\mathbf{r}_2}{(1 - \mu_1^2)^{\frac{1}{2}}}\right]\right\} \\ & + \frac{e^2(1 - \mathbf{v}_1 \cdot \mathbf{v}_2)\mathbf{v}}{[(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^2 - (\boldsymbol{\mu}_1 \times \boldsymbol{\mu}_2)^2][v^2 - (\mathbf{v}_1 \times \mathbf{v}_2)^2]}\{[\hat{\mathbf{r}} \cdot \mathbf{v} - \boldsymbol{\mu}_1 \cdot (\mathbf{v}_1 \times \mathbf{v}_2)](1 - \mu_2^2)^{\frac{1}{2}} \\ & - [\hat{\mathbf{r}} \cdot \mathbf{v} - \boldsymbol{\mu}_2 \cdot (\mathbf{v}_1 \times \mathbf{v}_2)](1 - \mu_1^2)^{\frac{1}{2}}\} + \frac{e^2(1 - v_1^2)(1 - v_2^2)\mathbf{v}}{[v^2 - (\mathbf{v}_1 \times \mathbf{v}_2)^2]^{\frac{3}{2}}} \\ & \times \ln\left\{\left(\frac{1 - v_1^2}{1 - v_2^2}\right)^{\frac{1}{2}}\left[\frac{(1 - \mu_2^2)^{\frac{1}{2}}[v^2 - (\mathbf{v}_1 \times \mathbf{v}_2)^2]^{\frac{1}{2}} + \hat{\mathbf{r}} \cdot \mathbf{v} - \boldsymbol{\mu}_2 \cdot (\mathbf{v}_1 \times \mathbf{v}_2)}{(1 - \mu_1^2)^{\frac{1}{2}}[v^2 - (\mathbf{v}_1 \times \mathbf{v}_2)^2]^{\frac{1}{2}} + \hat{\mathbf{r}} \cdot \mathbf{v} - \boldsymbol{\mu}_1 \cdot (\mathbf{v}_1 \times \mathbf{v}_2)}\right]\right\}.\end{aligned} \tag{26d}$$

A direct check by time differentiation verifies that this set of functions (26) is indeed constant in straight-line approximation, the accelerations being determined by the opening terms in (5).

Lorentz Transformation of the Newtonian Integrals

It should finally be noted that, to within their order of accuracy, the constants of the motion (26) have the required transformation character when passing from one Lorentz frame to another. The quantity $P^2 - E^2$ is a four-scalar, and **K** and **L** are the components of an antisymmetric four-tensor, the transformation rule for infinitesimally distinguished frames being

$$\mathbf{L}' = \mathbf{L} + \epsilon\hat{\mathbf{n}} \times \mathbf{K}, \tag{27a}$$

$$\mathbf{K}' = \mathbf{K} - \epsilon\hat{\mathbf{n}} \times \mathbf{L}. \tag{27b}$$

These statements may be tested by brute-force transformation or, more elegantly, in the same way that (5) may be proven covariant,[2,3] by developing certain necessary and sufficient conditions that they be true. For example, by direct transformation, using (17d),

$$\begin{aligned}\mathbf{L}' &= \mathbf{L}(\mathbf{r}_1', \mathbf{r}_2'; \mathbf{v}_1', \mathbf{v}_2') \\ &= \mathbf{L}[\mathbf{r}_1 + \epsilon(\mathbf{v}_1\hat{\mathbf{n}}\cdot\mathbf{r} - \hat{\mathbf{n}}t), \quad \mathbf{r}_2 - \epsilon\hat{\mathbf{n}}t; \\ &\quad \mathbf{v}_1 + \epsilon(\mathbf{v}_1\hat{\mathbf{n}}\cdot\mathbf{v}_1 - \hat{\mathbf{n}} + \dot{\mathbf{v}}_1\hat{\mathbf{n}}\cdot\mathbf{r}), \mathbf{v}_2 + \epsilon(\mathbf{v}_2\hat{\mathbf{n}}\cdot\mathbf{v}_2 - \hat{\mathbf{n}})],\end{aligned}$$

which to first order is

$$\begin{aligned}\mathbf{L}(\mathbf{r}_1, \mathbf{r}_2; \mathbf{v}_1, \mathbf{v}_2) + \epsilon\Big[&(\mathbf{v}_1\hat{\mathbf{n}}\cdot\mathbf{r} - \hat{\mathbf{n}}t)\cdot\frac{\partial\mathbf{L}}{\partial\mathbf{r}_1} \\ &- (\hat{\mathbf{n}}t)\cdot\frac{\partial\mathbf{L}}{\partial\mathbf{r}_2} + (\mathbf{v}_1\hat{\mathbf{n}}\cdot\mathbf{v}_1 - \hat{\mathbf{n}} + \dot{\mathbf{v}}_1\hat{\mathbf{n}}\cdot\mathbf{r})\cdot\frac{\partial\mathbf{L}}{\partial\mathbf{v}_1} \\ &+ (\mathbf{v}_2\hat{\mathbf{n}}\cdot\mathbf{v}_2 - \hat{\mathbf{n}})\cdot\frac{\partial\mathbf{L}}{\partial\mathbf{v}_2}\Big].\end{aligned}$$

Equating this to the right-hand side of (27a) gives

$$\begin{aligned}\hat{\mathbf{n}}\cdot\mathbf{r}\left(\mathbf{v}_1\cdot\frac{\partial\mathbf{L}}{\partial\mathbf{r}_1} + \dot{\mathbf{v}}_1\cdot\frac{\partial\mathbf{L}}{\partial\mathbf{v}_1}\right) &+ \hat{\mathbf{n}}\cdot\mathbf{v}_1\mathbf{v}_1\cdot\frac{\partial\mathbf{L}}{\partial\mathbf{v}_1} \\ &+ \hat{\mathbf{n}}\cdot\mathbf{v}_2\mathbf{v}_2\cdot\frac{\partial\mathbf{L}}{\partial\mathbf{v}_2} - \hat{\mathbf{n}}\cdot\left(\frac{\partial\mathbf{L}}{\partial\mathbf{v}_1} + \frac{\partial\mathbf{L}}{\partial\mathbf{v}_2}\right) \\ &- (\hat{\mathbf{n}}t)\cdot\left(\frac{\partial\mathbf{L}}{\partial\mathbf{r}_1} + \frac{\partial\mathbf{L}}{\partial\mathbf{r}_2}\right) = \hat{\mathbf{n}} \times \mathbf{K}.\end{aligned}$$

This result may be reduced further by factoring out the $\hat{\mathbf{n}}$. When **L** and **K** are introduced from (26), one has a dyadic equation in which sixteen linearly independent pieces can be identified, all of which, upon the substitution of $\dot{\mathbf{v}}_1$ from (5a), have been found to be satisfied to order e^2.

The transformation rules for all the constants have been tried out in this manner, the error in each case being a term of e^4 order.

IV. CONVENTIONAL ELECTRODYNAMICS

The treatment so far has been a sort of paper-and-pencil experiment attempting to see some of the actual features of a covariant Newtonian dynamics. In this last section we briefly comment on the overlap of the results of the "experiment" with what ought to result from the conventional retarded interaction of two charges.

The overlap can be perceived in the Eqs. (5): The right-hand side of each, to first order, is the force on one charge in time-symmetric interaction with its uniformly moving companion. But that same term is also, and independently, the force [excepting the relativistic factor $(1 - v_i^2)^{\frac{1}{2}}(\mathbf{I} - \mathbf{v}_i\mathbf{v}_i)$] on the one charge in *retarded* interaction with the other, moving uniformly—a commonplace calculation.[11]

In this present context, the latter result can be retrieved by repeating the steps of the first section starting with fully retarded potentials. The outcome will be an e^2 expansion of the forces differing from the time-symmetric case in every term but the first. Important consequences are as follows:

(a) The well-known extension of Coulomb's law to include uniform motion of charges yields an *interaction* approximately covariant in the sense already mentioned.

(b) Although like constants of the motion (energy, etc.) will differ in the two interactions, their coupling constant expansions will be identical to first order. Consequently, the integrals to e^2 of the dynamics which result from replacing Coulomb's law with its straight-line extension are just those presented in the last section, Eqs. (26).

Of course, as they contain no power of the coupling constant already neglected in the forces, it is just the approximate integrals of the corrected Coulomb dynamics which are physically consistent. Nevertheless it is interesting to regard that dynamics—the system (5) with e^4 and higher terms dropped—as exact, and to seek exact integrals; an example of this is a kind of internal angular momentum

$$\mathfrak{L} = \frac{\mathbf{r} \times \mathbf{v}}{(1 - v_1^2)^{\frac{1}{2}}(1 - v_2^2)^{\frac{1}{2}}}.$$

Because the coupling constant is missing, $\mathfrak{L}$ also happens to be an approximate integral of the full dynamics. When terms beyond e^2 are included in the forces, its time derivative will be of order e^4.

What can be said about the domain of validity of the approximation to the completely retarded dynamics by the straight-line equations of motion? Unhappily, little at present, but reflection on this concluding question raises interesting possibilities. One can see by dimensional analysis that, when the

[11] R. Becker, *Electromagnetic Fields and Interactions*, F. Sauter, Ed. (Blaisdell Publ. Co., New York, 1964), Vol. I, pp. 267–271.

masses are the same, the expansion in e^2 of either time-symmetric or completely retarded forces is (excepting a common factor of e^2/r^2) in fact an expansion in powers of the classical charge radius r_0 divided by the separation of the charges.

Then, because approximations to the fully retarded forces obtained by truncating the e^2 series depend upon both the size of c compared with charge speeds and r_0 with charge separations, the error introduced by keeping the opening term alone [which is also the opening term of (5)] of each charge's force will be minimized only by an appropriate set of restrictions on the charges' speeds and closest-approach distance. The meaning of "appropriate" will become clear when the behavior of higher-order terms at various speeds has been understood; so far, it has only been established[2] that, for the one-dimensional case, the second and third terms in the force law have logarithmic branch points, but no poles, at the speed of light. The implication is that the opening term of (5) might successfully comprehend the dynamics of such events as the collision of high-energy charges, the sole condition being the specification of some minimum impact parameter.

V. SUMMARY

For a pair of charges, application of the criterion that the world lines should straighten out as the coupling constant is turned off leads to frame-invariant Newtonian equations of motion, thus singling out a twelve-parameter set of world lines from among the infinitude admitted by classical electrodynamics.

The computation of the forces to e^4 order has been reduced to the evaluation of certain integrals. Either completely retarded or time-symmetric interactions may be treated, but as the latter both follow from an action principle and are form invariant under the inhomogeneous Lorentz group, Noether's theorem applies and delivers ten constants of the motion. The time-symmetric case has therefore been investigated here and the integrals presented explicitly to first order in e^2.

The equations of motion are covariant and the ten integrals are possessed of their characteristic transformation rules only approximately, that is, consistent to their degree of approximation in e^2. Only when terms of all orders in e^2 are summed will the dynamics be completely consistent with the principle of relativity.

All of the results, however, remain formal because the question of convergence of the expansions in e^2 is still unresolved. Until the problem is settled the whole regime is tentative; in the meantime such facts as the opening terms in the c^{-2} expansion of the forces being the Coulomb and Darwin interactions, respectively, suggest that the e^2 terms also have perturbative significance and that the expansions in that parameter might at least turn out to be asymptotic.

The opening term in the e^2 expansion of the forces for the fully retarded and time-symmetric cases is identical and is, furthermore, the classical result (multiplied by a relativistic kinematical factor) of computing the force on one charge via the retarded fields of a second whose velocity is fixed. This straight-line correction to Coulomb's law is then seen to be rather more significant than ordinarily remarked. It correctly approximates to e^2, in a frame-invariant way, the force on one of a pair of charges in retarded interaction. Finally, owing to the coincidence of all first-order results, the constants of the motion of the corrected Coulomb dynamics are just those of the time-symmetric equations.

A Hamiltonian formulation, with the inhomogeneous Lorentz-group canonical, of the dynamics sketched in this presentation is known to exist and to be unique. The next major step will be its construction. One can already anticipate from the approximate covariance of the theory that even to order e^2 the injunction of the zero-interaction theorem will only be short-circuited by eliminating physical positions as candidates for the q_i's. Although a commutation-rule puzzle then looms, it was not expected that the marriage of the canonical formulation of the instantaneous interaction of charges with the special theory of relativity would take place gratuitously. We just hope that the union proves fruitful.

ACKNOWLEDGMENTS

I am greatly obliged to Dr. Edward H. Kerner for his generous help and encouragement throughout the course of this work, and to the National Science Foundation for its partial support of the research.

APPENDIX A: ORDERING OF $\exp(sD_2)$

It has been shown[10] by Feynman's calculus for the noncommuting operators A, B that

$$\begin{aligned}\exp[s(A+B)] &= \exp(sA) + s\int_{\alpha=0}^{1} \exp[(1-\alpha)sA]B\exp(\alpha sA)\,d\alpha \\ &\quad + s^2\int_{\beta=0}^{1}\exp[(1-\beta)sA]B \\ &\quad \times \int_{\alpha=0}^{\beta}\exp[(\beta-\alpha)sA]B\exp(\alpha sA)\,d\alpha\,d\beta + \cdots,\end{aligned} \tag{A1}$$

where s is a c number.

Now, if $D_2 = A + B$ and

$$A = \mathbf{v}_2 \cdot \frac{\partial}{\partial \mathbf{r}_2},$$

$$B = \dot{\mathbf{v}}_2 \cdot \frac{\partial}{\partial \mathbf{v}_2} + \cdots + \mathbf{v}_2^{(n+1)} \cdot \frac{\partial}{\partial \mathbf{v}_2^{(n)}} + \cdots,$$

n being the order of the time derivative, the commutation rule

$$B \exp(ksA) = \exp(ksA)\left(ks\dot{\mathbf{v}}_2 \cdot \frac{\partial}{\partial \mathbf{r}_2} + B\right) \quad \text{(A2)}$$

is obeyed (k is another c number). Using (A2) in (A1) gives

$$\exp(sD_2) = \exp(sA) + \exp(sA)s\int_{\alpha=0}^{1}\left(\alpha s\dot{\mathbf{v}}_2 \cdot \frac{\partial}{\partial \mathbf{r}_2} + B\right) d\alpha + \exp(sA)s^2\int_{\beta=0}^{1}\int_{\alpha=0}^{\beta}\left(\beta s\dot{\mathbf{v}}_2 \cdot \frac{\partial}{\partial \mathbf{r}_2} + B\right) \times \left(\alpha s\dot{\mathbf{v}}_2 \cdot \frac{\partial}{\partial \mathbf{r}_2} + B\right) d\alpha\, d\beta + \cdots. \quad \text{(A3)}$$

When each of the integrands is expanded, the result can be arranged in a multinomial series in time derivatives of the velocities. The contribution to the linear terms by each integrand in (A3) is easily found by induction and is

$$\alpha s\mathbf{v}_2^{(n)} \cdot \frac{\partial}{\partial \mathbf{r}_2} + \mathbf{v}_2^{(n)} \cdot \frac{\partial}{\partial \mathbf{v}_2} + \mathbf{v}_2^{(n+1)} \cdot \frac{\partial}{\partial \dot{\mathbf{v}}_2} + \cdots, \quad \text{(A4)}$$

in which the order of the time derivative n is also equal to the number of products in the integrand; or, in other words, (A4) is the linear term for the integrand of s^n in (A3), whence

$$\exp(sD_2) = \exp(sA)\left[1 + s\int_{\alpha=0}^{1}\left(\alpha s\dot{\mathbf{v}}_2 \cdot \frac{\partial}{\partial \mathbf{r}_2} + B\right) d\alpha + s^2\int_{\beta=0}^{1}\int_{\alpha=0}^{\beta}\left(\alpha s\mathbf{v}_2^{(2)} \cdot \frac{\partial}{\partial \mathbf{r}_2} + \mathbf{v}_2^{(2)} \cdot \frac{\partial}{\partial \mathbf{v}_2} + \cdots\right) d\alpha\, d\beta + \cdots\right].$$

Integration gives

$$\exp(sD_2) = \exp(sA)\left\{1 + s\left(\frac{1}{2!}s\dot{\mathbf{v}}_2 \cdot \frac{\partial}{\partial \mathbf{r}_2} + B\right) + s^2\left[\frac{1}{3!}s\mathbf{v}_2^{(2)} \cdot \frac{\partial}{\partial \mathbf{r}_2} + \frac{1}{2}\left(\mathbf{v}_2^{(2)} \cdot \frac{\partial}{\partial \mathbf{v}_2} + \mathbf{v}_2^{(3)} \cdot \frac{\partial}{\partial \dot{\mathbf{v}}_2} + \cdots\right)\right] + \cdots\right\}. \quad \text{(A5)}$$

The coefficients of $\partial/\partial \mathbf{r}_2$ and $\partial/\partial \mathbf{v}_2$ are

$$\frac{1}{2!}s^2\dot{\mathbf{v}}_2 + \frac{1}{3!}s^3\mathbf{v}_2^{(2)} + \frac{1}{4!}s^4\mathbf{v}_2^{(3)} + \cdots$$

and

$$s\dot{\mathbf{v}}_2 + \frac{1}{2!}s^2\mathbf{v}_2^{(2)} + \frac{1}{3!}s^3\mathbf{v}_2^{(3)} + \cdots,$$

which are, respectively,

$$\int_{\eta=0}^{s}\int_{\xi=0}^{\eta}\dot{\mathbf{v}}_2(t+\xi)\, d\xi\, d\eta$$

and

$$\int_{\xi=0}^{s}\dot{\mathbf{v}}_2(t+\xi)\, d\xi.$$

The remaining terms in the { } of (A5) are

$$s\left(\mathbf{v}_2^{(2)} \cdot \frac{\partial}{\partial \dot{\mathbf{v}}_2} + \cdots + \mathbf{v}_2^{(l+1)} \cdot \frac{\partial}{\partial \mathbf{v}_2^{(l)}} + \cdots\right) + \frac{1}{2!}s^2\left(\mathbf{v}_2^{(3)} \cdot \frac{\partial}{\partial \dot{\mathbf{v}}_2} + \cdots + \mathbf{v}_2^{(l+2)} \cdot \frac{\partial}{\partial \mathbf{v}_2^{(l)}} + \cdots\right) + \cdots,$$

which may all be written

$$\sum_{n=1}^{\infty}\left(\mathbf{v}_2^{(n)}(t+s) \cdot \frac{\partial}{\partial \mathbf{v}_2^{(n)}} - \mathbf{v}_2^{(n)} \cdot \frac{\partial}{\partial \mathbf{v}_2^{(n)}}\right).$$

Thus, to terms linear in the time derivatives of the velocities,

$$\exp(sD_2) = \exp\left(s\mathbf{v}_2 \cdot \frac{\partial}{\partial \mathbf{r}_2}\right) \times \left[1 + \int_{\eta=0}^{s}\int_{\xi=0}^{\eta}\dot{\mathbf{v}}_2(t+\xi) \cdot \frac{\partial}{\partial \mathbf{r}_2}\, d\xi\, d\eta + \int_{\xi=0}^{s}\dot{\mathbf{v}}_2(t+\xi) \cdot \frac{\partial}{\partial \mathbf{v}_2}\, d\xi + \sum_{n=1}^{\infty}(\mathbf{v}_2^{(n)}(t+s) - \mathbf{v}_2^{(n)}) \cdot \frac{\partial}{\partial \mathbf{v}_2^{(n)}}\right]. \quad \text{(A6)}$$

In the present application the operand of (A6) contains no time derivatives of $\mathbf{v}_2$ beyond the first. Therefore, only $n = 1$ appears in (8).

APPENDIX B: COMPUTATION OF Ω_{LT}

It is first necessary to find an expression for the derivative of the Lagrangian (3) with respect to $D^n\mathbf{r}_i$. By the Cauchy integral formula, with the transformation $z = s + t$, the Lagrangian may be written

$$L = -m_1(1 - v_1^2)^{\frac{1}{2}} - m_2(1 - v_2^2)^{\frac{1}{2}} - e^2\sum_{p=0}^{\infty}\frac{(-D_2)^p}{(2p)!}\frac{p!}{2\pi i} \times \oint \frac{[1 - \mathbf{v}_2 \cdot \mathbf{v}_1(s+t)]}{s^{p+1}}|\mathbf{r}_1(s+t) - \mathbf{r}_2|^{2p-1}\, ds. \quad \text{(B1)}$$

The form (B1) is suitable for taking the derivative with respect to $D^n\mathbf{r}_1$:

$$\frac{\partial L}{\partial(D^n\mathbf{r}_1)} = \frac{m_1\mathbf{v}_1}{(1-v_1^2)^{\frac{1}{2}}}\delta_{n1} - e^2\sum_{p=0}^{\infty}\frac{(-D_2)^p}{(2p)!}\left(\frac{p!}{2\pi i}\oint\frac{1}{s^{p+1}} \times\left\{[1-\mathbf{v}_2\cdot\mathbf{v}_1(s+t)](2p-1)\frac{s^n}{n!} \times[\mathbf{r}_1(s+t)-\mathbf{r}_2]\,|\mathbf{r}_1(s+t)-\mathbf{r}_2|^{2p-3} - \mathbf{v}_2\frac{s^{n-1}}{(n-1)!}\,|\mathbf{r}_1(s+t)-\mathbf{r}_2|^{2p-1}\right\}ds\right), \tag{B2}$$

where the last term in the curly braces is zero for $n = 0$. Retransforming to z and again using the integral formula gives for the term in parentheses

$$\frac{p!}{n!\,(p-n)!}(2p-1)D_1^{p-n}(1-\mathbf{v}_1\cdot\mathbf{v}_2)r^{2p-3}\mathbf{r} - \frac{p!}{(n-1)!\,(p-n+1)!}\mathbf{v}_2 D_1^{p-n+1}r^{2p-1}$$

So, with factorials of negative numbers denoting vanishing terms, (B2) becomes

$$\frac{\partial L}{\partial(D^n\mathbf{r}_1)} = \frac{m_1\mathbf{v}_1}{(1-v_1^2)^{\frac{1}{2}}}\delta_{n1} - e^2\sum_{p=n}^{\infty}\frac{(-D_2)^p D_1^{p-n}}{(2p)!}(2p-1) \times\binom{p}{n}(1-\mathbf{v}_1\cdot\mathbf{v}_2)r^{2p-3}\mathbf{r} + e^2\sum_{p=n-1}^{\infty}\frac{(-D_2)^p D_1^{p-n+1}}{(2p)!}\binom{p}{n-1}\mathbf{v}_2 r^{2p-1}, \tag{B3}$$

which is the desired expression. The derivative of L with respect to $D^n\mathbf{r}_2$ follows by exchange of indices.

Now, in the equation for the determination of any Ω, Eq. (14), it suffices to treat $i = 1$; then, with $\mathbf{f}_i$ and g given by (17d),

$$(D\Omega_{\mathrm{LT}})_{i=1} = \sum_{n=0}^{\infty}\frac{\partial L}{\partial(D^n\mathbf{r}_1)}\cdot D^n(\mathbf{v}_1\hat{\mathbf{n}}\cdot\mathbf{r}_1 - \hat{\mathbf{n}}t). \tag{B4}$$

Substituting (B3) in (B4) gives

$$(D\Omega_{\mathrm{LT}})_1 = \frac{m_1\mathbf{v}_1}{(1-v_1^2)^{\frac{1}{2}}}\cdot D(\mathbf{v}_1\hat{\mathbf{n}}\cdot\mathbf{r}_1 - \hat{\mathbf{n}}t) - e^2\sum_{n=0}^{\infty}\left[\sum_{p=n}^{\infty}\frac{(-D_2)^p D_1^{p-n}}{(2p)!}(2p-1)\binom{p}{n}(1-\mathbf{v}_1\cdot\mathbf{v}_2)r^{2p-3}\mathbf{r}\right]\cdot D^n(\mathbf{v}_1\hat{\mathbf{n}}\cdot\mathbf{r}_1 - \hat{\mathbf{n}}t) + e^2\sum_{n=1}^{\infty}\left[\sum_{p=n-1}^{\infty}\frac{(-D_2)^p D_1^{p-n+1}}{(2p)!}\binom{p}{n-1}\mathbf{v}_2 r^{2p-1}\right]\cdot D^n(\mathbf{v}_1\hat{\mathbf{n}}\cdot\mathbf{r}_1 - \hat{\mathbf{n}}t). \tag{B5}$$

The first term is the exact derivative of

$$-m_1\hat{\mathbf{n}}\cdot\mathbf{r}_1(1-v_1^2)^{\frac{1}{2}}.$$

Also, with

$$D^n(\mathbf{v}_1\hat{\mathbf{n}}\cdot\mathbf{r}_1 - \hat{\mathbf{n}}t) = D_1^n(\mathbf{v}_1\hat{\mathbf{n}}\cdot\mathbf{r}_1) - \hat{\mathbf{n}}t\delta_{n0} - \hat{\mathbf{n}}\delta_{n1},$$

the sum rule

$$\sum_{n=0}^{\infty}\sum_{p=n}^{\infty} = \sum_{p=0}^{\infty}\sum_{n=0}^{p},$$

and shifting the index from $n-1$ to n in the last term, (B5) becomes

$$(D\Omega_{\mathrm{LT}})_1 = -D[m_1\hat{\mathbf{n}}\cdot\mathbf{r}_1(1-v_1^2)^{\frac{1}{2}}] - e^2\sum_{p=0}^{\infty}\sum_{n=0}^{p}\frac{(-D_2)^p}{(2p)!}(2p-1)\binom{p}{n}[D_1^{p-n}(1-\mathbf{v}_1\cdot\mathbf{v}_2)r^{2p-3}\mathbf{r}]\cdot D_1^n(\mathbf{v}_1\hat{\mathbf{n}}\cdot\mathbf{r}_1) + e^2\sum_{p=0}^{\infty}\sum_{n=0}^{p}\frac{(-D_2)^p}{(2p)!}\binom{p}{n}\mathbf{v}_2(D_1^{p-n}r^{2p-1})\cdot D_1^{n+1}(\mathbf{v}_1\hat{\mathbf{n}}\cdot\mathbf{r}_1) + e^2\sum_{p=1}^{\infty}\frac{(-D_2)^p D_1^{p-1}}{(2p)!}p(2p-1)(1-\mathbf{v}_1\cdot\mathbf{v}_2)r^{2p-3}\mathbf{r}\cdot\hat{\mathbf{n}} - e^2\sum_{p=0}^{\infty}\frac{(-D_1D_2)^p}{(2p)!}\mathbf{v}_2\cdot\hat{\mathbf{n}}r^{2p-1}. \tag{B6}$$

The term containing t explicitly has been dropped; being antisymmetric in the charge indices, it is cancelled by its opposite number upon addition of $(D\Omega_{\mathrm{LT}})_2$. The remaining pieces may now be simplied using Leibniz' rule

$$D_1^p(\mathbf{A}\cdot\mathbf{B}) = \sum_{n=0}^{p}\binom{p}{n}(D_1^{p-n}\mathbf{A})\cdot D_1^n\mathbf{B}.$$

Further, if the index p is shifted in the fourth term so

that counting begins at zero, the interaction pieces of (B6) become

$$-e^2\sum_{p=0}^{\infty}\frac{(-D_1D_2)^p}{(2p)!}\{(2p-1)(1-\mathbf{v}_1\cdot\mathbf{v}_2)\mathbf{r}\cdot\mathbf{v}_1\hat{\mathbf{n}}\cdot\mathbf{r}_1r^{2p-3} - r^{2p-1}D_1(\mathbf{v}_1\cdot\mathbf{v}_2\hat{\mathbf{n}}\cdot\mathbf{r}_1) + \tfrac{1}{2}D_2[(1-\mathbf{v}_1\cdot\mathbf{v}_2)\mathbf{r}\cdot\hat{\mathbf{n}}r^{2p-1}] + \mathbf{v}_2\cdot\hat{\mathbf{n}}r^{2p-1}\}, \quad \text{(B7)}$$

the curly brace of which is equal to

$$D_1[(1-\mathbf{v}_1\cdot\mathbf{v}_2)\hat{\mathbf{n}}\cdot\mathbf{r}_1r^{2p-1}] - \hat{\mathbf{n}}\cdot\mathbf{v}r^{2p-1} + \tfrac{1}{2}D_2[(1-\mathbf{v}_1\cdot\mathbf{v}_2)\hat{\mathbf{n}}\cdot\mathbf{r}r^{2p-1}]. \quad \text{(B8)}$$

The piece involving the relative speed is antisymmetric and can be discarded.

The interaction part of $(D\Omega_{LT})_2$ will be just (B7) with an index-exchanged curly brace. Then the interaction terms of $D\Omega_{LT}$ will be (B7) with a braced term the sum of (B8) and its exchanged counterpart, easily seen to be

$$\tfrac{1}{2}(D_1+D_2)[(1-\mathbf{v}_1\cdot\mathbf{v}_2)r^{2p-1}\hat{\mathbf{n}}\cdot(\mathbf{r}_1+\mathbf{r}_2)].$$

The right-hand side of (14) is thus an exact derivative for Lorentz transformations. The expression may be integrated to give

$$\Omega_{LT} = -m_1\hat{\mathbf{n}}\cdot\mathbf{r}_1(1-v_1^2)^{\frac{1}{2}} - m_2\hat{\mathbf{n}}\cdot\mathbf{r}_2(1-v_2^2)^{\frac{1}{2}} - \frac{e^2}{2}\sum_{p=0}^{\infty}\frac{(-D_1D_2)^p}{(2p)!}(1-\mathbf{v}_1\cdot\mathbf{v}_2)r^{2p-1}\hat{\mathbf{n}}\cdot(\mathbf{r}_1+\mathbf{r}_2). \quad \text{(B9)}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Motion of a Charged Particle in a Spatially Periodic Magnetic Field*

TIMOTHY P. COFFEY†
Department of Physics, University of Michigan

(Received 14 August 1967)

Degenerate perturbation theory is employed to discuss the motion of a charged particle in a constant magnetic field on which is superimposed a weak, transverse, spatially periodic magnetic field. A first-order solution of the equations of motion is presented. It is shown that the secular motion is periodic in time. The significance of this result with respect to the stability of protons in the inner Van Allen belt is discussed.

1. INTRODUCTION

In a previous paper[1] (henceforth cited as I) we presented a new formulation of classical perturbation theory. There we illustrated the nondegenerate form of this theory by discussing the van der Pol equation. The van der Pol equation has, of course, been adequately discussed by many authors using a variety of techniques. In the present paper, however, we discuss a problem which has not been adequately treated in previous publications. Here we employ degenerate perturbation theory to discuss the interaction between a charged particle and a constant magnetic field on which is superimposed a weak, transverse, spatially periodic magnetic field.

This interaction has played an important role in recent discussions of the stability of protons in the inner Van Allen belt. For example, Dragt[2] and Wentzel[3] have argued that a resonant interaction between the charged particle and a periodic magnetic field would cause a breakdown of the adiabatic invariance of the particle's orbital magnetic moment. They further argue that such a breakdown of the adiabatic invariance of the magnetic moment would destroy the magnetic trapping effect. This reasoning has led them to assert that a periodic disturbance (produced, for example, by a hydromagnetic wave) on the geomagnetic field is responsible for the removal of protons, which would otherwise be trapped, from the inner Van Allen belt.

In what follows we obtain a complete first-order solution of the equations of motion in the case where the periodic field is a sinusoid. We find that the secular changes produced by such a field are of bounded variation. In particular, the "average" magnetic moment is a periodic function of time. The relative fluctuation in the "average" magnetic moment depends upon the ratio of the particle's cyclotron radius to the wavelength of the periodic disturbance: the fluctuation is large when the ratio is small and small when the ratio is large. The essential point is that

* Supported in part by the National Science Foundation.
† Present address: EGG, Inc., Arlington, Virginia.
[1] T. P. Coffey and G. W. Ford, J. Math. Phys. **10**, 999 (1969).
[2] A. J. Dragt, J. Geophys. Res. **66**, 1641 (1961).
[3] D. Wentzel, J. Geophys. Res. **66**, 359 (1961).

that counting begins at zero, the interaction pieces of (B6) become

$$-e^2\sum_{p=0}^{\infty}\frac{(-D_1D_2)^p}{(2p)!}\{(2p-1)(1-\mathbf{v}_1\cdot\mathbf{v}_2)\mathbf{r}\cdot\mathbf{v}_1\hat{\mathbf{n}}\cdot\mathbf{r}_1r^{2p-3} - r^{2p-1}D_1(\mathbf{v}_1\cdot\mathbf{v}_2\hat{\mathbf{n}}\cdot\mathbf{r}_1) + \tfrac{1}{2}D_2[(1-\mathbf{v}_1\cdot\mathbf{v}_2)\mathbf{r}\cdot\hat{\mathbf{n}}r^{2p-1}] + \mathbf{v}_2\cdot\hat{\mathbf{n}}r^{2p-1}\}, \quad \text{(B7)}$$

the curly brace of which is equal to

$$D_1[(1-\mathbf{v}_1\cdot\mathbf{v}_2)\hat{\mathbf{n}}\cdot\mathbf{r}_1r^{2p-1}] - \hat{\mathbf{n}}\cdot\mathbf{v}r^{2p-1} + \tfrac{1}{2}D_2[(1-\mathbf{v}_1\cdot\mathbf{v}_2)\hat{\mathbf{n}}\cdot\mathbf{r}r^{2p-1}]. \quad \text{(B8)}$$

The piece involving the relative speed is antisymmetric and can be discarded.

The interaction part of $(D\Omega_{LT})_2$ will be just (B7) with an index-exchanged curly brace. Then the interaction terms of $D\Omega_{LT}$ will be (B7) with a braced term the sum of (B8) and its exchanged counterpart, easily seen to be

$$\tfrac{1}{2}(D_1+D_2)[(1-\mathbf{v}_1\cdot\mathbf{v}_2)r^{2p-1}\hat{\mathbf{n}}\cdot(\mathbf{r}_1+\mathbf{r}_2)].$$

The right-hand side of (14) is thus an exact derivative for Lorentz transformations. The expression may be integrated to give

$$\Omega_{LT} = -m_1\hat{\mathbf{n}}\cdot\mathbf{r}_1(1-v_1^2)^{\frac{1}{2}} - m_2\hat{\mathbf{n}}\cdot\mathbf{r}_2(1-v_2^2)^{\frac{1}{2}} - \frac{e^2}{2}\sum_{p=0}^{\infty}\frac{(-D_1D_2)^p}{(2p)!}(1-\mathbf{v}_1\cdot\mathbf{v}_2)r^{2p-1}\hat{\mathbf{n}}\cdot(\mathbf{r}_1+\mathbf{r}_2). \quad \text{(B9)}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Motion of a Charged Particle in a Spatially Periodic Magnetic Field*

TIMOTHY P. COFFEY†
Department of Physics, University of Michigan

(Received 14 August 1967)

Degenerate perturbation theory is employed to discuss the motion of a charged particle in a constant magnetic field on which is superimposed a weak, transverse, spatially periodic magnetic field. A first-order solution of the equations of motion is presented. It is shown that the secular motion is periodic in time. The significance of this result with respect to the stability of protons in the inner Van Allen belt is discussed.

1. INTRODUCTION

In a previous paper[1] (henceforth cited as I) we presented a new formulation of classical perturbation theory. There we illustrated the nondegenerate form of this theory by discussing the van der Pol equation. The van der Pol equation has, of course, been adequately discussed by many authors using a variety of techniques. In the present paper, however, we discuss a problem which has not been adequately treated in previous publications. Here we employ degenerate perturbation theory to discuss the interaction between a charged particle and a constant magnetic field on which is superimposed a weak, transverse, spatially periodic magnetic field.

This interaction has played an important role in recent discussions of the stability of protons in the inner Van Allen belt. For example, Dragt[2] and Wentzel[3] have argued that a resonant interaction between the charged particle and a periodic magnetic field would cause a breakdown of the adiabatic invariance of the particle's orbital magnetic moment. They further argue that such a breakdown of the adiabatic invariance of the magnetic moment would destroy the magnetic trapping effect. This reasoning has led them to assert that a periodic disturbance (produced, for example, by a hydromagnetic wave) on the geomagnetic field is responsible for the removal of protons, which would otherwise be trapped, from the inner Van Allen belt.

In what follows we obtain a complete first-order solution of the equations of motion in the case where the periodic field is a sinusoid. We find that the secular changes produced by such a field are of bounded variation. In particular, the "average" magnetic moment is a periodic function of time. The relative fluctuation in the "average" magnetic moment depends upon the ratio of the particle's cyclotron radius to the wavelength of the periodic disturbance: the fluctuation is large when the ratio is small and small when the ratio is large. The essential point is that

* Supported in part by the National Science Foundation.
† Present address: EGG, Inc., Arlington, Virginia.
[1] T. P. Coffey and G. W. Ford, J. Math. Phys. **10**, 999 (1969).
[2] A. J. Dragt, J. Geophys. Res. **66**, 1641 (1961).
[3] D. Wentzel, J. Geophys. Res. **66**, 359 (1961).

a resonant interaction between the particle and the periodic field is not sufficient to cause large changes in the magnetic moment.

Aside from its application to the question of stability of protons in the inner Van Allen belt, the example which we discuss is an interesting mathematical exercise. It illustrates quite nicely many of the phenomena which are characteristic of nonlinear oscillatory systems. For example, the ideas of secular growth, stability and instability, and synchronous and nonsynchronous behavior arise in a very natural way. The example also illustrates that a nonlinear resonance is considerably more complicated than a linear resonance.

Our program is as follows: In Sec. 2 we derive Hamilton's equations of motion which describe the interaction between the particle and the field; in Sec. 3 we introduce the appropriate perturbation theory and obtain the differential equations which describe the secular motion; in Sec. 4 we perform a phase-plane analysis in order to characterize the secular motion; in Sec. 5 we obtain an explicit solution of the differential equations which describe the secular motion; in Sec. 6 we discuss the behavior of the secular motion under resonance conditions. The final section summarizes the main conclusions of the paper.

2. THE EQUATIONS OF MOTION

In the Cartesian reference frame x, y, z the magnetic field is taken to have the form

$$\mathbf{B} = [B_1 \sin kz, 0, B_0]. \tag{2.1}$$

This field can be described by the vector potential

$$\mathbf{A} = [-B_0 y, (B_1/k) \cos kz, 0]. \tag{2.2}$$

The nonrelativistic Hamiltonian H which describes the system is

$$\begin{aligned} H &= (1/2m)\{\mathbf{p} - (e/c)\mathbf{A}\}^2 \\ &= (1/2m)\{[p_x + m\omega_0 y]^2 \\ &\quad + [p_y - (m\omega_1/k) \cos kz]^2 + p_z^2\}, \end{aligned} \tag{2.3}$$

where

$$\omega_0 = eB_0/mc, \quad \omega_1 = eB_1/mc. \tag{2.4}$$

In order to prepare the system for perturbation theory, we introduce the new canonical momenta J, p_r, P_Z and their conjugate coordinates ψ, r, and Z as follows:

$$\begin{aligned} x &= r - (2J/m\omega_0)^{\frac{1}{2}} \cos \psi, & p_x &= p_r, \\ y &= -(1/m\omega_0)p_r + (2J/m\omega_0)^{\frac{1}{2}} \sin \psi, & & \\ & & p_y &= (2m\omega_0 J)^{\frac{1}{2}} \cos \psi, \\ z &= Z, & p_z &= P_Z. \end{aligned} \tag{2.5}$$

The quantities r, $p_r/m\omega_0$, and Z are the Cartesian coordinates of the guiding center. In the unperturbed state the particle gyrates about this center with angular velocity ω_0 in a circle of radius $(2J/m\omega_0)^{\frac{1}{2}}$. We measure the time in units of the rotation period. To do this we introduce a new independent variable

$$\tau = \omega_0 t. \tag{2.6}$$

It is straightforward to show that the Hamiltonian h which is appropriate to the new variables is

$$\begin{aligned} h = J + (1/2m\omega_0)P_Z^2 - (\epsilon/k)(2m\omega_0 J)^{\frac{1}{2}} \cos \psi \cos kZ \\ + (\epsilon^2 m\omega_0/2k^2) \cos^2 kZ, \end{aligned} \tag{2.7}$$

where

$$\epsilon = \omega_1/\omega_0 = B_1/B_0. \tag{2.8}$$

Hamilton's equations of motion are found from Eq. (2.7) to be

$$\begin{aligned} J' &= -(\epsilon/2k)(2m\omega_0 J)^{\frac{1}{2}} \\ &\quad \times [\sin(\psi + kZ) + \sin(\psi - kZ)], && (2.9a) \\ P_Z' &= -(\epsilon/2)(2m\omega_0 J)^{\frac{1}{2}} \\ &\quad \times [\sin(\psi + kZ) - \sin(\psi - kZ)] \\ &\quad + (\epsilon^2 m\omega_0/k) \cos kZ \sin kZ, && (2.9b) \\ kZ' &= kP_Z/m\omega_0, && (2.9c) \\ \psi' &= 1 - (\epsilon/2k)(m\omega_0/2J)^{\frac{1}{2}} \\ &\quad \times [\cos(\psi + kZ) + \cos(\psi - kZ)], && (2.9d) \end{aligned}$$

where, for example,

$$J' = dJ/d\tau. \tag{2.10}$$

The system of differential equations (2.9) is in the standard form to which the perturbation theory of I is applicable. The parameter of smallness is $\epsilon = B_1/B_0$. We see from Eqs. (2.9) that the sum angle $\psi + kZ$ contributes only small amplitude, rapid fluctuations to the motion. However, the difference angle $\psi - kZ$ can give rise to secular motion when $\psi' - kZ' = O(\epsilon)$. The system (2.9) must, therefore, be treated by degenerate perturbation theory. We carry out this treatment in the next section.

3. PERTURBATION THEORY

In this section we perform first-order perturbation theory according to the formalism presented in I. Our object is to separate the rapidly fluctuating motion from the secular motion. In order to do this we introduce new variables U, V, K, and ϕ as follows:

$$\begin{aligned} kZ &= KU + \epsilon D_1(U, V, K, \phi) + O(\epsilon^2), && (3.1a) \\ P_Z &= V + \epsilon E_1(U, V, K, \phi) + O(\epsilon^2), && (3.1b) \\ J &= K + \epsilon F_1(U, V, K, \phi) + O(\epsilon^2), && (3.1c) \\ \psi &= \phi + \epsilon G_1(U, V, K, \phi) + O(\epsilon^2), && (3.1d) \end{aligned}$$

where D_1, E_1, F_1, and G_1 are required to be periodic functions of ϕ and of kU with period 2π. The variables U, V, K, and ϕ are to contain the secular motion and D_1, E_1, F_1, and G_1 are to contain the rapidly fluctuating motion. In order to guarantee that U, V, K, and ϕ represent the secular motion, we require that

$$kU' = (kV/m\omega_0) + \epsilon a_1(U, V, K, \phi) + O(\epsilon^2), \quad (3.2a)$$
$$V' = \epsilon b_1(U, V, K, \phi) + O(\epsilon^2), \quad (3.2b)$$
$$K' = \epsilon A_1(U, V, K, \phi) + O(\epsilon^2), \quad (3.2c)$$
$$\phi' = 1 + \epsilon B_1(U, V, K, \phi) + O(\epsilon^2). \quad (3.2d)$$

The functions a_1, b_1, A_1, and B_1 are to contain only those combinations of U and ϕ which can give rise to secular motion. The precise manner in which this choice is made is fully described in I.

If we substitute the ansatz (3.1) and (3.2) into Eqs. (2.9) and retain only terms through first order in ϵ, then we obtain the following set of equations:

$$A_1 + \hat{O}F_1 = -(1/2k)(2m\omega_0K)^{\frac{1}{2}}[\sin(\phi + kU) + \sin(\phi - kU)], \quad (3.3a)$$
$$b_1 + \hat{O}E_1 = -(\tfrac{1}{2})(2m\omega_0K)^{\frac{1}{2}}[\sin(\phi + kU) - \sin(\phi - kU)], \quad (3.3b)$$
$$a_1 + \hat{O}D_1 = (k/m\omega_0)E_1, \quad (3.3c)$$
$$B_1 + \hat{O}G_1 = -(1/2k)(m\omega_0/2K)^{\frac{1}{2}}[\cos(\phi + kU) + \cos(\phi - kU)], \quad (3.3d)$$

where the operator

$$\hat{O} = (kV/m\omega_0)\partial/\partial kU + \partial/\partial\phi. \quad (3.4)$$

The difference angle $\theta = \phi - kU$ can give rise to secular behavior when it is slowly varying. In order that U, V, K, and ϕ contain all of the secular motion, we must absorb the θ dependence into the functions A_1, a_1, b_1, and B_1. We therefore choose

$$A_1 = -(1/2k)(2m\omega_0K)^{\frac{1}{2}}\sin\theta, \quad (3.5a)$$
$$B_1 = -(1/2k)(m\omega_0/2K)^{\frac{1}{2}}\cos\theta, \quad (3.5b)$$
$$a_1 = 0, \quad (3.5c)$$
$$b_1 = (\tfrac{1}{2})(2m\omega_0K)^{\frac{1}{2}}\sin\theta. \quad (3.5d)$$

With this choice of A_1, B_1, a_1, and b_1 we find that

$$D_1 = (k/2\omega_2^2)(2K/m\omega_0)^{\frac{1}{2}}\sin(\phi + kU), \quad (3.6a)$$
$$E_1 = (1/2\omega_2^2)(2m\omega_0K)^{\frac{1}{2}}\cos(\phi + kU), \quad (3.6b)$$
$$F_1 = (1/2k\omega_2)(2m\omega_0K)^{\frac{1}{2}}\cos(\phi + kU), \quad (3.6c)$$
$$G_1 = -(1/2k\omega_2)(m\omega_0/2K)^{\frac{1}{2}}\sin(\phi + kU), \quad (3.6d)$$

where

$$\omega_2 = 1 + (kV/m\omega_0). \quad (3.7)$$

When our choice for A_1, B_1, a_1, and b_1 is substituted into Eqs. (3.2), we find that

$$K' = -(\epsilon/2k)(2m\omega_0K)^{\frac{1}{2}}\sin\theta, \quad (3.8a)$$
$$V' = (\epsilon/2)(2m\omega_0K)^{\frac{1}{2}}\sin\theta, \quad (3.8b)$$
$$\theta' = 1 - (kV/m\omega_0) - (\epsilon/2k)(m\omega_0/2K)^{\frac{1}{2}}\cos\theta, \quad (3.8c)$$

where $\theta' = \phi' - kU'$. It follows from Eqs. (3.8a) and (3.8b) that

$$K + (V/k) = I, \quad (3.9)$$

where I is a constant. The system of Eqs. (3.8), therefore, reduces to two equations relating the two variables K and θ. In the next section we use these equations to obtain some general information concerning K and θ without producing an explicit solution.

4. PHASE-PLANE ANALYSIS

We begin this section by introducing the new variables a and b which are defined by the equations

$$a = R\cos\theta, \quad b = R\sin\theta, \quad (4.1)$$

where

$$R = k(2K/m\omega_0)^{\frac{1}{2}}. \quad (4.2)$$

The quantity R measures the ratio of the cyclotron radius to the fundamental period of the disturbance. The equations of motion for a and b are found from Eqs. (3.8) to be

$$a' = -(1 - C + R^2/2)b, \quad (4.3a)$$
$$b' = (1 - C + R^2/2)a - \epsilon/2, \quad (4.3b)$$

where

$$C = k^2I/m\omega_0 \quad (4.3c)$$

is a constant. These equations give rise to the differential form

$$[1 - C + (R^2/2)]b\,db + \{[1 - C + (R^2/2)]a - (\epsilon/2)\}\,da = 0. \quad (4.4)$$

This is an exact differential whose integral is

$$R^4 + 4(1 - C)R^2 - 4\epsilon a = M, \quad (4.5)$$

where M is a constant. Equation (4.5) expresses the conservation of energy through first order. It follows from the Hamiltonian (2.7) that, to first order in ϵ,

$$M = (8k^2h/m\omega_0) - 4C^2. \quad (4.6)$$

The important aspects of the motion can be illustrated by plotting Eq. (4.5) in the a–b plane. Before doing this, it is useful to examine the points where a' and b' are simultaneously zero. These are the points of equilibrium and are usually termed singular points. It follows from Eqs. (4.3) that the singular points are

to be found from the equations

$$b = 0, \tag{4.7a}$$

$$a^3 + 2(1 - C)a - \epsilon = 0. \tag{4.7b}$$

If ϵ is sufficiently small, the cubic equation will have three real roots; we assume this to be the case. These roots, which we call a_1, a_2, a_3, are approximately as follows:

$$a_1 = 2[-2(1 - C)/3]^{\frac{1}{2}}[(\tfrac{3}{4})^{\frac{1}{2}} - \delta/6] + O(\epsilon^2) > 0, \tag{4.8a}$$

$$a_2 = 2[-2(1 - C)/3]^{\frac{1}{2}}[-(\tfrac{3}{4})^{\frac{1}{2}} - \delta/6] + O(\epsilon^2) < 0, \tag{4.8b}$$

$$a_3 = 2[-2(1 - C)/3]^{\frac{1}{2}}(\delta/3) + O(\epsilon^2) < 0, \tag{4.8c}$$

where

$$\delta = -(\epsilon/2)[-2(1 - C)/3]^{-\frac{3}{2}}. \tag{4.9}$$

The nature of these singular points can be determined by examining the behavior of the motion in their vicinity. In order to do this we let

$$a = a_i + \xi, \quad b = \eta, \tag{4.10}$$

where a_i is one of the singular points and ξ and η represent small displacements from this singular point. Upon substituting Eqs. (4.10) into Eqs. (4.3) and retaining terms through first order in ξ and η, we find that

$$\xi' = -[1 - C + (a_i^2/2)]\eta, \tag{4.11a}$$

$$\eta' = [1 - C + (3a_i^2/2)]\xi. \tag{4.11b}$$

We seek solutions in the form

$$\xi = \xi_0 e^{\lambda\tau}, \quad \eta = \eta_0 e^{\lambda\tau}. \tag{4.12}$$

These solutions are valid if

$$\lambda = \pm\{-[1 - C + (a_i^2/2)][1 - C + (3a_i^2/2)]\}^{\frac{1}{2}}. \tag{4.13}$$

When the values of a_i as expressed by Eqs. (4.8) are substituted into Eqs. (4.13), we find that we can classify the singular points as to their stability. This classification is given in Table I.

Table I shows that the trajectories must close about the point a_3 and they must also close about the point a_1. Furthermore, it follows from Eq. (4.5) that, for large values of R, the trajectories have the form

$$R^4 = \text{const} \tag{4.14}$$

and are, therefore, circles centered about the origin. The manner in which these requirements are satisfied is shown in Fig. 1. The several trajectories are plotted for a particular physical situation. The parameters such as the energy, background field strength, etc., have been assigned values which are appropriate to a proton which is moving in the inner Van Allen belt at a distance of two earth radii. However, the value of ϵ which was used in Fig. 1 was chosen to be about ten times larger than what one would expect at two earth radii.[4] This larger value of ϵ was used to facilitate the plotting of the trajectories.

TABLE I. Classification of singular points.

Singular Points	λ	Nature of Singular Point
a_1	imaginary	Center (stable)
a_2	real	Saddle point (unstable)
a_3	imaginary	Center (stable)

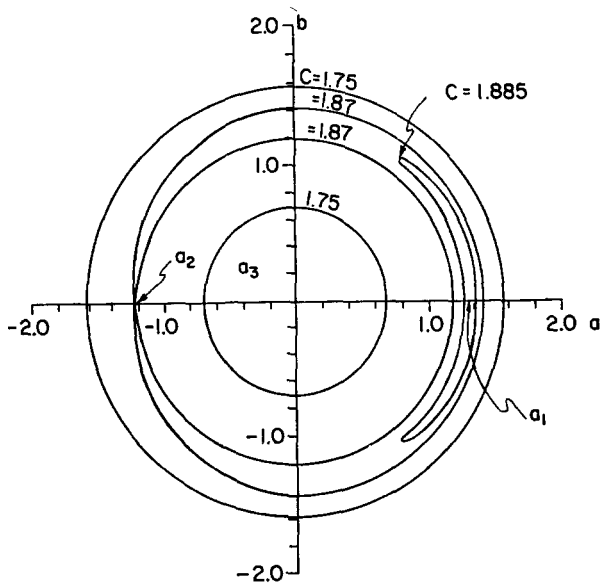

FIG. 1. Trajectories of a 50 MeV proton in the a–b plane; $k = 6.28 \times 10^{-8}$ cm^{-1}, $\omega_0 = 370$ rad/sec, $\epsilon = 0.01$.

The trajectories in Fig. 1 consist of a family of closed curves. This means that K, and consequently V, is a periodic function of time; we obtain the period in the next section. The trajectories can be divided into two groups: those which are centered about the point a_3 and those which are centered about the point a_1. The motion corresponding to the first group is nonsynchronous, since the difference angle θ increases without bound. The motion corresponding to the second group is synchronous since, for these trajectories, the angle θ oscillates between well-defined limits. The synchronous and nonsynchronous regions are separated by the trajectory, called a separatrix, which passes through the unstable point a_2.

The largest fluctuations in K occur on trajectories which pass close to the separatrix. As one moves

[4] See Ref. 2.

away from the separatrix into the nonsynchronous regions, the trajectories rapidly become circles centered about the point a_3. As one moves away from the separatrix into the synchronous region, the fluctuations in K and θ become smaller until, at the point a_1, they vanish. The separatrix, therefore, determines the range of values of K and θ for which maximum resonance occurs. In the following sections we obtain the time dependence of K and we estimate the total fluctuation in K under resonance conditions.

5. TIME DEPENDENCE OF THE MOTION

In order to find the explicit time dependence of K and an expression for the period T, we introduce a new variable

$$S = R^2 = (2k^2/m\omega_0)K. \tag{5.1}$$

It follows from Eqs. (3.8) that

$$S' = -\epsilon R \sin\theta. \tag{5.2}$$

The right-hand side of Eq. (5.2) can be expressed as a function of S alone by making use of Eqs. (4.5) and (4.6). A straightforward calculation gives

$$S' = -(\tfrac{1}{4})[-S^4 - 8LS^3 - (8N + 16L^2)S^2 - 16(2LN - \epsilon^2)S - 16N^2]^{\frac{1}{2}}, \tag{5.3}$$

where

$$L = 1 - C, \quad N = C^2 - 2Ch/I. \tag{5.4}$$

If we denote the roots of the quartic in the square bracket in (5.3) by S_1, S_2, S_3, and S_4, then Eq. (5.3) becomes

$$S' = -(\tfrac{1}{4})[(S_1 - S)(S - S_2)(S - S_3)(S - S_4)]^{\frac{1}{2}}. \tag{5.5}$$

This is a first-order differential equation for $S(\tau)$ which is solvable in terms of elliptic functions. The solutions depend upon the nature of the roots S_1, S_2, S_3, and S_4; we must distinguish the case of two real roots from the case of four real roots.

Case of Two Real Roots

It should be clear from Fig. 1 that two real roots corresponds to motion in the synchronous region. We arrange the real roots S_1 and S_2 such that $S_1 \geq S_2$. The complex roots S_3 and S_4 may be written as

$$S_3 = m + in, \quad S_4 = m - in. \tag{5.6}$$

In this case Eq. (5.5) has the solution[5]

$$S(\tau) = \frac{S_1B + S_2A + (S_2A - S_1B)\,\mathrm{cn}\,[(\tau - \tau_0)/4g]}{A + B + (A - B)\,\mathrm{cn}\,[(\tau - \tau_0)/4g]}, \tag{5.7}$$

where

$$A^2 = (S_1 - m)^2 + n^2, \quad B^2 = (S_2 - m)^2 + n^2, \tag{5.8}$$

and

$$g = (AB)^{-\frac{1}{2}}. \tag{5.9}$$

The function $\mathrm{cn}\,(x)$ is a Jacobi elliptic function. The modulus κ of the elliptic function is given by

$$\kappa^2 = \frac{(S_1 - S_2)^2 - (A - B)^2}{4AB}. \tag{5.10}$$

The constant τ_0 is chosen to satisfy the initial conditions. The function $\mathrm{cn}\,(x)$ is periodic in x with period 4K, K being the complete elliptic integral of the first kind. It follows that $S(\tau)$, and hence $K(\tau)$, is a periodic function of τ with period T given by

$$T = 16\mathrm{K}(AB)^{-\frac{1}{2}}, \tag{5.11}$$

where the modulus κ of the complete elliptic integral is given by Eq. (5.10).

Case of Four Real Roots

Four real roots correspond to motion in the nonsynchronous region. We consider first the case where $S_1 \geq S \geq S_2 \geq S_3 \geq S_4$. In this case the solution of Eq. (5.5) is[6]

$$S(\tau) = \frac{S_2(S_1 - S_3) - S_3(S_1 - S_2)\,\mathrm{sn}^2\,[(\tau - \tau_0)/4g]}{S_1 - S_3 - (S_1 - S_2)\,\mathrm{sn}^2\,[(\tau - \tau_0)/4g]}, \tag{5.12}$$

where

$$g = 2[(S_1 - S_3)(S_2 - S_4)]^{-\frac{1}{2}}. \tag{5.13}$$

The modulus κ of the Jacobi elliptic sine function is given by

$$\kappa^2 = \frac{(S_1 - S_2)(S_3 - S_4)}{(S_1 - S_3)(S_2 - S_4)}. \tag{5.14}$$

The function $\mathrm{sn}^2\,(x)$ is periodic in x with period 2K, K being the complete elliptic integral of the first kind. It follows that $S(\tau)$, and hence $K(\tau)$, is a periodic function of τ with period

$$T = 32\mathrm{K}[(S_1 - S_3)(S_2 - S_4)]^{-\frac{1}{2}}. \tag{5.15}$$

In the remaining case where $S_1 \geq S_2 \geq S_3 \geq S \geq S_4$, we find that[7]

$$S(\tau) = \frac{S_4(S_1 - S_3) + S_1(S_3 - S_4)\,\mathrm{sn}^2\,[(\tau - \tau_0)/4g]}{S_1 - S_3 + (S_3 - S_4)\,\mathrm{sn}^2\,[(\tau - \tau_0)/4g]}, \tag{5.16}$$

where g is given by Eq. (5.13) and where the modulus κ of the Jacobi elliptic sine function is given by

[5] P. F. Byrd and M. D. Friedman, *Handbook of Elliptic Integrals for Engineers and Physicists* (Springer-Verlag, Berlin, 1954), Eq. (259.00), p. 133.

[6] Reference 5, Eq. (256.00), p. 120.

[7] Reference 5, Eq. (252.00), p. 103.

Eq. (5.14). It follows that $S(\tau)$ is a periodic function of τ with period given by Eq. (5.15).

We have now determined the function $S(\tau)$ in the synchronous and nonsynchronous regions. A knowledge of the function $S(\tau)$ immediately determines $K(\tau)$, $V(\tau)$, and $\cos\theta(\tau)$. These latter functions allow us to find the explicit time dependence of $\phi(\tau)$ and $kU(\tau)$. We do not attempt to do this, since the resulting expressions would add little to our understanding of the motion. The significant point is that $K(\tau)$ and $V(\tau)$ are strictly periodic functions of τ. In the next section we estimate the maximum fluctuation in $K(\tau)$.

6. BEHAVIOR UNDER RESONANCE CONDITIONS

Exact resonance occurs when the particle traverses a single period of the sinusoidal field in one cyclotron period. This exact resonance is nearly fulfilled when

$$kV/m\omega_0 = 1. \tag{6.1}$$

If we substitute condition (6.1) into Eq. (2.7) and neglect the first-order term, we find that

$$k^2K/m\omega_0 = (k^2h/m\omega_0) - (\tfrac{1}{2}). \tag{6.2}$$

Equations (4.5), (4.6), (6.1), and (6.2) allow us to find a value of the constant C which corresponds to near resonance. The appropriate value of C is found to be

$$C = d + \tfrac{1}{2}, \tag{6.3}$$

where $d = k^2h/m\omega_0$. With this value of C the quartic in Eq. (5.3) becomes

$$\begin{aligned} S^4 + (4 - 8d)S^3 + (6 - 24d + 24d^2)S^2 \\ + (4 - 24d + 48d^2 - 32d^3 - 16\epsilon^2)S \\ + 1 - 8d + 24d^2 - 32d^3 + 16d^4. \end{aligned} \tag{6.4}$$

The roots of this quartic are approximately as follows:

$$S_{1,2} = (2d - 1) \pm 2(2d - 1)^{\frac{1}{4}}\epsilon^{\frac{1}{2}} + O(\epsilon), \tag{6.5}$$

$$S_{3,4} = (2d - 1) \pm i2(2d - 1)^{\frac{1}{4}}\epsilon^{\frac{1}{2}} + O(\epsilon). \tag{6.6}$$

Since two roots are complex, it follows that the trajectory corresponding to condition (6.3) lies in the synchronous region. Now, by definition,

$$d = k^2h/m\omega_0 = k^2H/m\omega_0^2, \tag{6.7}$$

where H is the energy. Thus

$$\begin{aligned} 2d - 1 &= (2k^2/m\omega_0^2)[H - (m\omega_0^2/2k^2)] \\ &= (2k^2/m\omega_0^2)[(mv_{\perp_0}^2/2) + (p_{z_0}^2/2m) \\ &\qquad - (m\omega_0^2/2k^2) + O(\epsilon)], \end{aligned} \tag{6.8}$$

where $v_{\perp_0}$ is the initial transverse velocity and p_{z_0} is the initial longitudinal momentum. According to Eq. (6.1),

$$p_{z_0} = m\omega_0/k + O(\epsilon). \tag{6.9}$$

Upon substituting Eq. (6.9) into Eq. (6.8), we find that

$$2d - 1 = k^2v_{\perp_0}^2/\omega_0^2 + O(\epsilon). \tag{6.10}$$

We now define the relative fluctuation ΔK in K as follows:

$$\Delta K = \frac{2(K_{\max} - K_{\min})}{(K_{\max} + K_{\min})} = \frac{2(S_1 - S_2)}{(S_1 + S_2)}. \tag{6.11}$$

Upon making use of Eqs. (6.5), (6.10), and (6.11), we find that

$$\Delta K = 4(kv_{\perp_0}/\omega_0)^{-\frac{3}{2}}\epsilon^{\frac{1}{2}} + O(\epsilon). \tag{6.12}$$

If we denote the initial value of K by K_0, then $v_{\perp_0}$ and K_0 are related by the expression

$$mv_{\perp_0}^2/2 = K_0\omega_0 + O(\epsilon). \tag{6.13}$$

It follows that

$$\begin{aligned} kv_{\perp_0}/\omega_0 &= k(2K_0/m\omega_0)^{\frac{1}{2}} + O(\epsilon) \\ &= R_0 + O(\epsilon), \end{aligned} \tag{6.14}$$

where R_0 is 2π times the ratio of the initial cyclotron radius to the wavelength of the disturbance. Upon substituting Eq. (6.14) into Eq. (6.12), we find that

$$\Delta K = 4R_0^{-\frac{3}{2}}\epsilon^{\frac{1}{2}} + O(\epsilon). \tag{6.15}$$

Thus the relative fluctuation in K under resonance conditions depends not only on ϵ but also on the ratio of the cyclotron radius to the wavelength of the periodic disturbance. This dependence of the relative change in K upon R_0 is not surprising. An increase of the wavelength of the disturbance requires a corresponding increase in the longitudinal particle velocity in order to achieve resonance. Associated with the increase in the longitudinal velocity is a decrease in the transverse velocity and hence a decrease in absolute value of K. This decrease in absolute value contributes to the increase in the relative change.

The ideas developed above are best illustrated through an example. We first observe from Eq. (6.15) that relative fluctuations of order unity will occur when

$$R_0 \approx (16\epsilon)^{\frac{1}{3}}. \tag{6.16}$$

Now consider the trajectories plotted in Fig. 1. These trajectories correspond to a 50 MeV proton moving in a disturbance whose wavelength is 10^8 cm (1000 km) and to a field strength ratio $\epsilon = 0.01$. It follows from Eq. (6.16) that, for $\epsilon = 0.01$, relative fluctuations of order unity will occur when

$$R_0 = (0.16)^{\frac{1}{3}} = 0.543 \tag{6.17}$$

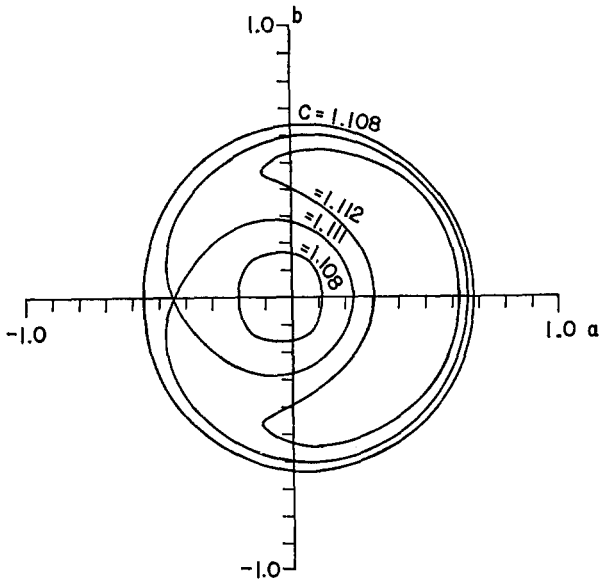

FIG. 2. Trajectories of a 50 MeV proton in the a–b plane; $k = 4.19 \times 10^{-8}$ cm^{-1}, $\omega_0 = 370$ rad/sec, $\epsilon = 0.01$.

lies in the resonance region. However, it is clear from Fig. 1 that $R_0 = 0.543$ lies well outside the resonance region. Thus $K(\tau)$ for a 50 MeV proton moving in a periodic disturbance whose wavelength is 1000 km undergoes only relatively small fluctuations.

Let us now increase the wavelength of the disturbance to 1.5×10^8 cm (1500 km) and hold the other parameters fixed. The trajectories in the a–b plane for this situation are plotted in Fig. 2. Inspection of Fig. 2 reveals that $R_0 = 0.543$ lies in the resonance region. Thus, as is evident from Fig. 2, $K(\tau)$ for a 50 MeV proton moving in a periodic disturbance whose wavelength is 1500 km can undergo relative fluctuations of order unity. Here we have graphic evidence of the influence of R_0 on the relative fluctuation of $K(\tau)$. We conclude this section with a few remarks relating our results to previously published work.

The average orbital magnetic moment $\mu(\tau)$ is related to $K(\tau)$ by the equation

$$\mu(\tau) = eK(\tau)/mc. \tag{6.18}$$

Therefore, what has been said above about $K(\tau)$ also holds for $\mu(\tau)$. This means that $\mu(\tau)$ is a periodic function of τ. Furthermore, the relative fluctuation in $\mu(\tau)$ (that is, the fluctuation measured with respect to the mean value of μ) depends upon the ratio of the cyclotron radius to the wavelength of the periodic disturbance. Neither of these results was employed by Dragt[2] or Wentzel[3] in their discussions of magnetospheric scattering.

The problem of magnetospheric scattering is certainly more complicated than the problem which we have discussed here. In magnetospheric scattering the effects of the mirror fields must be included as well as the fact that a particle encounters many waves of varying intensity, wavelength, and initial phase angle. These variations are typically accounted for by assuming that the particle's magnetic moment performs a random walk between mirror points.[2,3] The step size in the random walk calculation is found by estimating the change in the magnetic moment during one cyclotron period and multiplying this result by the number of cyclotron periods per transit between mirror points. The estimate is made by deriving an expression analogous to Eq. (3.8a) and integrating this expression by holding the phase angle fixed. Such a procedure neglects the subtleties such as synchronous and nonsynchronous behavior and the dependence of the motion on the ratio of the cyclotron radius to the period of the disturbance. These effects become increasingly more important the longer a particle remains attached to a given wave. It must be borne in mind here that the period of the secular motion varies as $\epsilon^{\frac{1}{2}}$ and not as ϵ. Thus the time interval over which one can employ linear perturbation theory is greatly curtailed under resonant conditions. We see then that, although the previous analyses may give an order-of-magnitude estimate of the change in the magnetic moment, they neglect some interesting and possibly important effects.

7. CONCLUSION

We have shown that the perturbation theory presented in I yields a complete first-order solution for the motion of a charged particle in a constant magnetic field on which is superimposed a weak, spatially periodic magnetic field. The significant result from a physical viewpoint is the periodic behavior of the secular motion. It is hoped that we have succeeded in showing that this periodic behavior points to some possible deficiencies in previous work concerning the stability of protons in the inner Van Allen belt.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Perturbation Method for a Nonlinear Wave Modulation. I

TOSIYA TANIUTI
Department of Physics, Nagoya University, Nagoya, Japan
AND
NOBUO YAJIMA
Research Institute for Fundamental Physics, Kyoto University, Kyoto, Japan

(Received 20 November 1968)

In this paper we consider a system of nonlinear wave equations which admits, in a linear approximation, a planewave solution with high-frequency oscillation. Then, for the wave of small but finite amplitude, we investigate how slowly varying parts of the wave such as the amplitude are modulated by nonlinear self-interactions. A stretching transformation shows that, in the lowest order of an asymptotic expansion, the original system of equations can be reduced to a tractable, single, nonlinear equation to determine the amplitude modulation.

1. INTRODUCTION

It was first shown by Gardner and Morikawa[1] that, by means of a coordinate stretching, the system of equations for a hydromagnetic wave in a cold plasma can be reduced to the Kortweg–deVries equation. The extension of the theory to general nonlinear systems, dissipative[2] as well as dispersive, has been established by Gardner and Su[3] and Taniuti and Wei.[4] However, their theories are restricted to the propagation of waves of small wavenumber and low frequency. For example, if they are applied to a nondissipative system, it is required that the dispersion relation for linearized equations takes a form

$$\omega = \lambda k + O(k^3) + \cdots,$$

and waves for small k are to be considered, where ω is the frequency, k the wavenumber, and λ is constant. This restriction excludes direct applications to systems involving oscillations such as are characterized by the dispersion relation

$$\omega = \omega_0 + \lambda k + O(k^2).$$

In order to establish a theory applicable to like systems, in the present paper we consider a system of nonlinear partial differential equations, which admits, in a linear approximation, a plane wave with high-frequency oscillation.

If the amplitude of the wave is small but finite, nonlinear terms give rise to a modulation of the amplitude as well as waves of higher harmonics. When the amplitude varies slowly over the period of the oscillation, a stretching transformation allows us to separate the system into a rapidly varying part associated with the oscillation and a slowly varying one such as the amplitude. Then a formal solution is given in an asymptotic expansion, and we derive, in the lowest order of the expansion, an equation to determine the modulation of the amplitude, which in certain cases becomes the nonlinear Schrödinger equation.

2. METHOD OF SOLUTION

In this paper, we consider a system of equations,

$$\frac{\partial U}{\partial t} + A(U)\frac{\partial U}{\partial x} + B(U) = 0, \tag{1}$$

where U is a column vector with n components, $u_1, u_2, \cdots, u_n$, and the $n \times n$ matrix A and the column vector B are functions of u_i's, being assumed sufficiently smooth. Let $U^{(0)}$ be a constant solution, satisfying

$$B(U^{(0)}) = 0, \tag{2}$$

and define the matrices A_0 and ∇B_0 by

$$A_0 \equiv A(U^{(0)}), \tag{3a}$$

$$(\nabla B_0)_{ij} = \left(\frac{\partial B_i}{\partial u_j}\right)_{U=U^{(0)}}. \tag{3b}$$

Then Eq. (1) linearized about $U^{(0)}$ takes the form

$$\frac{\partial U}{\partial t} + A_0\frac{\partial U}{\partial x} + \nabla B_0 \cdot U = 0, \tag{1'}$$

which admits a plane wave

$$\sim \exp\{\pm i(kx - \omega t)\}$$

subject to the dispersion relation

$$\det W_{\pm 1} \equiv \det|\mp i\omega I \pm ikA_0 + \nabla B_0| = 0, \tag{4}$$

where I is the unit matrix. Our object is to investigate how the plane wave is modulated by nonlinear

[1] C. S. Gardner and G. K. Morikawa, Courant Institute of Mathematical Sciences, New York University, Report No. NYO 9082, 1960.
[2] A successful application of a stretching transformation to a dissipative system was done earlier. See G. K. Morikawa, Comm. Pure Appl. Math. **10**, 291 (1957).
[3] C. S. Gardner and C. H. Su, 1966 Annual Report, Princeton Univ. Plasma Phys. Lab., Matt-Q-24, May 1967.
[4] T. Taniuti and C. C. Wei, J. Phys. Soc. (Japan) **24**, 941 (1968).

effects. For this aim we may restrict ourselves to the following case: For any real k, Eqs. (4) admit simultaneously at least a single real root ω, which changes smoothly as k changes so that $\omega(k)$ is continuously differentiable with respect to k. In addition, $l\omega$'s ($l = 0, \pm 2, \pm 3, \cdots$) are not roots of Eqs. (4) when ω is the root of Eqs. (4), that is,

$$\det W_l \neq 0 \quad \text{for} \quad |l| \neq 1, \tag{4'}$$

where W_l is the matrix $-il\omega I + ilkA_0 + \nabla B_0$.

We now assume the following: In a neighborhood of $U^{(0)}$ there exists a solution expanded in terms of a small parameter ϵ and of harmonics $\exp[il(kx - \omega t)]$,

$$U = \sum_{\alpha=0} \epsilon^\alpha U^{(\alpha)},$$

$$U^{(\alpha)} = \sum_{l=-\infty}^{+\infty} U_l^{(\alpha)}(\tau, \xi) \exp[il(kx - \omega t)], \quad \alpha \geq 1;$$

for example,

$$U = U^{(0)} + \sum_{l=-\infty}^{+\infty} \sum_{\alpha=1} \epsilon^\alpha U_l^{(\alpha)}(\tau, \xi) \exp[il(kx - \omega t)]. \tag{5}$$

Here τ and ξ are slow variables introduced through the stretching

$$\tau = \epsilon^2 t, \tag{6a}$$

$$\xi = \epsilon(x - \lambda t), \tag{6b}$$

where λ is the group velocity, that is,

$$\lambda = \partial\omega/\partial k. \tag{6c}$$

We also note that for the coefficients $U_l^{(\alpha)}$ the reality conditions

$$U_l^{(\alpha)} = U_{-l}^{(\alpha)*}$$

hold. (The asterisk denotes the complex conjugation.)

The idea of the expansion (5) is to include the fast local oscillations through the dependence on the harmonics, while dependence on τ and ξ takes care of the slow variation in amplitude.

The expansion and collection of powers of ϵ is straightforward. Substituting the expanded terms

$$\begin{aligned} A &= A_0 + \epsilon \nabla A_0 \cdot U^{(1)} + \epsilon^2 \{\nabla A_0 \cdot U^{(2)} \\ &\quad + \tfrac{1}{2} \nabla\nabla A_0 : U^{(1)} U^{(1)}\} + \cdots, \\ B &= B_0 + \epsilon \nabla B_0 \cdot U^{(1)} \\ &\quad + \epsilon^2 \{\nabla B_0 \cdot U^{(2)} + \tfrac{1}{2} \nabla\nabla B_0 : U^{(1)} U^{(1)}\} \\ &\quad + \epsilon^3 \{\nabla B_0 \cdot U^{(3)} + \nabla\nabla B_0 : U^{(1)} U^{(2)} \\ &\quad + \tfrac{1}{6} \nabla\nabla\nabla B_0 \vdots U^{(1)} U^{(1)} U^{(1)}\} + \cdots \end{aligned}$$

into Eq. (1) and equating the various powers of ϵ of the same harmonics to zero, we get

$$W_l U_l^{(1)} = 0, \tag{7a}$$

$$\begin{aligned} &W_l U_l^{(2)} + (-\lambda I + A_0) \frac{\partial U_l^{(1)}}{\partial \xi} \\ &\quad + \nabla A_0 \left\langle \cdot U^{(1)} \sum_{l'=-\infty}^{+\infty} il' k U_{l'}^{(1)} \exp\{il'(kx - \omega t)\} \right\rangle_l \\ &\quad + \tfrac{1}{2} \nabla\nabla B_0 : \langle U^{(1)} U^{(1)} \rangle_l = 0, \end{aligned} \tag{7b}$$

$$\begin{aligned} &W_l U_l^{(3)} + (-\lambda I + A_0) \frac{\partial U_l^{(2)}}{\partial \xi} + \frac{\partial U_l^{(1)}}{\partial \tau} \\ &\quad + \nabla A_0 \left\langle \cdot U^{(1)} \frac{\partial U^{(1)}}{\partial \xi} \right\rangle_l \\ &\quad + \nabla A_0 \langle \cdot U^{(1)} \sum il' k U_{l'}^{(2)} \exp\{il'(kx - \omega t)\} \rangle_l \\ &\quad + \tfrac{1}{2} \nabla\nabla A_0 \left\langle : U^{(1)} U^{(1)} \sum_{l'=-\infty}^{+\infty} il' k U_{l'}^{(1)} \right. \\ &\qquad \left. \times \exp\{il'(kx - \omega t)\} \right\rangle_l \\ &\quad + \nabla A_0 \left\langle \cdot U^{(2)} \sum_{l'=-\infty}^{+\infty} il' k U_{l'}^{(1)} \exp\{il'(kx - \omega t)\} \right\rangle_l \\ &\quad + \nabla\nabla B_0 : \langle U^{(1)} U^{(2)} \rangle_l \\ &\quad + \tfrac{1}{6} \nabla\nabla\nabla B_0 \vdots \langle U^{(1)} U^{(1)} U^{(1)} \rangle_l = 0, \end{aligned} \tag{7c}$$

in which the angle bracket $\langle\ \rangle_l$ denotes the coefficient of the lth harmonics; that is, for any function in Eqs. (7), say Q,

$$Q = \sum_{l=-\infty}^{+\infty} \langle Q \rangle_l \exp\{il(kx - \omega t)\}.$$

Also we have assumed the notations

$$\nabla A_0 \cdot U^{(1)} \equiv \sum_{i=1}^{n} \left(\frac{\partial A}{\partial u_i} \right)_{U=U_0} u_i^{(1)},$$

$$\nabla\nabla A_0 : U^{(1)} U^{(1)} \equiv \sum_{i,j} \left(\frac{\partial^2 A}{\partial u_i \partial u_j} \right)_{U=U_0} u_i^{(1)} u_j^{(1)},$$

and so on.

In view of the condition (4), Eq. (7a) yields

$$U^{(1)} = \varphi^{(1)} \mathbf{R}, \tag{8a}$$

and for $|l| \neq 1$,

$$U_l^{(1)} = 0, \tag{8b}$$

where $\mathbf{R}$ is a column vector satisfying

$$W_l \mathbf{R} = 0 \tag{9}$$

and $\varphi^{(1)}$ is a scalar function of τ and ξ to be determined later. Since Eqs. (8) imply

$$\left\langle \sum_l \sum_{l'} U_l^{(1)} U_{l'}^{(1)} l^p l'^{p'} \exp\{i(l + l')(kx - \omega t)\} \right\rangle_1 = 0$$

for any p and p' assumed zero or positive integer, then, from Eq. (7b) with $l = 1$, it follows that

$$W_1 U_1^{(2)} + (-\lambda I + A_0)\mathbf{R}\frac{\partial \varphi^{(1)}}{\partial \xi} = 0; \qquad (10)$$

hence Eq. (4) requires a compatibility condition for this equation to be solved. Introducing a row vector $\mathbf{L}$ corresponding to $\mathbf{R}$ through the equation

$$\mathbf{L}W_1 = 0 \qquad (9')$$

and multiplying Eq. (10) by $\mathbf{L}$, we have the explicit form of the condition,

$$\mathbf{L}(-\lambda I + A_0)\mathbf{R} = 0. \qquad (11)$$

It can easily be proved that, by means of the definition (6c), Eq. (11) is satisfied automatically; that is, differentiating Eq. (9) once with respect to k and multiplying the equation thus obtained by $\mathbf{L}$ from the left leads directly to Eq. (11). Solving Eq. (10) algebraically, one has

$$U_1^{(2)} = \varphi^{(2)}\mathbf{R} + Z(-\lambda I + A_0)\mathbf{R}\frac{\partial \varphi^{(1)}}{\partial \xi}. \qquad (10')$$

Here $\varphi^{(2)}$ is a function of ξ and τ to be determined in the higher order and Z is a matrix expressed by

$$Z_{ik} = \left(\frac{\partial D_{ik}}{\partial p}\right)_{p=\omega} \Big/ \left(\frac{\partial D}{\partial p}\right)_{p=\omega},$$

in which $D(p)$ is $(\det W_1)_{\omega=p}$, e.g.,

$$D(p) = \det|-ipI + ikA_0 + \nabla B_0|$$

and $D_{ik}(p)$'s are the cofactors of $D(p)$. The following identities, with Eq. (11), prove that Eq. (10') satisfies Eq. (10):

$$\sum_j W_1(p)_{kj} D(p)_{jq} = \delta_{kq} D(p),$$
$$\sum_j D(p)_{kj} W_1(p)_{jq} = \delta_{kq} D(p).$$

The latter of these implies $D(\omega)_{kj} \propto L_i$, from which differentiating with respect to p the former, at ω, yields

$$\sum_j W_{1,kj} Z_{jq} = \delta_{kq} - i\omega \sum \delta_{kj} D_{jq}(\omega) \Big/ \left(\frac{\partial D}{\partial p}\right)\omega.$$

Since $\det W_l \neq 0$ for $|l| = 1$, Eq. (7b) allows us to determine uniquely $U_l^{(2)}$ for $|l| \neq 1$. Noting that

$$\langle U^{(1)} U^{(1)} \rangle_l = 0, \quad \text{for} \quad |l| \geq 3,$$

we get

$$U_l^{(2)} = 0, \quad \text{for} \quad |l| \geq 3; \qquad (12a)$$

while substituting both Eqs. (8) into Eq. (7b) gives

$$U_0^{(2)} = -W_0^{-1}[ik\{(\nabla A_0 \cdot \mathbf{R}^*)\mathbf{R} - \text{c.c}\} + \tfrac{1}{2}(\nabla\nabla B : \mathbf{R}^*\mathbf{R} + \text{c.c})]\,|\varphi^{(1)}|^2, \qquad (12b)$$

$$U_2^{(2)} = -W_2^{-1}\{ik(\nabla A_0 \cdot \mathbf{R})\mathbf{R} + \tfrac{1}{2}\nabla\nabla B_0 : \mathbf{R}\mathbf{R}\}(\varphi^{(1)})^2, \qquad (12c)$$

$$U_{-2}^{(2)} = -W_{-2}^{-1}\{ik(\nabla A_0 \mathbf{R}^*)\mathbf{R}^* + \tfrac{1}{2}\nabla\nabla B_0 : \mathbf{R}^*\mathbf{R}^*\}(\varphi^{(1)*})^2. \qquad (12d)$$

We are now ready to determine $\varphi^{(1)}$. Multiply Eq. (7c) for $l = 1$ by $\mathbf{L}$ from the left, and introduce Eqs. (8), (10'), and (12) into the equation so obtained. Then, by means of Eq. (9'), the first term disappears, and thereby Eq. (11) eliminates $\partial\varphi^{(2)}/(\partial\xi)$ in the second term. In addition, Eqs. (8) imply $\langle U^{(1)} U_1^{(2)} \rangle_1 = 0$; consequently, the terms comprising $\varphi^{(2)}$ vanish. Hence it reduces to an equation for $\varphi^{(1)}$, such that

$$\alpha\frac{\partial \varphi^{(1)}}{\partial \tau} + \beta\frac{\partial^2 \varphi^{(1)}}{\partial \xi^2} + \gamma\,|\varphi^{(1)}|^2\,\varphi^{(1)} = 0. \qquad (13)$$

Here α, β, and γ are constant, being given as follows:

$$\alpha = \mathbf{L} \cdot \mathbf{R}, \qquad (14a)$$

$$\beta = \mathbf{L}(-\lambda I + A_0)Z(-\lambda I + A_0)\mathbf{R}, \qquad (14b)$$

$$\gamma = \mathbf{L}[ik\{2(\nabla A_0 \cdot \mathbf{R}^*)\mathbf{R}_2^{(2)} - (\nabla A_0 \cdot \mathbf{R}_2^{(2)})\mathbf{R}^* + (\nabla A_0 \cdot \mathbf{R}_0^{(2)})\mathbf{R} + (\nabla\nabla A_0 : \mathbf{R}\mathbf{R}^*)\mathbf{R} - \tfrac{1}{2}(\nabla\nabla A_0 : \mathbf{R}\mathbf{R})\mathbf{R}^*\} + \nabla\nabla B_0 : (\mathbf{R}\mathbf{R}_0^{(2)} + \mathbf{R}^*\mathbf{R}_2^{(2)}) + \tfrac{1}{2}\nabla\nabla\nabla B_0 \vdots \mathbf{R}\mathbf{R}^*\mathbf{R}], \qquad (14c)$$

where $\mathbf{R}_2^{(2)}$ and $\mathbf{R}_0^{(1)}$ are constant column vectors introduced through the equations

$$U_2^{(2)} = \mathbf{R}_2^{(2)}(\varphi^{(1)})^2,$$
$$U_0^{(2)} = \mathbf{R}_0^{(2)}\,|\varphi^{(1)}|^2;$$

for example,

$$\mathbf{R}_2^{(2)} = -W_2^{-1}\{ik(\nabla A_0 \cdot \mathbf{R})\mathbf{R} + \tfrac{1}{2}\nabla\nabla B_0 : \mathbf{R}\mathbf{R}\}, \qquad (15a)$$

$$\mathbf{R}_0^{(2)} = -W_0^{-1}[ik\{(\nabla A_0 \cdot \mathbf{R}^*)\mathbf{R} - \text{c.c}\} + \tfrac{1}{2}(\nabla\nabla B_0 : \mathbf{R}^*\mathbf{R} + \text{c.c})]. \qquad (15b)$$

If α is pure imaginary and β and γ are real, then, in terms of real constants $p = \beta/|\alpha|$ and $q = \gamma/|\alpha|$, Eq. (13) may be written as

$$i\frac{\partial \varphi^{(1)}}{\partial \tau} + p\frac{\partial^2 \varphi^{(1)}}{\partial \xi^2} + q\,|\varphi^{(1)}|^2\,\varphi^{(1)} = 0, \qquad (16)$$

which, for q equal to zero, reduces to the Schrödinger equation. In this sense, Eq. (16) may be called the nonlinear Schrödinger equation. If, in addition, these constants take the same sign, the solution of Eq. (16) which tends to zero for $|\xi| \to \infty$ is a solitary wave,[5-7] so that

$$\varphi^{(1)} = (-2\nu/q)^{\frac{1}{2}}\,\text{sech}\,\{(-p/\nu)^{-\frac{1}{2}}\xi\}\exp(-i\nu\tau); \qquad (17)$$

[5] P. G. Saffman, J. Fluid Mech. **11**, 16 (1961).
[6] R. Y. Chiao, E. Garmire, and C. H. Townes, Phys. Rev. Letters **13**, 479 (1964).
[7] T. Taniuti and H. Washimi, Phys. Rev. Letters **21**, 209 (1968).

hence if $|\varphi^{(1)}|$ approaches a constant φ_0 at infinity, the solution is given by a plane wave

$$\varphi^{(1)} = \varphi_0 \exp\,[i(\mu\xi - E\tau)], \tag{18}$$

where

$$E = p\mu^2 - q\varphi_0^2.$$

We note here that the plane wave is not stable but subject to the modulational instability.[7-9] In order to show this we introduce the real functions ρ and σ through[10,11]

$$\varphi^{(1)} = (\rho)^{\frac{1}{2}} \exp\left(i\int \sigma\, d\xi/2p\right). \tag{19}$$

Substituting Eq. (19) into Eq. (16) gives

$$\frac{\partial\rho}{\partial\tau} + \frac{\partial(\rho\sigma)}{\partial\xi} = 0, \tag{20a}$$

$$\frac{\partial\sigma}{\partial\tau} + \sigma\frac{\partial\sigma}{\partial\xi} = 2pq\frac{\partial\rho}{\partial\xi} + p^2\frac{\partial}{\partial\xi}\left(\rho^{-\frac{1}{2}}\frac{\partial}{\partial\xi}\left(\rho^{-\frac{1}{2}}\frac{\partial\rho}{\partial\xi}\right)\right). \tag{20b}$$

Since $pq > 0$, in the long wavelength limit Eqs. (20) are equivalent to a hydrodynamic system with negative pressure; in other words, the system becomes elliptic. Hence the perturbations modulating a constant amplitude ρ_0 and phase σ_0,

$$\rho_0 + \delta\rho \exp\,\{i(\tilde{k}\xi - \tilde{\omega}\tau)\},$$
$$\sigma_0 + \delta\sigma \exp\,\{i(\tilde{k}\xi - \tilde{\omega}\tau)\},$$

grow for small $\tilde{k}$; that is,

$$\tilde{\omega} = -\sigma_0 k \pm (-2pq\rho_0)^{\frac{1}{2}}\tilde{k} + O(k^3),$$

provided that ρ_0 is finite. A solution for finite-amplitude waves was considered by Kelley[12] to account for the self-focusing in nonlinear optics. Recently the evolution of the instability was investigated by Karpman and Krushkal.[9]

On the other hand, if p and q take the opposite signs, the plane wave is stable. In this case, Eqs. (20) satisfy the conditions given in Ref. 4 to be reducible to the Kortweg-deVries equation; that is, for the expansion in terms of a small parameter μ about the constant state,

$$\rho = \rho_0 + \mu\rho_1 + \mu^2\rho_2 + \cdots,$$
$$\sigma = \sigma_0 + \mu\sigma_1 + \mu^2\sigma_2 + \cdots,$$

the stretching

$$\xi' = \mu^{\frac{1}{2}}(\xi - (\sigma_0 + (-2pq\rho_0)^{\frac{1}{2}})\tau),$$
$$\tau' = \mu\,\tau,$$

transforms Eqs. (20) into the Kortweg-deVries equation for ρ_1 and σ_1.

Example

If A is symmetric and ∇B is antisymmetric, the matrix

$$H \equiv kA_0 - i\nabla B_0$$

becomes Hermitian. Let the eigenvalues of H be ω_m $(m = 1, 2, \cdots, n)$, and denote the corresponding left and right eigenvectors by $\mathbf{L}_m$ and $\mathbf{R}_m$ respectively. [The root of Eq. (4), ω, is identified with ω_1.] Then we may normalize $\mathbf{L}_m$ and $\mathbf{R}_m$, such that

$$(\mathbf{L}_m \cdot \mathbf{R}_m) = ic_m,$$

where c_m is real. On the other hand, the solution of Eq. (10), $U_1^{(2)}$, may be obtained by means of the expansion in terms of the eigenvectors $\{\mathbf{R}_m\}$, which yields

$$U_1^{(2)} = -\omega\sum_{m=1}(\omega - \omega_m)^{-1} \times c_m^{-1}\{\mathbf{L}_m(-\lambda I + A_0)\mathbf{R}_1\}\mathbf{R}_m\frac{\partial\varphi^{(1)}}{\partial\xi}.$$

Hence, noting that the constant β is equal to

$$\mathbf{L}_1(-\lambda I + A_0)U_1^{(2)}\Big/\left(\frac{\partial\varphi^{(1)}}{\partial\xi}\right) = \sum_{m=1}(\omega - \omega_m)^{-1}c_m^{-1}\,|\mathbf{L}_m(-\lambda I + A_0)\mathbf{R}_1|^2,$$

we find that β is real. Therefore, Eq. (13) takes the form

$$i\frac{\partial\varphi}{\partial\tau} + p\frac{\partial^2\varphi}{\partial\xi^2} + (\gamma/c_1)\,|\varphi|^2\,\varphi = 0,$$

where $p\ (= \beta/c_1)$ is real.

CONCLUSION

We finally note a validity of the assumed expansion (5) and a physical relevance of the conditions (4'). Since Eq. (1) may be a system of hyperbolic type which does not admit, in general, unique solutions for all time, it seems likely that the expansion (5) does not converge for a sufficiently long time. The convergence could be connected with a breaking time of the plane wave which is supposed to be larger for smaller k. Hence Eq. (13) may cease to be valid after a finite time, even if it admits solutions for all time.

The conditions (4'), which exclude the self-resonance, may be considered to be generally valid for $l \geq 2$, However, the condition for $l = 0$, that is, $\det(\nabla B_0) = 0$, is not satisfied in most physical systems. In such a case, however, we often find extraneous physical conditions to determine $U_0^{(\alpha)}$, such as boundary conditions and subsidiary conditions, so that the method of solution given in this paper is still applicable. This will be shown in Paper II by examples.

[8] M. J. Lighthill, J. Inst. Math. Appl. **1**, 269 (1965); Proc. Roy. Soc. (London) **A299**, 28 (1967).
[9] V. I. Karpman, Zh. Eksp. Teor. Fiz. Pis. Red. **6**, 829 (1967) [JETP Letters **6**, 277 (1967)]; V. I. Karpman and E. M. Krushkal, Zh. Eksp. Teor. Fiz. **55**, 530 (1968) [Sov. Phys.—JETP **28**, 277 (1969)].
[10] S. A. Akhmanov, A. P. Sukhorukov, and R. V. Khokhlov, Zh. Eksp. Teor. Fiz. **50**, 1537 (1966) [Sov. Phys.—JETP **23**, 1025 (1966)].
[11] L. A. Ostrovskiĭ, Zh. Eksp. Teor. Fiz. **51**, 1189 (1966) [Sov. Phys.—JETP **24**, 797 (1967)].
[12] P. L. Kelley, Phys. Rev. Letters **15**, 1005 (1965).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Critical Behavior of Several Lattice Models with Long-Range Interaction

MARK KAC
The Rockefeller University, New York, New York
AND
COLIN J. THOMPSON*
The Rockefeller University, New York, New York and Applied Mathematics Department
The University of New South Wales, Kensington, N.S.W., Australia

(Received 19 February 1969)

We consider a one-dimensional model with infinite-range interaction, a two-dimensional model, and a three-dimensional model, whose free energies can be expressed in terms of the largest eigenvalue of an integral equation. High- and low-temperature expansions in powers of the reciprocal of the range of the exponential part of the interaction, with the classical Curie–Weiss theory as leading term, are developed and studied in the critical region. We find that to leading order in the critical region the resummed high- and low-temperature expansions are analytic at the classical critical point but are nonanalytic at a displaced critical point. The modified singularities, which are no longer of Curie–Weiss type, give critical exponents which are identical with those obtained by Brout and others, and are almost surely not the true exponents. The technique, however, suggests a possible general method of successive approximation to true critical behavior.

1. INTRODUCTION

It is well known that critical behavior predicted by classical theories of phase transitions is in disagreement with experiment. On the other hand, Onsager's exact results for the two-dimensional Ising model and the extensive numerical results for the three-dimensional Ising model have striking similarities with experiment.[1] It is natural, therefore, that the classical theories have been neglected in favor of the short-range, or Onsagerian, theories.

The first step in the revival of the classical theories was taken by Brout,[2] who developed general perturbation expansions in the reciprocal range of interaction, with the classical theory as leading term. The hope was that the expansions in the critical region could be resummed to give results in accordance with the Onsagerian theories. Unfortunately, the expansions are only valid at temperatures above the classical critical point and to get beyond this point additional assumptions are required. Recent refinements[3] of Brout's method have failed to overcome this difficulty. A different tack has been taken by studying a class of models for which explicit expansions can be worked out at temperatures above and below the classical critical point as well as in the critical region. For a one-dimensional gas of hard rods attracting each other with a potential $-J\gamma \exp(-\gamma|x|)$ it has been shown[4] that for finite γ there is no phase transition, but that in the limit $\gamma \to 0$, the classical van der Waals theory results.[5] It has since been proved[6] that for a gas with attractive interaction in d dimensions of the form $\gamma^d \varphi(\gamma|\mathbf{r}|)$, the van der Waals theory results in the limit $\gamma \to 0$. Similar results hold for the corresponding lattice systems and in the magnetic case the $\gamma = 0$ limit gives the classical Curie–Weiss theory.

To avoid the necessity of going to the $\gamma = 0$ limit to produce a phase transition, and in so doing recover the classical theories, a number of two-dimensional lattice models were proposed[7] which could be reduced in essentially the same way as the one-dimensional model to the determination of the largest eigenvalue of an integral equation. High- and low-temperature expansions (in powers of the reciprocal range of interaction γ) were given in Ref. 7 and a preliminary study of the critical region (where the expansions break down) for one of the models (model A) was reported at the 1966 Brandeis summer school.[8] It was found that for a range of temperature of order $\gamma \log \gamma^{-1}$ around the classical critical point, the high- and low-temperature expansions become invalid and

* Present address: Applied Mathematics Department, Massachusetts Institute of Technology, Cambridge, Mass. 02139. Permanent address from September 1969: Northwestern University, Evanston, Ill.

[1] For comprehensive reviews and references see, for example, C. Domb, Advan. Phys. **9**, Nos. 34, 35 (1960); M. E. Fisher, *Boulder Lectures, 1964* (Univ. of Colorado Press, Boulder, Colo., 1965); and C. Domb, "Critical Phenomena," N.B.S. Misc. Pub. 273, Dec. 1966.

[2] R. Brout, Phys. Rev. **118**, 1009 (1960). For a review of this and later work see R. Brout, *Phase Transitions* (W. A. Benjamin, Inc., New York, 1965).

[3] For example: P. C. Hemmer, J. Math. Phys. **5**, 75 (1964); J. L. Lebowitz, G. Stell, and S. Baer, *ibid.* **6**, 1282 (1965); and G. Stell, J. L. Lebowitz, S. Baer, and W. Theumann, *ibid.* **7**, 1532 (1966).

[4] M. Kac, Phys. Fluids **2**, 8 (1959).

[5] M. Kac, G. E. Uhlenbeck, and P. C. Hemmer, J. Math. Phys. **4**, 216, 229 (1963); **5**, 60 (1964).

[6] J. L. Lebowitz and O. Penrose, J. Math. Phys. **7**, 98 (1966).

[7] M. Kac and E. Helfand, J. Math. Phys. **4**, 1078 (1963).

[8] M. Kac, *Brandeis Lectures, 1966* (Gordon and Breach Science Publishers, New York, 1968).

that a simple resummation gives new expansions which to first order [$\sim \gamma^2(\log \gamma^{-1})^2$] are analytic at the "old" classical singularity but yield a Curie–Weiss type singularity (i.e., a jump discontinuity in the specific heat) at a new "modified" critical point of order $\gamma \log \gamma^{-1}$ away from the old point. A one-dimensional model with a bona fide infinite range interaction $v(x)$, i.e., for which

$$\int_0^\infty xv(x)\,dx = \infty \tag{1.1}$$

independently of γ, was studied in the same way with essentially the same result. It was also shown that the resummation technique, when applied to the one exponential model and a one-dimensional model for which the integral in (1.1) is finite, yields new expansions which to leading order are analytic over the whole temperature range. It is felt, therefore, that by studying such perturbation expansions for various models it should be possible at least to "diagnose" a phase transition. It has in fact been proved[9] that the two-dimensional models of Ref. 7 have a phase transition, and quite recently[10,11] that the one-dimensional model considered here, for which $v(x)$ behaves essentially like $x^{-(1+a)}$ [see Eq. (3.1)], does not have a phase transition when $a > 1$, and does have a phase transition when $0 < a < 1$. There is no proof either way for $a = 1$, but the analysis given in Sec. 3 can be easily extended to $a = 1$ and one finds similar critical behavior to the two-dimensional model discussed in Sec. 4, suggesting that in this case there is also a phase transition. We remark that, contrary to a suggestion in Ref. 8, Dyson[11] has shown that (1.1) is, in general, not sufficient for a phase transition.

Our purpose here is to present calculations on the critical regions for a one-dimensional infinite-range interaction model, a two-dimensional model A, and a three-dimensional model. We carry the resummation procedure described in Ref. 8 to next order and find that to this order the singularity is no longer of Curie–Weiss type and that the specific heat actually diverges at the modified critical point. Unfortunately, the new critical behavior is almost surely incorrect. For example, in two dimensions one gets a logarithmically divergent energy, and in three dimensions a square-root specific-heat divergence, which are identical with the predictions of Brout's (and others) resummations.[2]

The advantages of the present method over the more general methods are these: (i) Low- as well as high-temperature expansions can be developed; and (ii) resummation shifts the critical point away from the classical value, presumably in the direction of the true critical point (the singularity in the general methods sticks at the classical value). Although our methods hint at a general underlying resummation scheme, it seems extremely difficult, if not impossible, to deduce the true nature of the singularity by such methods.

In the following section, we summarize the method used to reduce the problem of evaluating the free energy ψ for a class of models, to the determination of the largest eigenvalue of an integral equation, and present high- and low-temperature expansions for ψ. The critical regions for the one-dimensional model, for model A, and for the three-dimensional model are studied in Secs. 2, 3, and 4, respectively, and the main conclusions are summarized and discussed in the final section.

2. INTEGRAL EQUATION AND HIGH- AND LOW-TEMPERATURE EXPANSIONS FOR A ONE-DIMENSIONAL MODEL

Consider a one-dimensional chain of N spins, $\mu_i = \pm 1$, $i = 1, 2, \cdots, N$, with interaction energy given by

$$E = -J\gamma \sum_{1 \le i < j \le N} \rho(\gamma\,|i-j|)\mu_i\mu_j, \quad J > 0, \tag{2.1}$$

where

$$\rho(t) = \sum_{k=1}^{m} a_k \exp(-\sigma_k |t|); \quad a_k > 0, \quad \sigma_k > 0. \tag{2.2}$$

The problem is to evaluate the partition function

$$Q_N = \sum_{\mu_1=\pm 1} \cdots \sum_{\mu_N=\pm 1} \exp(-E/kT) \tag{2.3}$$

with E given by (2.1).

The simplifying feature in the reduction of (2.3) comes from the exponential nature of the interaction. Thus consider a set of m independent Ornstein–Uhlenbeck processes $x_1(t), \cdots, x_m(t)$ with zero mean and covariance

$$\langle x_k(t)x_k(t+\tau)\rangle = \exp(-\sigma_k|\tau|), \qquad k = 1, 2, \cdots, m, \tag{2.4}$$

where $\langle \cdots \rangle$ denotes probabilistic average. Independence implies that

$$\langle x_k(t)x_l(t+\tau)\rangle = 0, \quad \text{for} \quad k \neq l, \tag{2.5}$$

[9] R. B. Griffiths, J. Math. Phys. **8**, 478 (1967).

[10] D. Ruelle, Commun. Math. Phys. **9**, 267 (1968). Ruelle shows that if $v(n) \geq 0$ and if $\sum_{n=1}^{\infty} nv(n) < \infty$, the one-dimensional Ising model with interaction potential $v(n)$ has no phase transition to an ordered state.

[11] F. J. Dyson, to be published in Commun. Math. Phys. (1969). Dyson shows that if $v(n) \geq 0$ and is monotonically decreasing, there is no phase transition if $\lim_{N\to\infty} (\log \log N)^{-1} \sum_{n=1}^{N} nv(n) = 0$, extending Ruelle's result (Ref. 10), and if $v(n) > 0$ and

$$\sum_{n=1}^{\infty} \log\log(n+4)[n^3v(n)]^{-1} < \infty,$$

there is a phase transition.

and using (2.4) and (2.5), we can easily show that the stationary, Gaussian process $X(t)$ defined by

$$X(t) = \sum_{k=1}^{m} a_k^{\frac{1}{2}} x_k(t) \tag{2.6}$$

satisfies

$$\langle X(t)X(t+\tau)\rangle = \rho(\tau), \tag{2.7}$$

where $\rho(\tau)$ is given by (2.2). Symmetrizing (2.1), we can thus immediately write the partition function (2.3) in the form

$$\begin{aligned} Q_N &= \exp\left[-N\rho(0)\nu\gamma/2\right] \\ &\quad \times \sum_{\mu_1=\pm 1} \cdots \sum_{\mu_N=\pm 1} \left\langle \exp\left(\nu\gamma\right)^{\frac{1}{2}} \sum_{j=1}^{N} X(j)\mu_j \right\rangle \\ &= \{2\exp\left[-\rho(0)\nu\gamma/2\right]\}^N \left\langle \prod_{j=1}^{N} \cosh\left[(\nu\gamma)^{\frac{1}{2}} X(j)\right] \right\rangle, \end{aligned} \tag{2.8}$$

where ν is defined by

$$\nu = J/kT. \tag{2.9}$$

Now, since the vector process $\mathbf{x}(t) = (x_1(t), \cdots, x_m(t))$ is Markoffian, its joint probability density (by analogy with the one exponential model[7]) can be written as

$$W(\mathbf{x}_1, \cdots, \mathbf{x}_N) = W(\mathbf{x}_1) \prod_{k=1}^{N-1} P_\gamma(\mathbf{x}_k \mid \mathbf{x}_{k+1}), \tag{2.10}$$

where

$$W(\mathbf{x}) = \prod_{k=1}^{m} \{\exp\left(-x_k^2/2\right)/(2\pi)^{\frac{1}{2}}\} \tag{2.11}$$

and

$$\begin{aligned} &P_\gamma(\mathbf{x} \mid \mathbf{y}) \\ &= \prod_{k=1}^{m} \left\{ \exp\left[-\frac{(y_k - x_k e^{-\gamma\sigma_k})^2}{2(1 - e^{-2\gamma\sigma_k})} \right] \Big/ \left[2\pi(1 - e^{-2\gamma\sigma_k})\right]^{\frac{1}{2}} \right\}. \end{aligned} \tag{2.12}$$

The standard iteration argument[8] can now be applied to (2.8), and one has finally that the free energy ψ (in the thermodynamic limit) can be expressed in the form

$$\begin{aligned} -\frac{\psi}{kT} &= \lim_{N\to\infty} \frac{1}{N} \log Q_N \\ &= -\frac{\rho(0)\nu\gamma}{2} + \log 2 + \log \lambda_{\max}, \end{aligned} \tag{2.13}$$

where $\lambda_{\max}$ is the maximum eigenvalue of the integral equation

$$\int_{-\infty}^{\infty} \cdots \int K(\mathbf{x}, \mathbf{y})\phi(\mathbf{y})\, dy_1 \cdots dy_m = \lambda\phi(\mathbf{x}) \tag{2.14}$$

($\mathbf{x}$ and $\mathbf{y}$ denote m-dimensional vectors) with the kernel $K(\mathbf{x}, \mathbf{y})$ defined by

$$\begin{aligned} K(\mathbf{x}, \mathbf{y}) &= \left\{ \cosh\left[(\nu\gamma)^{\frac{1}{2}} \sum_{k=1}^{m} a_k^{\frac{1}{2}} x_k \right] \right\}^{\frac{1}{2}} \left[\frac{W(\mathbf{x})}{W(\mathbf{y})} \right]^{\frac{1}{2}} P_\gamma(\mathbf{x} \mid \mathbf{y}) \\ &\quad \times \left\{ \cosh\left[(\nu\gamma)^{\frac{1}{2}} \sum_{k=1}^{m} a_k^{\frac{1}{2}} y_k \right] \right\}^{\frac{1}{2}}. \end{aligned} \tag{2.15}$$

For small γ, it can be shown from (2.14) and (2.15) (details can be found in Ref. 8) that to order γ^2

$$\lambda_{\max} = \exp\left(\frac{\gamma}{2} \sum_{k=1}^{m} \sigma_k - \gamma E_0 \right), \tag{2.16}$$

where E_0 is the smallest eigenvalue of the differential equation

$$\begin{aligned} &\sum_{k=1}^{m} \frac{\partial^2\psi}{\partial x_k^2} - \left\{ \tfrac{1}{4} \sum_{k=1}^{m} \sigma_k^2 x_k^2 \right. \\ &\qquad \left. - \frac{1}{\gamma} \log\cosh\left[(\nu\gamma)^{\frac{1}{2}} \sum_{k=1}^{m} (a_k\sigma_k)^{\frac{1}{2}} x_k \right] \right\} \psi = -E\psi. \end{aligned} \tag{2.17}$$

The first thing to note is that in the limit $\gamma \to 0$ one recovers the classical Curie–Weiss theory from the above results. A simple calculation shows that the minimum of the potential in (2.17) (the term in curly brackets) occurs at $\mathbf{x} = \mathbf{x}^{(0)}$, where

$$x_k^{(0)} = 2(\nu/\gamma)^{\frac{1}{2}} [(a_k\sigma_k)^{\frac{1}{2}}/\sigma_k^2] \tanh\eta \tag{2.18}$$

and η is a solution of the equation

$$\eta = (2\nu\nu_0) \tanh\eta \tag{2.19}$$

with

$$\nu_0 = \sum_{k=1}^{m} \frac{a_k}{\sigma_k} = \int_0^\infty \rho(t)\, dt < \infty. \tag{2.20}$$

The condition $\nu_0 < \infty$ is simply a stability condition which is necessary to ensure the existence of the thermodynamic limit. $2\nu_c = \nu_0^{-1}$ is seen immediately to be the classical Curie point and in the limit $\gamma \to 0$, E_0 is just the minimum of the potential. Thus, by substituting (2.18) into the potential of (2.17) and using (2.16), we have that

$$\begin{aligned} \lim_{\gamma\to 0} \lambda_{\max} = \mathcal{K}(\nu) &= 1, \quad 2\nu \le \nu_0^{-1}, \\ &= \exp\left[\log\cosh\eta_0 - (\nu\nu_0)\tanh^2\eta_0\right], \\ &\qquad 2\nu > \nu_0^{-1}, \end{aligned} \tag{2.21}$$

where η_0 is the positive solution of (2.19).

For small but finite γ, standard perturbation theory can be applied to (2.17). Thus, for high temperatures ($2\nu < \nu_0^{-1}$), the log cosh function in (2.17) can be expanded directly, giving to order γ

$$\begin{aligned} &\sum_{k=1}^{m} \frac{\partial^2\psi}{\partial x_k^2} - \left\{ \tfrac{1}{4} \sum_{k=1}^{m} \sigma_k^2 x_k^2 - \frac{\nu}{2} \left[\sum_{k=1}^{m} (a_k\sigma_k)^{\frac{1}{2}} x_k \right]^2 \right. \\ &\qquad \left. + \frac{\nu^2\gamma}{12} \left[\sum_{k=1}^{m} (a_k\sigma_k)^{\frac{1}{2}} x_k \right]^4 \right\} \psi = -E\psi. \end{aligned} \tag{2.22}$$

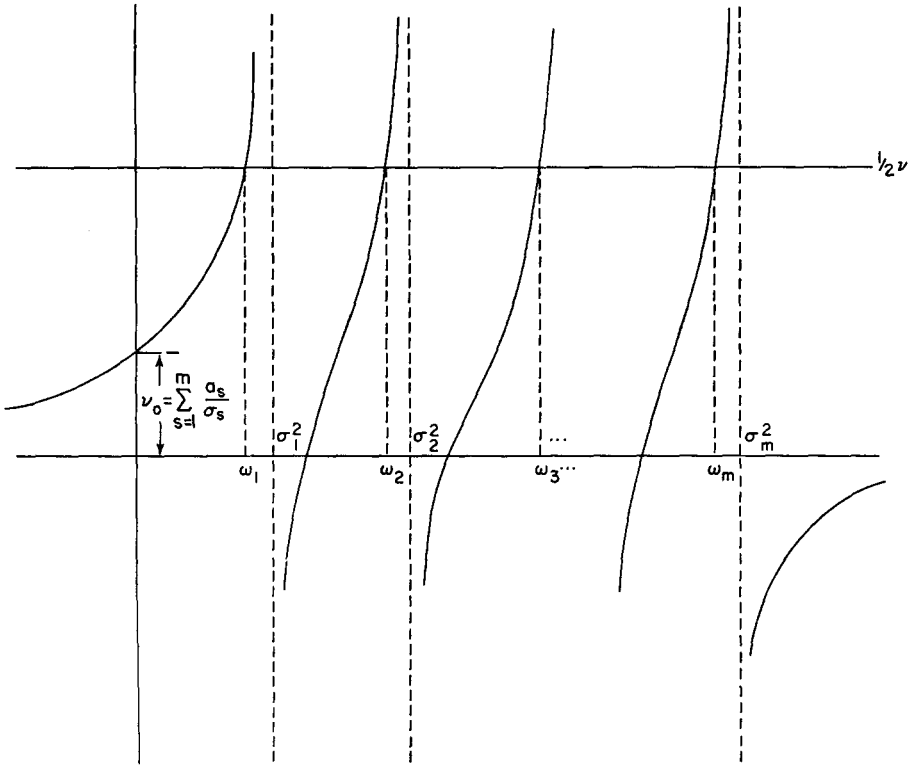

FIG. 1. Eigenvalue solutions of Eq. (2.24).

The eigenvalues and eigenfunctions of the quadratic form in (2.22) are obtained easily by minimizing

$$\tfrac{1}{4}\sum_{k=1}^{m}\sigma_k^2 x_k^2 - \frac{\nu}{2}\left[\sum_{k=1}^{m}(a_k\sigma_k)^{\frac{1}{2}}x_k\right]^2 - \frac{\omega}{4}\sum_{k=1}^{m}x_k^2. \quad (2.23)$$

One thus finds that the eigenvalues ω_s are solutions of the equation (shown schematically in Fig. 1)

$$\frac{1}{2\nu} = \sum_{k=1}^{m}\frac{a_k\sigma_k}{\sigma_k^2 - \omega}, \quad (2.24)$$

and the corresponding normalized eigenfunctions are $\mathbf{a}^s = (a_1^s, a_2^s, \cdots, a_m^s)$, where

$$a_k^s = \frac{1}{N_s}\frac{(a_k\sigma_k)^{\frac{1}{2}}}{(\sigma_k^2 - \omega_s)}; \quad N_s^2 = \sum_{k=1}^{m}\frac{a_k\sigma_k}{(\sigma_k^2 - \omega_s)^2}. \quad (2.25)$$

Applying the orthogonal transformation

$$x_k = \sum_{s=1}^{m} a_k^s y_s, \quad (2.26)$$

we find that (2.17) becomes

$$\sum_{s=1}^{m}\frac{\partial^2\phi}{\partial y_s^2} - \left[\tfrac{1}{4}\sum_{s=1}^{m}\omega_s y_s^2 + \alpha\gamma\left(\sum_{s=1}^{m}\frac{y_s}{N}\right)^4\right]\phi = -E\phi, \quad (2.27)$$

where

$$\alpha = \nu^2/12(2\nu)^4 \quad (2.28)$$

and

$$\phi(\{y_s\}) = \psi\left(\left\{\sum_{k=1}^{m}a_k^s y_s\right\}\right). \quad (2.29)$$

The unperturbed part of (2.27), namely

$$\sum_{s=1}^{m}\frac{\partial^2}{\partial y_s^2} - \tfrac{1}{4}\sum_{s=1}^{m}\omega_s y_s^2, \quad (2.30)$$

is the Schrödinger operator for a set of m uncoupled oscillators, with eigenvalues

$$\lambda_{\mathbf{k}}^0 = \sum_{s=1}^{m}(k_s + \tfrac{1}{2})\omega_s^{\frac{1}{2}}, \quad k_s = 0, 1, 2, \cdots, \quad (2.31)$$

and corresponding normalized eigenfunctions

$$\psi_{\mathbf{k}}^0(x_1, \cdots, x_m) = \prod_{s=1}^{m}\left[\frac{\omega_s^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}k_s!}\right]^{\frac{1}{2}} H_{k_s}(\omega_s^{\frac{1}{4}}x_s) \times \exp(-\tfrac{1}{4}\omega_s^{\frac{1}{2}}x_s^2), \quad (2.32)$$

where the $H_n(x)$ are Hermite polynomials [note that because $2\nu < \nu_0^{-1}$, it follows from (2.24) that all the ω_s are positive].

In terms of $\lambda_{\mathbf{k}}^0$ and $\psi_{\mathbf{k}}^0$, the ground state of (2.27) can be written as

$$E_0 = \lambda_0^0 + \lambda_0^1 + \lambda_0^2 + \cdots, \quad (2.33)$$

where, from (2.31),

$$\lambda_0^0 = \tfrac{1}{2}\sum_{s=1}^{m}\omega_s^{\frac{1}{2}}, \quad (2.34)$$

and by ordinary perturbation theory

$$\lambda_0^1 = \alpha\gamma \int_{-\infty}^{\infty} \cdots \int [\psi_0^0(\mathbf{y})]^2 \left(\sum_{s=1}^{m} \frac{y_s}{N_s}\right)^4 = 3\alpha\gamma\rho^2, \quad (2.35)$$

where ρ is defined by

$$\rho = \sum_{s=1}^{m} \frac{1}{N_s^2 \omega_s^{\frac{1}{2}}}, \quad (2.36)$$

and

$$\lambda_0^2 = -\sum_{\mathbf{k}\neq 0}^{m} (\lambda_{\mathbf{k}}^0 - \lambda_0^0)^{-1} \left(\int_{-\infty}^{\infty} \cdots \int \psi_0^0(\mathbf{y}) \alpha\gamma \left(\sum_{s=1}^{m} \frac{y_s}{N_s}\right)^4 \times \psi_{\mathbf{k}}^0(\mathbf{y})\, dy_1 \cdots dy_m \right)^2, \quad (2.37)$$

etc.

Making use of the results

$$(\lambda_{\mathbf{k}}^0 - \lambda_0^0)^{-1} = \left(\sum_{s=1}^{m} k_s \omega_s^{\frac{1}{2}}\right)^{-1} = \int_0^{\infty} \exp\left(-t \sum_{s=1}^{m} k_s \omega_s^{\frac{1}{2}}\right) \quad (2.38)$$

and

$$\exp\left[-\tfrac{1}{2}(x^2 + y^2)\right] \sum_{k\neq 0}^{\infty} \frac{H_k(x) H_k(y) e^{-k\tau}}{k!\,(2\pi)^{\frac{1}{2}}} = L(x, y, \tau) - L(x, y, \infty), \quad (2.39)$$

where

$$L(x, y, \tau) = [2\pi(1 - e^{-2\tau})]^{-\frac{1}{2}} \times \exp\left[-(x^2 - 2xye^{-\tau} + y^2)/2(1 - e^{-2\tau})\right], \quad (2.40)$$

we have from (2.37) that

$$\begin{aligned}
\lambda_0^2 = -\int_0^{\infty} dt \Bigg\{ & \text{the coefficient of } (4!)^2(\alpha\gamma)^2(uv)^4 \text{ in} \\
& \int_{-\infty}^{\infty} \cdots \int dx_1 \cdots dx_m \\
& \times \exp\left(u \sum_{s=1}^{m} x_s / N_s^4 \omega_s^{\frac{1}{2}}\right) \int_{-\infty}^{\infty} \cdots \int dy_1 \cdots dy_m \\
& \times \exp\left(v \sum_{s=1}^{m} y_s / N_s^4 \omega_s^{\frac{1}{2}}\right) \\
& \times \left[\prod_{s=1}^{m} L(x_s, y_s, t\omega_s^{\frac{1}{2}}) - \prod_{s=1}^{m} L(x_s, y_s, \infty) \right] (2\pi)^{-m/2} \Bigg\} \\
= -(4!\,\alpha\gamma)^2 \times & \text{ the coefficient of } (uv)^4 \text{ in} \\
& \exp\left[\tfrac{1}{2}\rho(u^2 + v^2)\right] \\
& \times \int_0^{\infty} dt \left\{ \exp\left(uv \sum_{s=1}^{m} e^{-\omega_s^{\frac{1}{2}} t} / N_s^2 \omega_s^{\frac{1}{2}}\right) - 1 \right\} \\
= -24(\alpha\gamma)^2 & \{T_4 + 3\rho^2 T_2\}, \quad (2.41)
\end{aligned}$$

where ρ is defined by (2.36) and T_n by

$$T_n = \int_0^{\infty} \left(\sum_{s=1}^{m} e^{-\omega_s^{\frac{1}{2}} t} / N_s^2 \omega_s^{\frac{1}{2}} \right)^n dt. \quad (2.42)$$

Finally, combining (2.41), (2.35), (2.34), and (2.33), we have, to order γ^2,

$$E_0 = \tfrac{1}{2} \sum_{s=1}^{m} \omega_s^{\frac{1}{2}} + 3\alpha\gamma\rho^2 - 24(\alpha\gamma)^2 [T_4 + 3\rho^2 T_2] + \cdots, \quad (2.43)$$

where ρ and T_n are defined by (2.36) and (2.42), respectively.

To develop low-temperature expansions ($2\nu \geq \nu_0^{-1}$), the potential in (2.17) must first be expanded around its minimum. Equation (2.17) then becomes, after making the change of variables (2.26),

$$\sum_{s=1}^{m} \frac{\partial^2 \phi}{\partial y_s^2} - \left\{ \tfrac{1}{4} \sum_{s=1}^{m} \omega_s' y_s^2 + \beta' \gamma^{\frac{1}{2}} \left(\sum_{s=1}^{m} \frac{y_s}{N_s'} \right)^3 + \alpha' \gamma \left(\sum_{s=1}^{m} \frac{y_s}{N_s'} \right)^4 \right\} \phi = -(E - \gamma^{-1} \log \mathcal{K}(\nu))\phi, \quad (2.44)$$

where ω_s' and N_s' are obtained from ω_s and N_s by replacing ν by ν', which is defined by

$$\nu' = \nu(1 - \eta^2 / 2\nu\nu_0); \quad (2.45)$$

α', β' are defined, respectively, by

$$\alpha' = \frac{\nu^2}{12(2\nu')^4} \left(1 - \frac{\eta^2}{2\nu\nu_0}\right) \left(1 - \frac{3\eta^2}{2\nu\nu_0}\right) \quad (2.46)$$

and

$$\beta' = \frac{\nu^{\frac{3}{2}}}{3(2\nu')^3} \eta (2\nu\nu_0)^{-\frac{1}{2}} \left(1 - \frac{\eta^2}{2\nu\nu_0}\right), \quad (2.47)$$

and, recalling (2.21), $\gamma^{-1} \log \mathcal{K}(\nu)$ is just the minimum of the potential.

It follows from (2.45) and (2.24) that $\omega_s' > 0$ when $2\nu > \nu_0^{-1}$; so straightforward perturbation theory can now be applied to (2.44) with the result

$$\begin{aligned}
E_0 = \gamma^{-1} \log \mathcal{K}(\nu) & + \tfrac{1}{2} \sum_{s=1}^{m} \omega_s'^{\frac{1}{2}} \\
& + 3\alpha'\gamma\rho'^2 - 6\beta'^2\gamma [T_3' + \tfrac{1}{2} 3\rho'^2 T_1'] \\
& - 24(\alpha'\gamma)^2 [T_4' + 3\rho'^2 T_2'] + \cdots. \quad (2.48)
\end{aligned}$$

When $|2\nu - \nu_0^{-1}|$ is sufficiently small (depending on γ), the expansions (2.43) and (2.48) for E_0 become invalid in the sense that terms of the same order are present in all orders of perturbation. We call the range of ν's for which the expansions are invalid the critical region, and this is the subject of the following sections.

3. CRITICAL REGION OF A ONE-DIMENSIONAL MODEL

For the sake of definiteness, we take the interaction potential $\rho(t)$ in (2.1) to be

$$\rho(t) = \int_0^1 \lambda^a e^{-\lambda|t|}\, d\lambda, \quad a > 0. \quad (3.1)$$

Discretized, this becomes

$$\rho(t) \sim \sum_{k=1}^{m} (k\Delta\lambda)^a \Delta\lambda e^{-(k\Delta\lambda)|t|}, \quad \Delta\lambda = m^{-1}, \quad (3.2)$$

i.e., in the notation of the previous section,

$$a_k = (k\Delta\lambda)^a \Delta\lambda \quad \text{and} \quad \sigma_k = k\Delta\lambda, \quad (3.3)$$

and the free energy is obtained from (2.13) by letting $m \to \infty$. It is easily verified that

$$\nu_0 = \int_0^\infty \rho(t)\, dt = a^{-1}$$

and

$$\int_0^\infty t\rho(t)\, dt = \int_0^1 \lambda^{a-2}\, d\lambda < \infty, \quad a > 1,$$
$$= \infty, \quad a < 1. \quad (3.4)$$

Since the integral in (3.4) is in some sense a measure of the range of interaction, one might expect peculiar phenomena for the case $a < 1$ when the integral diverges. In fact, it has been proved[11] that in this case the system exhibits a phase transition. We demonstrate how this comes about by studying the critical region for $0 < a < 1$. $a > 1$ is discussed in Ref. 8 and the following analysis can be easily extended to $a = 1$ [where the integral (3.4) also diverges]. In this case ρ [Eq. (3.6)] diverges logarithmically at $2\nu = a$, so the analysis is almost identical with that given in the following section for the two-dimensional model A [see Eq. (4.15)]; this suggests that for $a = 1$ there is also a phase transition. (There is as yet no proof that this is or is not the case.) Another demonstration, given in Ref. 8 and perhaps of some interest, shows that for $a \leq 1$ and sufficiently low temperatures, the maximum eigenvalue of the integral equation (2.14) is asymptotically degenerate, which implies, as for the ordinary nearest-neighbor Ising model, that long-range order exists. We concern ourselves here, however, with an analysis of the critical region.

We define the critical region to be that range of ν values for which the ν-dependent part of the unperturbed ground-state energy for E_0 is of the same order as the first-order perturbation correction. At high temperatures, from (2.43), this means

$$\sum_{s=1}^{m} \{[\omega_s(2\nu)]^{\frac{1}{2}} - [\omega_s(a)]^{\frac{1}{2}}\} \sim 3\alpha\gamma\rho^2.$$

In the Appendix it is shown that, for small $a - 2\nu > 0$ and $0 < a < 1$,

$$\lim_{m\to\infty} \sum_{s=1}^{m} \{[\omega_s(2\nu)]^{\frac{1}{2}} - [\omega_s(a)]^{\frac{1}{2}}\}$$
$$= A(a - 2\nu) + B(a - 2\nu)^{1/a} + \cdots \quad (3.5)$$

and, in the limit $m \to \infty$,

$$\rho = 2a^2(A + (B/a)(a - 2\nu)^{(1/a)-1} + \cdots), \quad (3.6)$$

where

$$A = \pi^{-1} \int_0^\infty d\xi \left[1 - a\xi^a \int_0^{1/\xi} \frac{v^{a-1}}{1+v^2}\, dv\right]$$
$$\times \left[a^2\xi^a \int_0^{1/\xi} \frac{v^{a-1}}{1+v^2}\, dv\right]^{-1} \quad (3.7)$$

and

$$B = \operatorname{cosec}(\pi/a)[\tfrac{1}{2}a^2\pi \operatorname{cosec}(\tfrac{1}{2}a\pi)]^{1/a}. \quad (3.8)$$

For $a < 1$, therefore, the critical region is defined by

$$a - 2\nu \sim \gamma, \quad (3.9)$$

and it is appropriate in this region to define a new temperature variable ν_1 by

$$\nu_1 = (2\nu - a)\gamma^{-1}. \quad (3.10)$$

If one examines the low-temperature expansion (2.48), one arrives at the same result (3.9) with $2\nu - a$ in place of $a - 2\nu$, so that (3.10) is appropriate for high ($\nu_1 < 0$) and low ($\nu_1 > 0$) temperatures.

Substitution of (3.10) into (2.43) gives the high-temperature ($\nu_1 < 0$) critical-region expansion for E_0. It is not difficult to show, however, that terms of order $\gamma^{1/a}$ are present in all orders of perturbation so that to calculate the precise coefficient of $\gamma^{1/a}$ one has to calculate and sum the "most divergent terms" to all orders of perturbation. To second order (which is as far as we have gone), using the fact that

$$T_2 = -2Ba^2(1 - a)(a - 2\nu)^{1/a-2} + \cdots \quad (3.11)$$

(see the Appendix), where B is given by (3.8), and in general that

$$T_n \sim \gamma^{((n-1)/a-n)} \quad \text{for} \quad a - 2\nu \sim \gamma$$
$$\text{and} \quad n = 1, 2, \cdots, \quad (3.12)$$

and in addition the expansions (3.5) and (3.6), we find that Eq. (2.43) for E_0 becomes[12]

$$E_0 = \tfrac{1}{2} \sum_{s=1}^{m} [\omega_s(a)]^{\frac{1}{2}} + \gamma\left[\frac{(aA)^2}{4} - \frac{\nu_1 A}{2}\right]$$
$$+ \frac{B}{2}|\nu_1\gamma|^{1/a}\left[1 - \frac{aA}{\nu_1} + \frac{a^{-1}(a^{-1}-1)}{2!}\left(\frac{a^2A}{\nu_1}\right)^2 + \cdots\right]; \quad (3.13)$$

[12] In the computation of $\lambda_{\max}$ [Eq. (2.16)], and therefore ψ, we have a term $\lim_{m\to\infty} \frac{1}{2}\gamma \sum_{s=1}^{m} (s\Delta\lambda - \omega_s^{\frac{1}{2}})$, which is finite.

$$\lim_{m\to\infty} \tfrac{1}{2} \sum_{s=1} \omega_s^{\frac{1}{2}}$$

is infinite, so that in (3.13) and subsequent formulas it must be remembered that $\frac{1}{2}\sum_{s=1}^{m} [\omega_s(a)]^{\frac{1}{2}}$ is to be combined with $-\frac{1}{2}\sum_{s=1}^{m} s\Delta\lambda$ before letting $m \to \infty$.

and without actually calculating the remaining coefficients in the expansion of the $\gamma^{1/a}$ term (coming from higher-order perturbation), it seems clear that the coefficient in question (remembering that $\nu_1 < 0$) is simply $(a^2A - \nu_1)^{1/a}$. We, therefore, make the following *conjecture*:

$$E_0 = \tfrac{1}{2}\sum_{s=1}^{m}[\omega_s(a)]^{\frac{1}{2}} + \gamma\left[\frac{(aA)^2}{4} - \frac{\nu_1 A}{2}\right] + \frac{B\gamma^{1/a}}{2}(a^2A - \nu_1)^{1/a} + \cdots, \quad \text{for } \nu_1 < 0. \tag{3.14}$$

If the conjecture is correct, then clearly the range of validity of the expansion should be $\nu_1 < a^2A$, which is beyond the range of validity of the original high-temperature expansion (2.43). Note also that (to order $\gamma^{1/a}$, at least) (3.14) is analytic at the old classical singularity ($\nu_1 = 0$) but now has a singularity at $\nu_1 = a^2A$, which is precisely the position of the modified singularity deduced rigorously in Ref. 8.

Let us now, as a check on the validity of (3.14), examine the low-temperature expansion (2.48). In the critical region we have, to leading order, from (2.19), (2.21), (2.45), (2.46), (2.47), and (3.10), that

$$\begin{aligned} \log \mathcal{K}(v) &= -3\nu_1^2\gamma^2/4a^2, \\ 2v' &= a - 2\nu_1\gamma, \\ \alpha' &= (48a^2)^{-1} = \alpha, \end{aligned} \tag{3.15}$$

and

$$\beta' = (2a^2)^{-1}(\nu_1\gamma/16)^{\frac{1}{2}} = \beta\gamma^{\frac{1}{2}}.$$

The differential equation (2.44), in the critical region, then becomes

$$\sum_{s=1}^{m}\frac{\partial^2\phi}{\partial y_s^2} - \left\{\tfrac{1}{4}\sum_{s=1}^{m}\omega_s' y_s^2 + \beta\gamma\left(\sum_{s=1}^{m}\frac{y_s}{N_s'}\right)^3 + \alpha\gamma\left(\sum_{s=1}^{m}\frac{y_s}{N_s'}\right)^4\right\}\phi = -\left(E + \frac{3\nu_1^2\gamma}{4a^2}\right)\phi, \tag{3.16}$$

and straightforward perturbation gives Eq. (2.48) for E_0 with the replacements (3.15). Note, however, from (3.12), that $T_1' \sim \gamma^{-1}$ and, since $\beta' = \beta\gamma^{\frac{1}{2}}$ to leading order, there is a first-order term in (2.48), viz., $-9\beta^2\gamma^2 T_1'$, coming from second-order perturbation. The reason for this is that in the critical region, the minimum of the potential is no longer the correct point to expand around. To remedy this difficulty, we perform a shift in (2.44), i.e., replace y_s by $y_s + y_s^{(0)}$, before applying perturbation, and then choose the $y_s^{(0)}$ to eliminate the linear term in second order (the resulting $y_s^{(0)}$'s, as it turns out, also minimize E_0 to that order). Thus substituting $y_s + y_s^{(0)}$ for y_s in (3.16) gives

$$\sum_{s=1}^{m}\frac{\partial^2\phi}{\partial y_s^2} - \left\{\tfrac{1}{4}\sum_{s=1}^{m}\omega_s' y_s^2 + \tfrac{1}{4}\sum_{s=1}^{m}\omega_s'(y_s^{(0)})^2 + \left[\tfrac{1}{2}\sum_{s=1}^{m}\omega_s' y_s y_s^{(0)} + \beta\gamma\left(S + \sum_{s=1}^{m}\frac{y_s}{N_s}\right)^3 + \alpha\gamma\left(S + \sum_{s=1}^{m}\frac{y_s}{N_s'}\right)^4\right]\right\}\phi = -\left(E + \frac{3\nu_1^2\gamma}{4a^2}\right)\phi, \tag{3.17}$$

where the square-bracketed term is to be considered as the perturbation, and S is defined by

$$S = \sum_{s=1}^{m} y_s^{(0)}/N_s'. \tag{3.18}$$

Application of ordinary perturbation theory to (3.17) gives

$$E_0 = -\frac{3\nu_1^2\gamma}{4a^2} + E_0^{(0)} + E_0^{(1)} + E_0^{(2)} + \cdots, \tag{3.19}$$

where, by applying the same method used in obtaining (2.43), we have

$$E_0^{(0)} = \tfrac{1}{2}\sum_{s=1}^{m}\omega_s'^{\frac{1}{2}} + \tfrac{1}{4}\sum_{s=1}^{m}\omega_s'(y_s^{(0)})^2, \tag{3.20}$$

$$E_0^{(1)} = \beta\gamma(S^3 + 3\rho' S) + \alpha\gamma(S^4 + 6\rho' S^2 + 3\rho'^2), \tag{3.21}$$

and

$$\begin{aligned} E_0^{(2)} = -\Big\{&\tfrac{1}{4}\sum_{s=1}^{m}\omega_s'(y_s^{(0)})^2 + 3\beta\gamma S(S^2 + \rho') \\ &+ 4\alpha\gamma S(S^3 + 3\rho' S) \\ &+ T_1'[4\alpha\gamma(S^3 + 3\rho' S) + 3\beta\gamma(S^2 + \rho')]^2\Big\} \\ &- 18T_2'[2\alpha\gamma(S^2 + \rho') + \beta\gamma S]^2 \\ &- 6T_3'[4\alpha\gamma S + \beta\gamma]^2 - 24(\alpha\gamma)^2 T_4'. \end{aligned} \tag{3.22}$$

Minimizing $E_0^{(0)} + E_0^{(1)}$ with respect to $y_s^{(0)}$ gives

$$\tfrac{1}{2}y_s^{(0)} = -(N_s'\omega_s')^{-1}[4\alpha\gamma S^3 + 12\alpha\gamma\rho' S + 3\beta\gamma S^2 + 3\beta\gamma\rho'], \tag{3.23}$$

and multiplying both sides by $(N_s')^{-1}$ and summing over s, using (3.15) and the fact that

$$T_1' = \sum_{s=1}^{m}\frac{1}{N_s'^2\omega_s'} = \frac{(2v')^2}{a - 2v'} = \frac{a^2}{2\nu_1\gamma} \tag{3.24}$$

in the critical region, gives

$$S^3 + 3(6\nu_1)^{\frac{1}{2}}S^2 + (12\nu_1 + 3\rho')S + 3\rho'(6\nu_1)^{\frac{1}{2}} = 0, \tag{3.25}$$

or

$$S = -(6\nu_1)^{\frac{1}{2}} \quad \text{or} \quad -(6\nu_1)^{\frac{1}{2}} \pm (6\nu_1 - 3\rho')^{\frac{1}{2}}. \tag{3.26}$$

For either value of S in (3.26), the term in braces in (3.22), which is the unwanted linear term in second order, vanishes.

To obtain the minimum value for E_0 to second order, substitution of $S = -(6\nu_1)^{\frac{1}{2}}$ in the above is appropriate if $\nu_1 \leq \frac{1}{2}\rho'$; and if $\nu_1 \geq \frac{1}{2}\rho'$,

$$S = (6\nu_1)^{\frac{1}{2}} \pm (6\nu_1 - 3\rho')^{\frac{1}{2}}$$

is appropriate (minimization of $E_0^{(0)} + E_0^{(1)} + E_0^{(2)}$ gives the same result to second order). We, therefore, have that

$$E_0 = \tfrac{1}{2}\sum_{s=1}^{m}\omega_s'^{\frac{1}{2}} - \frac{3\nu_1\rho'\gamma}{4a^2} + 3\alpha\gamma\rho'^2 - 72(\alpha\gamma)^2 T_2'(6\nu_1 - \rho')^2 + \cdots, \quad \text{for } 0 < \nu_1 \leq \tfrac{1}{2}\rho', \tag{3.27}$$

and

$$E_0 = \tfrac{1}{2}\sum_{s=1}^{m}\omega_s'^{\frac{1}{2}} - \frac{3\nu_1^2\gamma}{4a^2} - \frac{\rho'^2\gamma}{8a^2} - 72(\alpha\gamma)^2 T_2'(-2\rho')^2 + \cdots, \quad \text{for } \nu_1 \geq \tfrac{1}{2}\rho', \tag{3.28}$$

and it is not difficult to show that, to order γ, these formulas are identical with those obtained rigorously in Ref. 8 [Eqs. (6.112) and (6.115), respectively]. Notice that the modified singularity occurs at $2\nu = a + \nu_{1c}\gamma$, where from (3.6), to leading order,

$$\nu_{1c} = \tfrac{1}{2}\rho' = a^2 A, \tag{3.29}$$

and this is the value predicted in the conjectured critical-region expansion (3.14). Clearly, as a final check on our conjectured resummation, (3.27) should reduce to (3.14).

It is not difficult to show that terms of order $\gamma^{1/a}$ occur in all orders of perturbation below the classical critical point just as for above, so that, again to compute the precise coefficient of $\gamma^{1/a}$, one must sum the most divergent terms to all orders of perturbation. We have, however, only carried the perturbation to second order. Nevertheless, if we substitute the expansions (3.5), (3.6), and (3.11), with ν replaced by ν' given in (3.15), into the expansion (3.27), we find that, for $0 < \nu_1 \leq a^2 A$,

$$E_0 = \tfrac{1}{2}\sum_{s=1}^{m}[\omega_s(a)]^{\frac{1}{2}} + \gamma\left[\frac{(aA)^2}{4} - \frac{\nu_1 A}{2}\right] + (2\nu_1\gamma)^{1/a}\left\{1 - \frac{3}{2a} + \frac{aA}{2\nu_1} + \frac{a^{-1}(a^{-1}-1)}{2!}\left(\frac{3}{2}\right)^2 - \frac{3A}{4\nu_1}(1-a) + \frac{a^{-1}(a^{-1}-1)}{2!}\left(\frac{a^2A}{2\nu_1}\right)^2 + \cdots\right\}. \tag{3.30}$$

By writing

$$(a^2A - \nu_1)^{1/a} = (2\nu_1)^{1/a}\left\{1 - \left(\frac{3}{2} - \frac{a^2A}{2\nu_1}\right)\right\}^{1/a},$$

the terms of the coefficient of $\gamma^{1/a}$ given in (3.30) are recognized immediately as the first three terms in the expansion of $(a^2A - \nu_1)^{1/a}$ in powers of $(\frac{3}{2} - a^2A/2\nu_1)$. This observation, we believe, supports very strongly the conjecture that (3.14) is the correct expansion of E_0 to order $\gamma^{1/a}$ for $\nu_1 \leq a^2 A$.

The conjectures leading to (3.14) can be stated more clearly (and hence be generalized to the models considered in the following two sections), and the peculiar nature of the expansion for $0 < \nu_1 \leq a^2 A$ [in powers of $(\frac{3}{2} - a^2A/2\nu_1)$!] can be better understood if one goes back to the expansions (3.5), (3.6), and (3.11) and notes that to the order recorded

$$\rho = -2a^2\frac{df}{d\nu} \tag{3.31}$$

and

$$T_2 = -a^4\frac{d^2f}{d\nu^2}, \tag{3.32}$$

where $f(\nu)$ is defined by

$$f(\nu) = \tfrac{1}{2}\sum_{s=1}^{m}\omega_s^{\frac{1}{2}} = \tfrac{1}{2}\sum_{s=1}^{m}[\omega_s(a)]^{\frac{1}{2}} + A(a - 2\nu) + B(a - 2\nu)^{1/a} + \cdots. \tag{3.33}$$

Using these facts, we find that the high-temperature expansion (2.43) can be written as

$$E_0 = f(\nu) - \frac{1}{2}\left(\frac{\rho\gamma}{4}\right)\frac{df}{d\nu} + \frac{1}{2!}\left(\frac{\rho\gamma}{4}\right)^2\frac{d^2f}{d\nu^2} + \cdots, \tag{3.34}$$

and the conjecture leading to (3.14) can now be stated as follows: The $\gamma^{1/a}$ term (or the most divergent term) in nth-order perturbation is given by

$$\frac{(-1)^n}{n!}\left(\frac{\rho\gamma}{4}\right)^n\frac{d^nf}{d\nu^n}, \quad \text{for } n \geq 3. \tag{3.35}$$

If this is so, then (3.34) becomes, to order $\gamma^{1/a}$,

$$E_0 = f\left(\nu - \frac{\gamma\rho}{4}\right) + \frac{1}{2}\left(\frac{\gamma\rho}{4}\right)\frac{df}{d\nu}, \tag{3.36}$$

and it should be noted that the second term in (3.36) cancels against the $\gamma^{1/a}$ term proportional to A in $f(\nu - \gamma\rho/4)$ [see Eq. (3.5)]. It is then straightforward to reduce (3.36) to (3.14).

Now when $0 < \nu_1 \leq a^2 A$, E_0 given by (3.27) can be written as

$$E_0 = f(\nu') - \gamma\left(\frac{\rho' - 6\nu_1}{4}\right)\frac{df}{d\nu'} + \frac{\gamma^2}{2!}\left(\frac{\rho' - 6\nu_1}{4}\right)^2\frac{d^2f}{d\nu'^2} + \cdots + \frac{\gamma}{8}(\rho' - 6\nu_1)\frac{df}{d\nu'} - \frac{3\nu_1\gamma\rho}{8a^2}, \quad (3.37)$$

and the conjecture, stated implicitly above, says that the most divergent term in nth-order perturbation (when $0 < \nu_1 \leq a^2 A$) is

$$\frac{(-\gamma)^n}{n!}\left(\frac{\rho' - 6\nu_1}{4}\right)^n\frac{d^nf}{d\nu'^n}. \quad (3.38)$$

If this is so, then [using (3.31)] (3.37) becomes

$$E_0 = f(\nu' - \tfrac{1}{4}\gamma[\rho' - 6\nu_1]) - (\gamma\rho'^2/16a^2). \quad (3.39)$$

Now since $2\nu' = 2\nu - 3\nu_1\gamma$ in the critical region [Eqs. (3.10) and (3.15)]

$$\nu' - \frac{\gamma(\rho' - 6\nu_1)}{4} = \nu - \frac{\gamma\rho'}{4} \quad (3.40)$$

and since, to order $\gamma^{1/a}$, ρ' can be replaced by ρ, (3.39) is identical with (3.36) and, therefore, with (3.14).

To complete the critical-region analysis for the present model, we investigate the expansion (3.28) for E_0 below the modified critical point, i.e., for $\nu_1 > a^2 A$. Using the results (3.31) to (3.33), we can write (3.28) as

$$E_0 = f(\nu') + \left(\frac{\gamma\rho'}{2}\right)\frac{df}{d\nu'} + \frac{1}{2!}\left(\frac{\gamma\rho'}{2}\right)^2\frac{d^2f}{d\nu'^2} + \cdots - \frac{1}{2}\left(\frac{\gamma\rho'}{2}\right)\frac{df}{d\nu'} - \frac{3\nu_1'\gamma}{4a^2}. \quad (3.41)$$

Assuming that the most divergent term in nth order is

$$\frac{1}{n!}\left(\frac{\gamma\rho'}{2}\right)^n\frac{d^nf}{d\nu'^n}, \quad \text{for} \quad n \geq 3, \quad (3.42)$$

Eq. (3.41) becomes

$$E_0 = f\left(\nu' + \frac{\gamma\rho'}{2}\right) - \frac{3\nu_1^2\gamma}{4a^2} + \frac{\gamma\rho'^2}{8a^2}, \quad (3.43)$$

which reduces to

$$E_0 = \tfrac{1}{2}\sum_{s=1}^{m}[\omega_s(a)]^{\frac{1}{2}} + \gamma\left[\nu_1 A - \frac{(aA)^2}{2} - \frac{3\nu_1^2}{4a^2}\right] + \frac{B\gamma^{1/a}}{2}(2\nu_1 - 2a^2A)^{1/a} + \cdots, \quad \text{for } \nu_1 > a^2A. \quad (3.44)$$

In summary, if one accepts the conjectures regarding the form of the most divergent term in nth-order perturbation ($n \geq 3$), one finds on resumming that the classical critical point (at $2\nu = a$) is displaced to the point $2\nu_c = a + a^2A\gamma$ and that for $\nu < \nu_c$, E_0 is given by (3.14) to order $\gamma^{1/a}$, and for $\nu > \nu_c$ by (3.44). From these expansions and Eqs. (2.13) and (2.16) for the free energy in terms of E_0, we see that the specific heat behaves like $|\nu - \nu_c|^{1/a-2}$ on both sides of the modified critical point, with a superimposed (Curie–Weiss type) jump discontinuity. In other words, if $\frac{1}{2} < a < 1$, the specific heat diverges like $|\nu - \nu_c|^{1/a-2}$, and if $0 < a < \frac{1}{2}$, the singularity remains basically of Curie–Weiss type. This behavior is almost surely not the exact critical behavior; especially, as we see in the following section, since the same technique predicts a logarithmically divergent energy for a two-dimensional model, which is certainly wrong. We return to this point in the final section.

4. CRITICAL REGION OF MODEL A

The model considered in the previous section is special in the sense that it is one-dimensional with an infinite-range interaction. Two- and higher-dimensional systems, on the other hand, exhibit phase transitions already with finite-range interactions, which is in a sense more realistic. In this section, we consider a particular two-dimensional lattice model (model A, which was first introduced in Ref. 7) with M rows and N columns and with interaction energy

$$E = -\sum_{\substack{1\leq k<k'\leq N \\ 1\leq l<l'\leq M}} v(kl, k'l')\mu_{kl}\mu_{k'l'}, \quad (4.1)$$

where

$$v(kl, k'l') = J\gamma e^{-\gamma|k-k'|}\{\delta_{l',l} + \tfrac{1}{2}[\delta_{l',l+1} + \delta_{l',l-1}]\}, \quad J > 0, \quad (4.2)$$

i.e., the spins interact with each other in the same row and in the two neighboring rows according to the exponential of the row distance between spins. Griffiths has proved recently[9] that this model has a phase transition (in fact, he proves that any two- or higher-dimensional lattice model undergoes a transition to an ordered state, provided that the interaction is ferromagnetic and that at least nearest-neighbor spins in a plane interact). Our concern here is the behavior of the model in the critical region.

Because of the exponential nature of the interaction potential (4.2), the technique described in Sec. 2 can be used to reduce the evaluation of the partition function $Q_{N,M}$ to the problem of finding the largest eigenvalue of an M-dimensional integral equation. Details of the reduction are given in Refs. 7 and 8, and

the final result is that the free energy ψ is given by

$$-\frac{\psi}{kT} = \lim_{M\to\infty}\lim_{N\to\infty}\frac{1}{MN}\log Q_{N,M}$$
$$= \log 2 - \frac{\nu\gamma}{2} + \lim_{M\to\infty}\frac{1}{M}\log\Lambda_{\max}, \quad (4.3)$$

where

$$\Lambda_{\max} = \exp\,[M\gamma/2 - \gamma E_0], \quad (4.4)$$

and to order γ^2, E_0 is the smallest eigenvalue of the differential equation

$$\sum_{k=1}^{M}\frac{\partial^2\psi}{\partial x_k^2} - \left\{\tfrac{1}{4}\sum_{k=1}^{M}x_k^2 - \gamma^{-1} \times \sum_{k=1}^{M}\log\cosh\left(\frac{\nu\gamma}{2}\right)^{\frac{1}{2}}(x_k + x_{k+1})\right\}\psi = -E\psi. \quad (4.5)$$

At high temperatures the logcosh function in (4.5) can be expanded, and to order γ, (4.5) becomes

$$\sum_{k=1}^{M}\frac{\partial^2\psi}{\partial x_k^2} - \left\{\tfrac{1}{4}\sum_{k=1}^{M}x_k^2 - \frac{\nu}{4}\sum_{k=1}^{M}(x_k + x_{k+1})^2 + \frac{\nu^2\gamma}{48}\sum_{k=1}^{M}(x_k + x_{k+1})^4\right\}\psi = -E\psi. \quad (4.6)$$

The quadratic form in braces can be diagonalized by the transformation

$$x_k = \sum_{r=1}^{M} a_{kr}y_{r-1}, \quad (4.7)$$

where

$$\begin{aligned} a_{kr} &= M^{-\frac{1}{2}}, && \text{for } r = 1,\\ &= (2/M)^{\frac{1}{2}}\sin\,[\pi r(k-1)/M], && \text{for } r \text{ even},\\ &= (2/M)^{\frac{1}{2}}\cos\,[\pi(r-1)(k-1)/M], && \text{for } r \text{ odd}; \end{aligned} \quad (4.8)$$

and after making the change of variables (4.7), Eq. (4.6) becomes

$$\sum_{r=0}^{M-1}\frac{\partial^2\phi}{\partial y_r^2} - \left\{\tfrac{1}{4}\sum_{r=0}^{M-1}\omega_{\bar{r}}y_r^2 + \alpha\gamma\sum_{k=1}^{M}\left(\sum_{r=0}^{M-1}b_{kr}y_r\right)^4\right\}\phi = -E\phi, \quad (4.9)$$

where $\bar{r}$ denotes the integral part of $(r+1)/2$ and α, ω_r, and b_{kr} are defined, respectively, by

$$\alpha = \nu^2/48, \quad (4.10)$$

$$\omega_r = 1 - 4\nu + 4\nu\sin^2(\pi r/M), \quad (4.11)$$

and

$$b_{kr} = a_{k,r+1} + a_{k+1,r+1}. \quad (4.12)$$

Note that the ω_r's are all positive provided $4\nu < 1$, so that $\nu = \frac{1}{4}$ in this case represents the classical critical point.

By comparing (4.9) with the corresponding Eq. (2.27) for the one-dimensional model, we can immediately write down the high-temperature expansion for E_0, which to second order [cf. (2.43)] is

$$\frac{E_0}{M} = \frac{1}{2M}\sum_{r=0}^{M-1}\omega_{\bar{r}}^{\frac{1}{2}} + 3\alpha\gamma\rho^2 - 24(\alpha\gamma)^2 \times \{T_4 + 3\rho^2T_2\} + \cdots, \quad (4.13)$$

and in the limit $M\to\infty$, $M^{-1}\sum_{r=0}^{M-1}\omega_{\bar{r}}^{\frac{1}{2}}$, ρ, and T_2, etc., are given by

$$\lim_{M\to\infty}\frac{1}{M}\sum_{r=0}^{M-1}\omega_{\bar{r}}^{\frac{1}{2}} = \frac{1}{\pi}\int_0^{\pi/2}(1 - 4\nu + 4\nu\sin^2\theta)^{\frac{1}{2}}\,d\theta,$$
$$\rho = \lim_{M\to\infty}\sum_{r=0}^{M-1}\frac{b_{k,r}^2}{\omega_{\bar{r}}^{\frac{1}{2}}} = \frac{8}{\pi}\int_0^{\pi/2}\cos^2\theta\{1 - 4\nu + 4\nu\sin^2\theta\}^{\frac{1}{2}}\,d\theta, \quad (4.14)$$

and

$$T_2 = \lim_{M\to\infty}M^{-1}\sum_{p=1}^{M}\sum_{q=1}^{M}\int_0^{\infty}dt\left(\sum_{r=0}^{M-1}\frac{b_{p,r}b_{q,r}\exp\,(-\omega_{\bar{r}}^{\frac{1}{2}}t)}{\omega_{\bar{r}}}\right)^2 = \frac{16}{\pi}\int_0^{\pi/2}\cos^4\theta\{1 - 4\nu + 4\nu\sin^2\theta\}^{-\frac{3}{2}}\,d\theta,$$

etc.

The integrals (4.14) can be expressed in terms of elementary elliptic integrals, and when $1 - 4\nu > 0$ is small, they have the expansions

$$f(\nu) = \lim_{M\to\infty}\frac{1}{2M}\sum_{r=0}^{M}\omega_{\bar{r}}^{\frac{1}{2}} = \frac{1}{\pi} + \frac{1}{4\pi}(1 - 4\nu)(\log x - 1) + \cdots,$$
$$\rho = -4\frac{df}{d\nu} = -\frac{8}{\pi}(1 - \tfrac{1}{2}\log x) + \cdots, \quad (4.15)$$

and

$$T_2 = -4\frac{d^2f}{d\nu^2} = \frac{x}{\pi} + \cdots,$$

where x is defined by

$$x = 16/(1 - 4\nu). \quad (4.16)$$

The critical region, from (4.15) and by analogy with Eq. (3.5), is obtained for that range of ν's for which

$$(1 - 4\nu)\log x \sim \gamma(\log x)^2. \quad (4.17)$$

In this region, again by analogy with the analysis of the previous section [(3.10), in particular], we therefore define a new temperature variable ν_1 by

$$4\nu - 1 = \tfrac{1}{4}(\nu_1\gamma)\log x, \quad (4.18)$$

and to obtain the high-temperature critical-region expansions for E_0 we substitute the expansions (4.15) into (4.13).

The calculation is simplified somewhat by using the relations (4.15) between ρ, T_2, and the derivatives of f [which hold exactly for all $4\nu < 1$ and are essentially equivalent to Eqs. (3.31) and (3.32) for the one-dimensional model]. Thus, if we make the same conjecture (3.35) (with $\gamma\rho/32$ instead of $\gamma\rho/4$) regarding the most divergent terms, we have, using (4.15) [cf. (3.36)],

$$\lim_{M\to\infty}\frac{E_0}{M}=\epsilon_0=f\left(\nu-\frac{\gamma\rho}{32}\right)+\frac{1}{2}\left(\frac{\gamma\rho}{32}\right)\frac{df}{d\nu}+\cdots$$
$$=\frac{1}{\pi}+\frac{1}{4\pi}\left(1-4\nu+\frac{\gamma\rho}{8}\right)\times\left\{\log\left(\frac{16}{1-4\nu+(\gamma\rho/8)}\right)-1\right\}-\frac{\gamma}{4\pi^2}\left[1-\tfrac{1}{2}\log\left(\frac{16}{1-4\nu}\right)\right]^2+\cdots. \tag{4.19}$$

Substituting (4.18), and to leading order ρ by $(4/\pi)\log x$, (4.19) becomes

$$\epsilon_0=\frac{1}{\pi}+\frac{\gamma(\log\gamma^{-1})^2}{16\pi^2}[1-\pi\nu_1]-\frac{\gamma\log\gamma^{-1}}{16\pi^2}(2-\pi\nu_1)[\log(2-\pi\nu_1)-1]+\cdots,\quad\text{for}\quad\nu_1<0. \tag{4.20}$$

Again (4.20) exists for $\nu_1 < 2/\pi$ and is analytic at the old singularity $\nu_1 = 0$. The modified singularity at $\nu_1 = 2/\pi$ is from (4.15) to leading order $\rho/2$, and is the value obtained rigorously in Ref. 8. The $\gamma(\log 1/\gamma)^2$ in (4.20) also agrees with that obtained in Ref. 8.

As a check on (4.20) we obtain, as in the previous section, low-temperature ($\nu_1 > 0$) expansions in the critical region and compare, in particular, the expansion for $0 < \nu_1 < 2/\pi$ with (4.20). They should, of course, be identical.

We proceed exactly as before, expanding the potential in (4.5) around its minimum and changing to the transformed variables (4.7). In the critical region, as before, we must first shift before applying perturbation theory, and we find, by essentially repeating word for word the argument from (3.16) to (3.27), that ϵ_0 for $\nu_1 > 0$ is given by

$$\epsilon_0=f(\nu')-3\nu_1\rho'\gamma\log x'/64+3\alpha\gamma\rho'^2-72(\alpha\gamma)^2T_2'(\rho'-6\nu_1\log x')^2+\cdots,\quad\text{for}\quad 0<\nu_1\leq 2/\pi, \tag{4.21}$$

and

$$\epsilon_0=f(\nu')-\frac{3(\nu_1\log x')^2\gamma}{64}-\frac{\gamma\rho'^2}{128}-72(\alpha\gamma)^2T_2'(-2\rho')^2+\cdots,\quad\text{for}\quad\nu_1>2/\pi, \tag{4.22}$$

where the primed quantities are obtained from the corresponding unprimed quantities by replacing ν by ν', which, in the critical region, is given to leading order by

$$1-4\nu'=\tfrac{1}{2}(\nu_1\gamma)\log x'. \tag{4.23}$$

Using the derivative relations (4.15) and essentially repeating the argument from (3.37) to (3.39), (4.21) gives

$$\epsilon_0=f\left(\nu'-\frac{\gamma(\rho'-6\nu_1\log x')}{32}\right)-\frac{\gamma(\rho')^2}{256}+\cdots; \tag{4.24}$$

and since, from (4.23),

$$\nu'-\frac{\gamma(\rho'-6\nu_1\log x')}{32}=\nu-\frac{\gamma\rho}{32} \tag{4.25}$$

to leading order (ρ' can be replaced by ρ), (4.24) is identical with (4.19) and, therefore, with (4.20).

Similarly for $\nu_1 > 2/\pi$, repeating the argument (3.41) to (3.44), (4.22) gives

$$\epsilon_0=f\left(\nu'+\frac{\gamma\rho'}{16}\right)-\frac{3(\nu_1\log x')^2\gamma}{64}+\frac{\gamma\rho'^2}{128}+\cdots$$
$$=\frac{1}{\pi}+\frac{\gamma(\log\gamma^{-1})^2}{8\pi^2}[\pi\nu_1-1-\tfrac{3}{8}(\pi\nu_1)^2]-\frac{\gamma\log\gamma^{-1}}{8\pi^2}(\pi\nu_1-2)[\log(\pi\nu_1-2)+1]+\cdots,\quad\text{for}\quad\nu_1>2/\pi. \tag{4.26}$$

It follows from formulas (4.20), (4.26), (4.3), and (4.4) that in the critical region, to order $\gamma^2\log\gamma^{-1}$,

$$-\psi/kT\sim C_\pm|\nu-\nu_c|\log|\nu-\nu_c|+D_\pm, \tag{4.27}$$

where C_+, D_+ and C_-, D_- are constants (depending on γ) appropriate for $\nu < \nu_c$ and $\nu > \nu_c$, respectively, where from (4.18) and (4.22) $\nu_c = \frac{1}{4} + (\gamma\log\gamma^{-1})/8\pi$ is the modified critical point. From (4.27) we see that to order $\gamma^2\log\gamma^{-1}$, the energy diverges logarithmically and the specific heat has a simple pole at ν_c. This is definitely not the true critical behavior of the model, although one might be tempted to conjecture from this result that the specific heat does in fact diverge logarithmically, just as for the two-dimensional nearest-neighbor model.

5. CRITICAL REGION OF A THREE-DIMENSIONAL MODEL

We conclude with a brief discussion of a three-dimensional model, which may be considered as a natural extension of model A.

Consider a three-dimensional lattice of spins with interaction energy

$$E = -\sum_{\substack{1\le k<k'\le N\\ 1\le l<l'\le M\\ 1\le m<m'\le M}} v(klm, k'l'm')\mu_{klm}\mu_{k'l'm'}, \quad (5.1)$$

where

$$\begin{aligned} v(klm, k'l'm') &= J\gamma e^{-\gamma|k-k'|}\{\tfrac{1}{3}\delta_{l'l}(\delta_{m',m-1}+\delta_{m',m+1}) \\ &\quad + \tfrac{1}{3}\delta_{m',m}(\delta_{l',l-1}+\delta_{l',l+1}) + \tfrac{1}{3}(\delta_{l',l+1}\delta_{m',m-1} \\ &\quad + \delta_{l',l-1}\delta_{m',m+1}) + \delta_{l',l}\delta_{m',m}\}. \end{aligned} \quad (5.2)$$

It is easily verified that

$$\langle X(klm)X(k'l'm')\rangle = v(klm, k'l'm'), \quad (5.3)$$

where

$$X(klm) = (J\gamma/\sqrt{3})(U_{lm}(k) + U_{l+1,m}(k) + U_{l,m+1}(k)) \quad (5.4)$$

and $U_{l,m}(k)$ are independent Ornstein–Uhlenbeck processes with zero mean and covariance

$$\langle U_{lm}(k)U_{l'm'}(k')\rangle = \delta_{l',l}\delta_{m',m}e^{-\gamma|k-k'|}. \quad (5.5)$$

So by the standard argument

$$-\psi/kT = \log 2 - \nu\gamma/2 + \lim_{M\to\infty} M^{-2}\log\Lambda_{\max}, \quad (5.6)$$

where

$$\Lambda_{\max} = \exp(M^2\gamma - \gamma E_0), \quad (5.7)$$

and to order γ^2, E_0 is the smallest eigenvalue of the differential equation

$$\sum_{l,m=1}^{M}\frac{\partial^2\psi}{\partial x_{l,m}^2} - \left\{\tfrac{1}{4}\sum_{l,m=1}^{M}x_{l,m}^2 + \gamma^{-1}\sum_{l,m=1}^{M}\log\cosh\left(\frac{\nu\gamma}{3}\right)^{\frac{1}{2}} \times (x_{l,m}+x_{l+1,m}+x_{l,m+1})\right\}\psi = -E\psi. \quad (5.8)$$

At high temperatures, the log cosh can be expanded as before, and it is straightforward, but rather tedious, to show that to second-order perturbation ($\alpha = \nu^2/108$)

$$\begin{aligned} \epsilon_0 &= \lim_{M\to\infty} M^{-2}E_0 \\ &= f(\nu) + 3\alpha\gamma\rho^2 - 24(\alpha\gamma)^2(T_4 + 3\rho^2T_2) + \cdots, \end{aligned} \quad (5.9)$$

where now

$$\begin{aligned} f(\nu) &= \lim_{M\to\infty}(2M)^{-2}\sum_{r,s=0}^{M-1}\tilde{\omega}_{r,s}^{\frac{1}{2}} \\ &= (2\pi)^{-2}\int_0^\pi\int_0^\pi d\theta_1\, d\theta_2 \\ &\quad \times\left\{\left[1 - 2\nu - \frac{4\nu}{3}(\cos\theta_1 + \cos\theta_2 \right.\right. \\ &\quad \left.\left. + \cos(\theta_1 - \theta_2)\right]^{\frac{1}{2}} + \theta_2 \to -\theta_2\right\}, \end{aligned} \quad (5.10)$$

$$\begin{aligned} \tilde{\omega}_{\alpha,\beta} &= \omega_{r,s}, \quad \text{for } (\alpha,\beta) = (2r, 2s) \text{ or } (2r-1, 2s), \\ &= \omega_{r,-s}, \quad \text{for } (\alpha,\beta) = (2r, 2s-1) \\ &\qquad \text{or } (2r-1, 2s-1), \end{aligned} \quad (5.11)$$

$$\omega_{r,s} = 1 - 2\nu - \frac{4\nu}{3}\left[\cos\left(\frac{2\pi r}{M}\right) + \cos\left(\frac{2\pi s}{M}\right) + \cos\frac{2\pi}{M}(r-s)\right], \quad (5.12)$$

and

$$\rho = -6\frac{df}{d\nu}, \quad (5.13)$$

$$T_2 = -9\frac{d^2f}{d\nu^2}, \quad (5.14)$$

etc.

Equation (5.9) is valid for $6\nu < 1$, i.e., $\nu = \frac{1}{6}$ corresponds to the classical critical point, and when $1 - 6\nu > 0$ is small, $f(\nu)$ has the expansion

$$f(\nu) = f(\tfrac{1}{6}) + A(1-6\nu) + B(1-6\nu)^{\frac{3}{2}} + \cdots. \quad (5.15)$$

In the critical region, therefore [where $f(\nu) - f(\frac{1}{6}) \sim 3\alpha\gamma\rho^2$], we define a new temperature variable ν_1 by

$$6\nu = 1 + \nu_1\gamma/9, \quad (5.16)$$

and by repeating the previous argument we conjecture from (5.9) that

$$\epsilon_0 = f(\tfrac{1}{6}) + \gamma(A^2 - \tfrac{1}{9}A\nu_1) + B(2A - \tfrac{1}{9}\nu_1)^{\frac{3}{2}}\gamma^{\frac{3}{2}} + \cdots, \quad \text{for } \nu_1 < 18A. \quad (5.17)$$

The low-temperature analysis can be performed in the same way as above, but the essential result is contained in (5.17); namely, that to order $\gamma^{\frac{3}{2}}$, the critical point is shifted to

$$\nu_c = \tfrac{1}{6} + A\gamma/3, \quad (5.18)$$

and the specific heat diverges like $|\nu - \nu_c|^{-\frac{1}{2}}$ for $\nu < \nu_c$ and $\nu > \nu_c$ in the neighborhood of ν_c. The square-root singularity, as we have remarked, is also predicted by the Brout and related expansions.[2]

6. DISCUSSION

We have studied a one-dimensional model, a two-dimensional model, and a three-dimensional model, all with a long-range exponential interaction, and have developed high- and low-temperature expansions in the reciprocal "range of interaction" γ, with the classical theory as leading term. The expansions break down for a range of temperatures (defining the critical region) around the classical critical point, and in this region a resummation must be effected. We have carried out the perturbation calculations at high

and low temperatures and have conjectured the form of the most divergent term in nth-order perturbation ($n \geq 3$). Summing these "most divergent terms" shifts the classical critical point by a small amount, and one finds that the resummed expansions are analytic at the "old" classical critical point, but are nonanalytic at the modified critical point (the high- and low-temperature critical-region expansions actually match up at the old singularity, which lends some support to our conjectures). The resummed expansions predict critical exponents which are identical with those of Brout's and similar schemes, and are almost surely incorrect. Unlike our expansions, however, the latter are valid only at high temperatures (above the classical critical point), and in the resummed expansions the singularity "sticks" at the classical value.

The resummation described in the previous sections is only the first step in a succession of resummations. One should now investigate the resummed expansions given here and determine a new critical region, obtained when these expansions break down. This then gives us a new critical point and presumably, on resumming once more, new critical exponents. The procedure, of course, gets rapidly out of hand, and in any case it is not at all clear at the moment how the correct exponents are approached by this successive approximation scheme. There is also the added difficulty that in replacing the integral equations by differential equations we have neglected terms of order γ^2, and in making our conjectures about the form of the most divergent terms, we are assuming implicitly that these are the same for the integral equation and the differential equation.

Nevertheless, our results hint at a general resummation method, starting from the classical theories, and it is encouraging that the essential information about the modified critical points and critical exponents can be deduced from the high-temperature expansions alone. However, the inability (at the moment) of our method and more general methods to give information about the true critical exponents remains the most serious difficulty (or drawback) with this type of approach.

ACKNOWLEDGMENTS

One of us (C. J. T.) is grateful to the Australian Commonwealth Government for the award of a Queen Elizabeth Fellowship, and to The Rockefeller University for support and hospitality during 1965–66 and on a short visit early in 1968, during which time most of the work reported here was carried out.

Addendum: After this work was completed it was noticed that in all cases the high-temperature expansions for the models were formally identical with the general high-temperature expansions obtained by Siegert. By a slight extension of Siegert's argument it is possible actually to prove the most divergent term conjecture (3.35) in general, but only for temperatures above the classical critical point. Details will be published subsequently.

APPENDIX

We derive here the expansions (3.5), (3.6), and (3.11) for the one-dimensional model.

Our starting point is the result [Eq. (6.85) of Ref. 8]

$$2f(v) = \lim_{m\to\infty} \sum_{s=1}^{m} (\omega_s^{\frac{1}{2}} - s\Delta\lambda) = \frac{1}{\pi}\int_0^\infty \log\left[1 - \frac{2v}{a} + 2v\xi^a \int_0^{1/\xi} \frac{v^{a-1}}{1+v^2}\,dv\right] d\xi. \tag{A1}$$

For $\delta = a - 2v$ small and $0 < a < 1$, we can write (A1) as

$$2f(v) = 2f(a/2) + A\delta + F(\delta), \tag{A2}$$

where

$$A = \frac{1}{\pi}\int_0^\infty [(1 - a\xi^a g(\xi))/a^2\xi^a g(\xi)]\,d\xi \tag{A3}$$

and

$$F(\delta) = \frac{1}{\pi}\int_0^\infty \log\left[1 + \frac{\delta(1 - a\xi^a g(\xi))}{a^2\xi^a g(\xi)}\right] d\xi - \frac{\delta}{\pi}\int_0^\infty \frac{(1 - a\xi^a g(\xi))}{a^2\xi^a g(\xi)}\,d\xi, \tag{A4}$$

and $g(\xi)$ is defined by

$$g(\xi) = \int_0^{1/\xi} \frac{v^{a-1}}{1+v^2}\,dv. \tag{A5}$$

Making the substitution $\xi = \delta^{1/a}\eta$ in (A4) gives

$$F(\delta) = B\delta^{1/a} + \cdots, \tag{A6}$$

where

$$B = \frac{1}{\pi}\int_0^\infty \{\log[1 + D\xi^{-a}] - D\xi^{-a}\}\,d\xi \tag{A7}$$

and

$$D^{-1} = a^2\int_0^\infty \frac{v^{a-1}}{1+v^2}\,dv = \frac{a^2\pi}{2}\operatorname{cosec}\left(\frac{a\pi}{2}\right). \tag{A8}$$

Now

$$\frac{\partial B}{\partial D} = \frac{1}{\pi}\int_0^\infty \left(\frac{1}{\eta^a + D} - \frac{1}{\eta^a}\right) d\eta = -(D^{a^{-1}-1}/a)\operatorname{cosec}(a^{-1} - 1)\pi, \tag{A9}$$

and it follows that

$$B = D^{1/a} \operatorname{cosec}(\pi/a). \tag{A10}$$

Combining (A2), (A3), (A6), and (A10) gives the expansion (3.5).

From the definition (2.36) of ρ and Eqs. (2.24) and (2.25), it follows that

$$\rho = -(2\nu)^2 \frac{d}{d\nu} \sum_{s=1}^{m} \omega_s^{\frac{1}{2}}, \tag{A11}$$

and the expansion (3.6) for ρ follows from (3.5) and (A11).

To derive the expansion (3.11) for T_2, we use the result [which follows from Eq. (6.70) of Ref. 8]

$$\lim_{m\to\infty} \sum_{s=1}^{m} \frac{\exp(-\omega_s^{\frac{1}{2}} t)}{N_s^2 \omega_s^{\frac{1}{2}}} = \frac{(2\nu)^2}{2} \int_0^1 \frac{x^{a/2} \exp(-x^{\frac{1}{2}} t)\, dx}{x^{\frac{1}{2}} \left\{ \left[1 - 2\nu/a - \nu \log(1-x) + (\nu a/2) x^{a/2} \int_x^\infty (\log|1-\mu|/\mu^{1+a/2})\, d\mu \right]^2 + (\pi\nu)^2 x^a \right\}}. \tag{A12}$$

From the definition (2.42) of T_2, it then follows trivially that (in the limit $m \to \infty$)

$$T_2 = C(a - 2\nu)^{a^{-1}-2} + \cdots, \tag{A13}$$

where

$$C = \left(\frac{a^2}{2}\right)^2 \int_0^\infty d\tau \times \left\{ \int_0^\infty \frac{a^2 \exp(-\tau y^{\frac{1}{2}}) y^{(a-1)/2}\, dy}{[1 + (a^3/4) y^{a/2} E]^2 + (a^2\pi/2)^2 y^a} \right\}^2 \tag{A14}$$

and

$$E = \int_0^\infty \frac{\log|\mu - 1|}{\mu^{1+a/2}}\, d\mu = \frac{2\pi}{a} \cot\left(\frac{a\pi}{2}\right). \tag{A15}$$

Making the substitution $y \to [yG^{-1/2a}]^2$, where

$$G = D^{-2} = [a^2\pi/2 \sin(a\pi/2)]^2, \tag{A16}$$

and performing the integral over τ gives

$$C = a^4 G^{-(1/2a)-1} \int_0^\infty dy \int_0^\infty dz \frac{(yz)^a}{[1 + 2\cos(\pi a/2) y^a + y^{2a}][1 + 2\cos(\pi a/2) z^a + z^{2a}](y+z)}, \tag{A17}$$

and making the further substitution $y \to y$, $z \to \xi y$ (with Jacobian y) reduces (A17) to

$$C = a^4 G^{-(1/2a)-1} \int_0^\infty \frac{\xi^a\, d\xi}{1+\xi} \int_0^\infty \frac{y^{2a}\, dy}{[1 + 2\cos(\pi a/2) y^a + y^{2a}][1 + 2\cos(\pi a/2)(\xi y)^a + (\xi y)^{2a}]}. \tag{A18}$$

Substituting $y = x^{1/a}$ in (A18) and using the result stated on p. 118 of Ref. 13 gives for the y-integral in (A18),

$$-\frac{\pi \operatorname{cosec}(\pi/a)}{a} \cdot \frac{1}{2\sin(\pi a/2)} \times \frac{1+\xi}{\xi} \cdot \frac{2\cos(\pi a/2)}{(e^{i\pi a} - \xi^a)(e^{-i\pi a} - \xi^a)}; \tag{A19}$$

and finally, making the substitution $\xi = x^{1/a}$ gives

$$C = -\frac{2a^2}{\pi} G^{-1/2a} \operatorname{cosec}\left(\frac{\pi}{a}\right) \sin(\pi a) \times \int_0^\infty \frac{dx}{x^2 - 2\cos(\pi a) x + 1} = -2D^{-1/a} a^2 (1-a) \operatorname{cosec}(\pi/a), \tag{A20}$$

and combining (A13), (A20), and (A10) gives (3.11).

[13] E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge University Press, London, 1927).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

On Perturbation Expansions for Real-Time Green's Functions*

VICTOR KORENMAN
Department of Physics and Astronomy
and
Center for Theoretical Physics, University of Maryland, College Park, Maryland

(Received 19 September 1968)

A proof by Craig that the perturbation expansion for the real-time self-energy of a particle in a many-particle system has the same form in any statistical state is shown to be invalid and the stated theorem is shown to be untrue.

In a recent paper,[1] Craig claims to demonstrate that the perturbation expansion for the real-time self-energy (mass operator) of a particle in a many-particle system has the same form independent of the statistical state of the system. In other words, the self-energy can be expanded as a unique power series in the potential whose terms are the irreducible diagrams composed of interacting particle lines, and the nature of the ensemble is reflected only in the single-particle correlation function itself. It is the purpose of the present note to point out that the purported proof is incorrect and that additional ensemble-dependent terms must appear in the expansion.

The nature of the difficulty is easily understood when one realizes that the successive terms in the perturbation expansion contain expressions for higher correlations of the system as functionals of the one-particle correlation function.[2] The "proof" above is then tantamount to stating that higher correlations are a unique functional of one-particle correlations independent of the statistical ensemble which is appropriate to a given situation.[3] But surely an ensemble can be found where the various correlations are initially chosen to be quite different from those implied by the "universal" functional form.[4] As a trivial example, consider an electron system with no interactions. The "proof" would imply the Hartree–Fock relation

$$\langle\psi(r_1t_1)\psi(r_2t_2)\psi^\dagger(r_3t_3)\psi^\dagger(r_4t_4)\rangle = \langle\psi(r_1t_1)\psi^\dagger(r_3t_3)\rangle \times \langle\psi(r_2t_2)\psi^\dagger(r_4t_4)\rangle - \langle\psi(r_1t_1)\psi^\dagger(r_4t_4)\rangle \times \langle\psi(r_2t_2)\psi^\dagger(r_3t_3)\rangle \quad (1)$$

for any ensemble. But let the ensemble be the pure state which is the BCS ground state and the relation above is clearly incorrect.

In general, then, there must be additional terms in the expansion to take account of the relaxation[4] of the given initial condition to the "universal" functional form when, in an interacting system, this relaxation actually occurs. These are presumably the terms found by Fujita[5] which Craig claims to remove. It remains to see just where Craig's proof breaks down.

The difficulty is found after Eq. (3.12) of Ref. 1, where the interacting Green's function G is expanded in a power series whose terms involve noninteracting Green's functions of all orders computed in the relevant ensemble. Craig claims that these can be decomposed into products of single-particle Green's functions in the usual generalization of the Hartree–Fock procedure.[6,7] The example above shows that such a decomposition is not correct for an arbitrary ensemble.

The proof cited for the claim of decomposability is that of Ambegaokar.[7] In brief, Ambegaokar takes the correct equation for the two-particle function[8]

$$G_0^{-1}(1,\bar{1})G_0(\bar{1},2;1',2') = \delta(1,1')G_0(2,2') - \delta(1,2')G_0(2,1') \quad (2)$$

and operates on the left with G_0 to find

$$G_0(1,2;1',2') = G_0(1,1')G_0(2,2') - G_0(1,2')G_0(2,1'). \quad (3)$$

* Work supported in part by the U.S. Office of Naval Research (contract N000 14-67-A-0239-003) and the Advanced Research Projects Agency (contract DAH004-67-C0023).

[1] R. A. Craig, J. Math. Phys. **9**, 605 (1968).

[2] This connection of terms in the expansion with higher-order correlations is pointed out in Appendix B of the author's thesis (Harvard University, 1965), where the perturbation expansion found by Craig is derived in yet another way. See also V. Korenman, Ann. Phys. (N.Y.) **39**, 72 (1966).

[3] That this is implied by Craig's proof can also be seen by following the type of argument used in Ref. 2. The higher-correlation functions are found by functionally differentiating lower-order functions with respect to an appropriately defined external field. It is easily shown that the functional expression for higher correlations in terms of single-particle correlations is determined completely by the functional expression for the self-energy in terms of the single-particle correlations. See in this respect also G. Baym and L. P. Kadanoff, Phys. Rev. **124**, 287 (1961).

[4] Bogoliubov has discussed the functional dependence of higher correlations on single-particle functions in the context of a classical system. See N. N. Bogoliubov, "Problems of a Dynamical Theory in Statistical Physics," translated by E. K. Gora, Geophysical Research Papers No. 70, AFCRL-TR-60-279, 1960.

[5] S. Fujita, J. Math. Phys. **6**, 1877 (1965).

[6] See, for example, P. C. Martin and J. Schwinger, Phys. Rev. **115**, 1342 (1959), Sec. VI.

[7] V. Ambegaokar, *Astrophysics and the Many Body Problem*, K. W. Ford, Ed. (W. A. Benjamin, Inc., New York, 1963), p. 349.

[8] See Ref. 7 for notation.

This equation and its simple generalizations are the results Craig needs. But G_0^{-1} has a vanishing eigenvalue so that the solution to Eq. (2) is only given modulo the addition of an arbitrary amount of the solution of the homogeneous equation $G_0^{-1}(1, \bar{1}) \times f(\bar{1}) = 0$. This added term will serve to correct Eq. (3) for the mismatch between the value of the two-particle correlation function implied by the decomposition and that appropriate to the statistical state of interest.[9] These are just the terms which are neglected in Craig's expansion and which are needed to describe the complete effect of the chosen statistical state on the system properties.

[9] This is not to say that Ambegaokar's proof is not valid in the context in which it was introduced. In thermal equilibrium the boundary conditions on the correlation functions are such that no additional terms can be appended to Eq. (3).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

On Next-Nearest-Neighbor Interaction in Linear Chain. I

CHANCHAL K. MAJUMDAR AND DIPAN K. GHOSH
Tata Institute of Fundamental Research, Bombay, India

(Received 2 September 1967)

Ground-state properties of the Hamiltonian

$$H = \tfrac{1}{2}J\sum_{i=1}^{N}\boldsymbol{\sigma}_i\cdot\boldsymbol{\sigma}_{i+1} + \tfrac{1}{2}J\alpha\sum_{i=1}^{N}\boldsymbol{\sigma}_i\cdot\boldsymbol{\sigma}_{i+2}$$

($\boldsymbol{\sigma}_{N+1}\equiv\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_{N+2}\equiv\boldsymbol{\sigma}_2$) are studied for both signs of J and $-1\leq\alpha\leq 1$ to gain insight into the stability of the ground state with nearest-neighbor interactions only ($\alpha = 0$) in the presence of the next-nearest-neighbor interaction. Short chains of up to 8 particles have been exactly studied. For $J > 0$, the ground state for even N belongs always to spin zero, but its symmetry changes for certain values of α. For $J < 0$, the ground state belongs either to the highest spin (ferromagnetic state) or to the lowest spin and so to zero for even N. The trend of the results suggests that these facts are true for arbitrary N and that the critical value of α is probably zero. Upper and lower bounds to the ground-state energy per spin of the above Hamiltonian are obtained. Such bounds can also be obtained for the square lattice with the nearest- as well as the next-nearest-neighbor interaction.

I. INTRODUCTION

The Heisenberg linear chain with the Hamiltonian

$$H = \tfrac{1}{2}J\sum_{i=1}^{N}\boldsymbol{\sigma}_i\cdot\boldsymbol{\sigma}_{i+1} \qquad (1)$$

($J < 0$ ferromagnetic, $J > 0$ antiferromagnetic) was thoroughly investigated by Bethe[1] and the ground-state energy determined by Bethe and Hulthén.[2] des Cloizeaux and Pearson[3] discussed the low-lying excitation spectrum and Griffiths[4] calculated the magnetization at zero temperature. While the 1-dimensional version is an interesting many-body problem, the general Heisenberg Hamiltonian as a description of magnetic phenomena belongs, to quote Herring,[5] "more to the world of thought." An extensive criticism of the exchange integral and its relevance to magnetic properties of solids is given by Herring.[6]

Two obvious criticisms leveled against (1) or its 3-dimensional analog are the neglect of anisotropy and the restriction to nearest-neighbor interaction only. Taking the isotropic Hamiltonian, Mermin and Wagner[7] showed that there was no spontaneous magnetization in one and two dimensions. In three dimensions, spontaneous magnetization is believed to exist. An attempt to incorporate anisotropy is the study of the Hamiltonian considered in detail by Orbach[8]:

$$H = \tfrac{1}{2}J\sum_{i=1}^{N}(\sigma_i^x\sigma_{i+1}^x + \sigma_i^y\sigma_{i+1}^y + \Delta\sigma_i^z\sigma_{i+1}^z). \qquad (2)$$

Many interesting properties of this Hamiltonian are known from the recent extensive work by Yang and Yang,[9] who give references to earlier works. One might expect that in two dimensions, for sufficiently large Δ, a spontaneous magnetic moment exists, although in the 1-dimensional case, at finite temperatures, it probably does not.

[1] H. Bethe, Z. Physik **71**, 205 (1931).
[2] L. Hulthén, Arkiv Mat. Astron. Fysik **26A**, No. 11 (1938).
[3] J. des Cloizeaux and J. J. Pearson, Phys. Rev. **128**, 2131 (1962).
[4] R. B. Griffiths, Phys. Rev. **133**, A768 (1964).
[5] C. Herring, Rev. Mod. Phys. **34**, 631 (1962).
[6] C. Herring, in *Magnetism*, G. T. Rado and H. Suhl, Eds. (Academic Press Inc., New York, 1966), Vol. IV.
[7] N. D. Mermin and H. Wagner, Phys. Rev. Letters **17**, 1133 (1966).
[8] R. Orbach, Phys. Rev. **112**, 309 (1958); **115**, 1181 (1959).
[9] C. N. Yang and C. P. Yang, Phys. Rev. **147**, 303 (1966); **150**, 321, 327 (1966); **151**, 258 (1966).

This equation and its simple generalizations are the results Craig needs. But G_0^{-1} has a vanishing eigenvalue so that the solution to Eq. (2) is only given modulo the addition of an arbitrary amount of the solution of the homogeneous equation $G_0^{-1}(1, \bar{1}) \times f(\bar{1}) = 0$. This added term will serve to correct Eq. (3) for the mismatch between the value of the two-particle correlation function implied by the decomposition and that appropriate to the statistical state of interest.[9] These are just the terms which are neglected in Craig's expansion and which are needed to describe the complete effect of the chosen statistical state on the system properties.

[9] This is not to say that Ambegaokar's proof is not valid in the context in which it was introduced. In thermal equilibrium the boundary conditions on the correlation functions are such that no additional terms can be appended to Eq. (3).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

On Next-Nearest-Neighbor Interaction in Linear Chain. I

CHANCHAL K. MAJUMDAR AND DIPAN K. GHOSH
Tata Institute of Fundamental Research, Bombay, India

(Received 2 September 1967)

Ground-state properties of the Hamiltonian

$$H = \tfrac{1}{2}J\sum_{i=1}^{N}\boldsymbol{\sigma}_i\cdot\boldsymbol{\sigma}_{i+1} + \tfrac{1}{2}J\alpha\sum_{i=1}^{N}\boldsymbol{\sigma}_i\cdot\boldsymbol{\sigma}_{i+2}$$

($\boldsymbol{\sigma}_{N+1}\equiv\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_{N+2}\equiv\boldsymbol{\sigma}_2$) are studied for both signs of J and $-1\leq\alpha\leq 1$ to gain insight into the stability of the ground state with nearest-neighbor interactions only ($\alpha = 0$) in the presence of the next-nearest-neighbor interaction. Short chains of up to 8 particles have been exactly studied. For $J > 0$, the ground state for even N belongs always to spin zero, but its symmetry changes for certain values of α. For $J < 0$, the ground state belongs either to the highest spin (ferromagnetic state) or to the lowest spin and so to zero for even N. The trend of the results suggests that these facts are true for arbitrary N and that the critical value of α is probably zero. Upper and lower bounds to the ground-state energy per spin of the above Hamiltonian are obtained. Such bounds can also be obtained for the square lattice with the nearest- as well as the next-nearest-neighbor interaction.

I. INTRODUCTION

The Heisenberg linear chain with the Hamiltonian

$$H = \tfrac{1}{2}J\sum_{i=1}^{N}\boldsymbol{\sigma}_i\cdot\boldsymbol{\sigma}_{i+1} \qquad (1)$$

($J < 0$ ferromagnetic, $J > 0$ antiferromagnetic) was thoroughly investigated by Bethe[1] and the ground-state energy determined by Bethe and Hulthén.[2] des Cloizeaux and Pearson[3] discussed the low-lying excitation spectrum and Griffiths[4] calculated the magnetization at zero temperature. While the 1-dimensional version is an interesting many-body problem, the general Heisenberg Hamiltonian as a description of magnetic phenomena belongs, to quote Herring,[5] "more to the world of thought." An extensive criticism of the exchange integral and its relevance to magnetic properties of solids is given by Herring.[6]

Two obvious criticisms leveled against (1) or its 3-dimensional analog are the neglect of anisotropy and the restriction to nearest-neighbor interaction only. Taking the isotropic Hamiltonian, Mermin and Wagner[7] showed that there was no spontaneous magnetization in one and two dimensions. In three dimensions, spontaneous magnetization is believed to exist. An attempt to incorporate anisotropy is the study of the Hamiltonian considered in detail by Orbach[8]:

$$H = \tfrac{1}{2}J\sum_{i=1}^{N}(\sigma_i^x\sigma_{i+1}^x + \sigma_i^y\sigma_{i+1}^y + \Delta\sigma_i^z\sigma_{i+1}^z). \qquad (2)$$

Many interesting properties of this Hamiltonian are known from the recent extensive work by Yang and Yang,[9] who give references to earlier works. One might expect that in two dimensions, for sufficiently large Δ, a spontaneous magnetic moment exists, although in the 1-dimensional case, at finite temperatures, it probably does not.

[1] H. Bethe, Z. Physik **71**, 205 (1931).
[2] L. Hulthén, Arkiv Mat. Astron. Fysik **26A**, No. 11 (1938).
[3] J. des Cloizeaux and J. J. Pearson, Phys. Rev. **128**, 2131 (1962).
[4] R. B. Griffiths, Phys. Rev. **133**, A768 (1964).
[5] C. Herring, Rev. Mod. Phys. **34**, 631 (1962).
[6] C. Herring, in *Magnetism*, G. T. Rado and H. Suhl, Eds. (Academic Press Inc., New York, 1966), Vol. IV.
[7] N. D. Mermin and H. Wagner, Phys. Rev. Letters **17**, 1133 (1966).
[8] R. Orbach, Phys. Rev. **112**, 309 (1958); **115**, 1181 (1959).
[9] C. N. Yang and C. P. Yang, Phys. Rev. **147**, 303 (1966); **150**, 321, 327 (1966); **151**, 258 (1966).

Our purpose is to direct attention to the second of the two restrictions—only nearest-neighbor interactions. We propose to study the Hamiltonian

$$H = \tfrac{1}{2}J\sum_{i=1}^{N}\boldsymbol{\sigma}_i\cdot\boldsymbol{\sigma}_{i+1} + \tfrac{1}{2}J\alpha\sum_{i=1}^{N}\boldsymbol{\sigma}_i\cdot\boldsymbol{\sigma}_{i+2} \qquad (3)$$

($N+1\equiv 1$, $N+2\equiv 2$), where the next-nearest neighbors interact with a strength αJ, $-1\leq\alpha\leq 1$. In general, the next-nearest-neighbor interaction will be smaller than the nearest-neighbor interaction. However, their ratio may not be completely negligible.[10] An exact solution of this Hamiltonian, not to speak of its 3-dimensional variant, is probably very difficult. We shall, however, present lower and upper bounds on the ground-state energy per spin in one dimension for any J and for α in $[-1, 1]$. To get an idea of the possibilities involved we have studied short chains of 4, 6, 8 and 3, 5, 7 spins and we shall present the data with the corresponding conjectures about large N.

The question that we hope to answer by studying (3) in this way is the following: How stable are the ground states of (1) with regard to the presence of the next-nearest-neighbor interaction? In particular, if the "classical" ground states become unstable, what is the nature of instability and what is the new ground state? Another purpose is to examine the mathematical difficulties one might encounter in the extension to a realistic situation of forces of intermediate range.

The sign of the exchange integral is a difficult question and without any prior prejudice we should allow both possibilities for next-nearest-neighbor interactions. We thus have the following four different cases:

(i) $J<0$, $\alpha>0$: all the interactions are ferromagnetic, and the classical ground state of aligned spins is expected to be stable;

(ii) $J>0$, $\alpha<0$: the nearest neighbors interact antiferromagnetically, while the next-nearest neighbors tend to align themselves—again a stabilization of the classically accepted picture;

(iii) $J<0$, $\alpha<0$: the next-nearest-neighbors have a tendency to align in opposite direction and the ferromagnetic alignment is likely to be destroyed for large $|\alpha|$ (An interesting question here is: what is this critical α?);

(iv) $J>0$, $\alpha>0$: the alignment of the next-nearest neighbors is opposing that of the nearest-neighbor interaction and the ground state, although of spin zero, may have alignments different from that of the classical ground state, that is, different symmetry.

[10] J. S. Smart, in *Magnetism*, G. T. Rado and H. Suhl, Eds. (Academic Press Inc., New York, 1963), Vol. III.

The study of short linear chains corroborates the expectations. In cases (i) and (ii), nothing untoward happens. In the presence of strong antiferromagnetic next-nearest-neighbor interaction, the ferromagnetic ground state becomes unstable and the lowest state has spin zero (N even). When all interactions are antiferromagnetic, the ground state definitely has spin zero, but the symmetry of the ground state changes for strong antiferromagnetic next-nearest-neighbor interaction.

Concerned with this last point we have a theorem due to von Neumann and Wigner[11] about Hamiltonians such as (3) that depend on a single parameter. The theorem forbids crossing of two levels of identical symmetry. By symmetry all possible symmetries are to be included. The levels that cross in the case of all interactions antiferromagnetic, differ by a kind of permutation symmetry, which we shall describe in detail for the short chains. We must also recall a theorem of Lieb and Mattis[12] for the ground state when $J>0$. According to this theorem, the ground state always has spin zero. This we have found to be true.

II. SHORT LINEAR CHAINS. NOTATION

Let us start with a few generalities and notations. We note that the total spin S^2 as well as its z component S_z are constants of motion. Hence, to get a complete picture of the eigenvalues, it will be enough to investigate the $S_z=0$ subspace (N even) since every state can be rotated to this subspace without change in energy. The number of states with $S_z=0$ for N spin-$\frac{1}{2}$ particles is given by $\binom{N}{\frac{1}{2}N}$. It is also desirable to know the number of states with total spin $S=0$, which is

$$\binom{N}{\frac{1}{2}N}\Big/(\tfrac{1}{2}N+1).$$

This is obtained by considering the difference in the total number of states in the $S_z=0$ subspace and in the $S_z=1$ subspace, a procedure with obvious generalization.[13] It is possible to derive it in a somewhat circuitous way, which, however, has certain advantages in that this new method gives also the structure of $S=0$ states. Consider the problem group-theoretically. Each particle of spin $\frac{1}{2}$ is associated with the basic representation of the SU_2 group. For N particles, the possible spins can be obtained by constructing direct products of the basic representations of the SU_2 group, the spin $\frac{1}{2}$ representation.

[11] E. P. Wigner and J. von Neumann, Physik. Z. **30**, 467 (1929).
[12] E. H. Lieb and D. C. Mattis, J. Math. Phys. **3**, 749 (1962).
[13] F. Bloch, Z. Physik **59**, 208 (1930).

The various representations are characterized by Young's tableaux, with at most two rows. It is well known that, because of the nature of the special unitary group SU_2, only one number may be used to characterize the representation. However, the tableaux with two rows are also certain special representations of the symmetric or permutation group of N particles. The dimensionality of the representation is obviously the possible number of linearly independent states with the spin characteristic of the tableau. This dimensionality is nothing but the character of the unit element in that representation and can also be obtained as the number of ways of filling up the tableaux in the standard order.[14] All this simply depends on the correspondence of the representations of the permutation groups and the general linear group discussed extensively by Weyl.[15] Now the two-rowed tableaux with numerical characters filling them obviously suggest a way of writing down the spin $S = 0$ states. The number of total $S = 0$ states are fewer than the number of $S_z = 0$ states and the states in $S = 0$ subspace may be studied for slightly longer chains than those possible for $S_z = 0$ subspace. Hulthén,[2] who first studied short linear chains, used the construction suggested by the above procedure, but found it convenient not to use the states of the tableaux filled in the standard fashion. Rather he used a set of states which had the same structure, but took better advantage of the cyclic nature of the Hamiltonian with only nearest-neighbor interaction. We shall use the same set of states used by Hulthén, but we shall find that the presence of next-nearest-neighbor interaction introduces certain difficulties in the computations as soon as the chain becomes moderately long.

We shall follow Hulthén's article[2] closely in notation. Let α and β be the up- and down-spin states and σ_k^x, σ_k^y, σ_k^z be the usual Pauli spinors for the particle numbered k. Now introduce

$$[l, m] = \alpha(l)\beta(m) - \beta(l)\alpha(m),$$
$$\{l, m\} = \alpha(l)\beta(m) + \beta(l)\alpha(m). \tag{4}$$

In general, a symmetric function in particles $k, l, m, \cdots$ is $\{k, l, m \cdots\}$. The basic functions of (4) fulfil certain algebraic relations:

$$[k, l]\{m, n\} + [l, m]\{n, k\} + [m, n]\{k, l\} + [n, k]\{l, m\} = 0,$$
$$[k, l]\,[m, n] + [k, n]\,[l, m] + [k, m]\,[n, l] = 0. \tag{5}$$

Hulthén's basis functions are constructed out of such units as (4). Hence it is important to know the following properties for calculating eigenvalues of the Hamiltonian:

$$\tfrac{1}{2}(1 - \boldsymbol{\sigma}_l \cdot \boldsymbol{\sigma}_m)[l, m] = 2[l, m],$$
$$\tfrac{1}{2}(1 - \boldsymbol{\sigma}_l \cdot \boldsymbol{\sigma}_m)\{l, m, k, \cdots\} = 0,$$
$$\tfrac{1}{2}(1 - \boldsymbol{\sigma}_l \cdot \boldsymbol{\sigma}_m)[k, l]\{m, n\} = -[l, m]\{k, n\},$$
$$\tfrac{1}{2}(1 - \boldsymbol{\sigma}_l \cdot \boldsymbol{\sigma}_m)[k, l]\,[m, n] = [l, m]\,[n, k]. \tag{6}$$

III. EVEN NUMBER OF SPINS

A. 4 Spins

We shall start by considering 4 spins. This is a somewhat degenerate case since the next-nearest-neighbor interactions are not fully developed. Nevertheless, certain interesting features are present. The Hamiltonian is

$$H = \tfrac{1}{2}J \sum_{i=1}^{4} \boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_{i+1} + \tfrac{1}{2}J\alpha \sum_{i=1}^{2} \boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_{i+2}. \tag{7}$$

Here the particle 5 is equivalent to particle numbered 1. Define now

$$H' = -\frac{H - \alpha J - 2J}{J}$$
$$= \sum_{i=1}^{4} \tfrac{1}{2}(1 - \boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_{i+1}) + \alpha \sum_{i=1}^{2} \tfrac{1}{2}(1 - \boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_{i+2}). \tag{8}$$

There are 6 states with $S_z = 0$, and are of the forms $\alpha(1)\,\beta(2)\,\alpha(3)\,\beta(4)$, $\alpha(1)\,\alpha(2)\,\beta(3)\,\beta(4)$, etc. The 6×6 matrix of the Hamiltonian can easily be constructed. The diagonalization problem is trivial and can be done by inspection. We shall rather follow Hulthén's construction procedure described above, so as to illustrate our remarks in Sec. II.

With 4 spin-$\frac{1}{2}$ particles we have three tableaux to consider corresponding to partitions [4], [31], [22]. [4] corresponds to only completely symmetric states of spin 2, $\psi_6 = \{1\ 2\ 3\ 4\}$ and using (6) one has

$$H'\psi_6 = H'\{1\ 2\ 3\ 4\} = 0. \tag{9}$$

The tableau [31] can be filled in the standard order in 3 different ways:

$$\begin{array}{cccc} 1 & 3 & 4 & \\ 2 & & & \end{array}, \quad \begin{array}{cccc} 1 & 2 & 4 & \\ 3 & & & \end{array}, \quad \text{and} \quad \begin{array}{cccc} 1 & 2 & 3 & \\ 4 & & & \end{array}.$$

These have spin 1. As Hulthén remarked, it is convenient to choose a function suggested by the first labeling and then use the nearest-neighbor interaction part of the Hamiltonian (7) to generate the other states. Hence we take as the basis

$$\psi_3 = [12]\{34\},$$
$$\psi_4 = [23]\{41\}, \tag{10}$$
$$\psi_5 = [34]\{12\}.$$

[14] D. Littlewood, *The Theory of Group Characters* (Oxford University Press, London, 1940).
[15] H. Weyl, *Classical Groups* (Princeton University Press, Princeton, N.J., 1946).

It is especially important to notice that the basis functions are not normalized and are not necessarily orthogonal. Using (6) to operate with H' and using (5), we get

$$H'\psi_3 = 3\psi_3 + \psi_5 + \alpha(\psi_3 - \psi_5),$$
$$H'\psi_4 = -\psi_3 + 2\psi_4 - \psi_5 + \alpha(\psi_3 + 2\psi_4 + \psi_5), \quad (11)$$
$$H'\psi_5 = \psi_3 + 3\psi_5 - \alpha(\psi_3 - \psi_5).$$

Hence, we obtain the eigenfunctions

$$H'(\psi_3 + \psi_5) = 4(\psi_3 + \psi_5),$$
$$H'(\psi_3 + \psi_4) = (2 + 2\alpha)(\psi_3 + \psi_4), \quad (12)$$
$$H'(\psi_4 + \psi_5) = (2 + 2\alpha)(\psi_4 + \psi_5).$$

We have also the scalar product $(\psi_3, \psi_3) = (\psi_4, \psi_4) = (\psi_5, \psi_5) = 4, (\psi_3, \psi_4) = (\psi_4, \psi_5) = -2,$ and $(\psi_3, \psi_5) = 0$. Hence, the eigenfunctions $(\psi_3 + \psi_5)$, $(\psi_3 + \psi_4)$, and $(\psi_4 + \psi_5)$ are mutually orthogonal.

The tableau [22] can be filled in two ways:

$$\begin{matrix} 1 & 3 \\ 2 & 4 \end{matrix} \quad \text{and} \quad \begin{matrix} 1 & 2 \\ 3 & 4 \end{matrix}.$$

They have spin zero. Again we choose Hulthén's basis functions, cyclically generated:

$$\varphi_1 = [12]\,[34], \quad \varphi_2 = [23]\,[41]. \quad (13)$$

Then, by (6) and (5),

$$H'\varphi_1 = 4\varphi_1 + 2\varphi_2 + 2\alpha[13]\,[24],$$
$$H'\varphi_2 = 2\varphi_1 + 4\varphi_2 - 2\alpha[13]\,[24], \quad (14)$$

so that

$$H'(\varphi_1 + \varphi_2) = 6(\varphi_1 + \varphi_2),$$
$$H'(\varphi_1 - \varphi_2) = (2 + 4\alpha)(\varphi_1 - \varphi_2). \quad (15)$$

In deriving the last equation we have used (5). We have obtained all the eigenvalues which are given in Fig. 1. It must be pointed out that this is not a "spectrum of the chain" of 4 spin-$\frac{1}{2}$ particles. As des Cloizeaux and Pearson[3] have explained, the spectrum of the chain necessarily implies a consideration of the wave vector k. (Identical remarks hold for Figs. 3 and 5.)

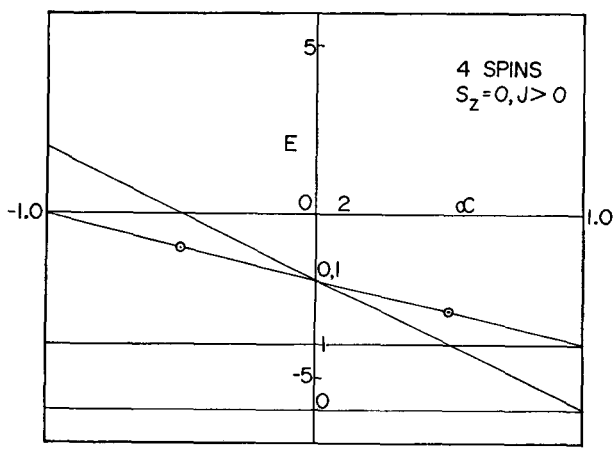

FIG. 1. Eigenvalues of $-H'$ for 4 spins as functions of α. The small circle denotes a doubly degenerate level. Spins of the levels are also indicated.

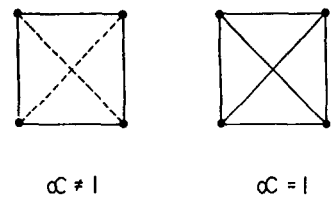

FIG. 2. Symmetries of the 4-spin complex. ($\alpha = 1$) is a complete graph and represents a tetrahedron.

Figure 1 has been drawn for the $S_z = 0$ subspace, with $J > 0$, antiferromagnetic case. For ferromagnetic case ($J < 0$) the picture should be inverted (same for Figs. 3 and 5). For antiferromagnetic case the ground state always has spin zero. For ferromagnetic case, however, the ground state changes from one having spin 2 to one having spin zero at $\alpha = -\frac{1}{2}$.

There are certain interesting multiplet structures and level crossings in the diagram, which may be related to the symmetries of the problem. Since we have already taken full rotational symmetry into account by going over to $S_z = 0$ subspace, we must look for the explanation of the multiplet structure to certain other invariances. In particular, the structure at $\alpha = 1$, where all states of the same spin come together, is really interesting. To investigate the symmetries, it is convenient to have geometrical pictures, which make the permutation symmetries inherent in (7) rather intuitive. So let us place the 4 spins at the corners of a square and imagine that the sides represent bonds of strength J connecting nearest neighbors. Join the diagonals with dotted lines to represent the next-nearest-neighbor bonds of strength αJ. When $\alpha = 1$, the dotted lines are replaced by full lines. In the language of graph theory we have now the "square of the original graph."[16] If we call two vertices connected by a nearest-neighbor bond to be at a "distance" 1, the next-nearest-neighbor interaction inserts a bond between vertices at distance 2. For the square diagram of 4 spins, the square of the graph is a "complete" graph,[17] a somewhat degenerate situation from the physical stand point. To obtain the symmetries of our Hamiltonian, consider the automorphisms of the graphs in Fig. 2, which leave the connections invariant. For $\alpha = 1$, the

[16] A. Mukhopadhyaya, J. Combinatorial Theory **2**, 290 (1967).

[17] G. Uhlenbeck and G. W. Ford, in *Statistical Mechanics*, J. de Boer and G. E. Uhlenbeck, Eds. (North-Holland Publishing Co., Amsterdam, 1962), Vol. I, p. 119.

graph is complete and the automorphisms constitute the full symmetric group S_4. The representations of the group are well known[14] and the degeneracies of the states at $\alpha = 1$ are immediately explained. In fact, the very construction of our states φ_1, φ_2, ψ_3, ψ_4, ψ_5, and ψ_6 make their transformation properties obvious. When $\alpha \neq 1$, the group of automorphisms is clearly the group of the square. It is a subgroup of S_4 with 8 elements and is, in fact, the dihedral group D_4. The irreducible representations of D_4 fall into five classes A_1, A_2, B_1, B_2, and E with dimensions 1, 1, 1, 1, 2 (see the Appendix for the character table). It is easy to verify that the doubly degenerate states $(\psi_3 + \psi_4)$ and $(\psi_4 + \psi_5)$ form the representation of dimension 2. The ground state $(\varphi_1 + \varphi_2)$ belongs to the representation A_1, as also the completely symmetric state ψ_6. The state $(\varphi_1 - \varphi_2)$ belongs to the representation B_1. The state $(\psi_3 + \psi_5)$ transforms as B_2. The triple degeneracy at $\alpha = 0$ of the eigenvalues is accidental.

B. 6 Spins

Henceforth we shall concentrate on the $S_z = 0$, $S = 0$ subspace. For 6 spin-$\frac{1}{2}$ particles there are five $S = 0$ states, which we take as

$$\begin{aligned} \varphi_1 &= [12]\,[34]\,[56], \quad \varphi_2 = [23]\,[45]\,[61], \\ \psi_3 &= [23]\,[41]\,[56], \quad \psi_4 = [12]\,[45]\,[63], \\ \psi_5 &= [34]\,[61]\,[25]. \end{aligned} \tag{16}$$

The Hamiltonian is written as

$$\begin{aligned} H' &= -\frac{H - 3J - 3\alpha J}{J} \\ &= \tfrac{1}{2}\sum_{i=1}^{6}(1 - \boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_{i+1}) + \tfrac{1}{2}\alpha\sum_{i=1}^{6}(1 - \boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_{i+2}), \end{aligned} \tag{17}$$

with usual identification $\boldsymbol{\sigma}_7 \equiv \boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_8 \equiv \boldsymbol{\sigma}_2$. With (6) and (5) we get

$$\begin{aligned} H'\varphi_1 &= (6 + 6\alpha)\varphi_1 + (1 - 2\alpha)(\psi_3 + \psi_4 + \psi_5), \\ H'\varphi_2 &= (6 + 6\alpha)\varphi_2 - (1 - 2\alpha)(\psi_3 + \psi_4 + \psi_5), \\ H'\psi_3 &= 4\psi_3 + 2\varphi_1 - 2\varphi_2 \\ &\quad + 2\alpha\{-\varphi_1 + \varphi_2 + 3\psi_3 - [25]\,[14]\,[63]\}, \\ H'\psi_4 &= 4\psi_4 + 2\varphi_1 - 2\varphi_2 \\ &\quad + 2\alpha\{-\varphi_1 + \varphi_2 + 3\psi_4 - [25]\,[14]\,[63]\}, \\ H'\psi_5 &= 4\psi_5 + 2\varphi_1 - 2\varphi_2 \\ &\quad + 2\alpha\{-\varphi_1 + \varphi_2 + 3\psi_5 - [25]\,[14]\,[63]\}. \end{aligned} \tag{18}$$

From (18), after some obvious manipulations, we get

$$\begin{aligned} H'(\varphi_1 + \varphi_2) &= (6 + 6\alpha)(\varphi_1 + \varphi_2), \\ H'(\varphi_1 - \varphi_2) &= (6 + 6\alpha)(\varphi_1 - \varphi_2) \\ &\quad + 2(1 - 2\alpha)(\psi_3 + \psi_4 + \psi_5), \\ H'(\psi_3 - \psi_4) &= (4 + 6\alpha)(\psi_3 - \psi_4), \\ H'(\psi_3 - \psi_5) &= (4 + 6\alpha)(\psi_3 - \psi_5), \\ H'(\psi_3 + \psi_4 + \psi_5) &= (4 + 6\alpha)(\psi_3 + \psi_4 + \psi_5) \\ &\quad + (6 - 6\alpha)(\varphi_1 - \varphi_2) \\ &\quad - 6\alpha[25]\,[14]\,[63]. \end{aligned} \tag{19}$$

We have found 3 eigenvalues $(6 + 6\alpha)$, $(4 + 6\alpha)$, and $(4 + 6\alpha)$. The remaining 2 must be found by studying the secular equation connecting $(\varphi_1 - \varphi_2)$ and $(\psi_3 + \psi_4 + \psi_5)$. The first difficulty of using the next-nearest-neighbor interaction is apparent in the existence of the state [25] [14] [63]. With $\alpha = 0$, the five linearly independent states of the Hulthén basis (16) were the only ones regenerated by the operation of the Hamiltonian. However, now another state [25] [14] [63] appears which has to be re-expressed in terms of the basis (16). The basis functions (16) were neither normalized nor mutually orthogonal in general. Hence we write

$$\begin{aligned} \psi &\equiv [25]\,[14]\,[63] \\ &= c_1\varphi_1 + c_2\varphi_2 + c_3\psi_3 + c_4\psi_4 + c_5\psi_5, \end{aligned} \tag{20}$$

and consider the set of linear equations for the c's:

$$\begin{aligned} (\varphi_1, \psi) = c_1(\varphi_1, \varphi_1) &+ c_2(\varphi_1, \varphi_2) + c_3(\varphi_1, \psi_3) \\ &+ c_4(\varphi_1, \psi_4) + c_5(\varphi_1, \psi_5), \end{aligned} \tag{21}$$

etc. The matrix of the scalar product is easy to construct:

$$\begin{aligned} &(\varphi_i, \varphi_i) = (\psi_i, \psi_i) = 8, \\ &(\varphi_1, \varphi_2) = -2, \\ &(\varphi_1, \psi_i) = -(\varphi_2, \psi_i) = 4, \\ &(\psi_i, \psi_j) = 2 \quad (i \neq j), \\ &(\varphi_1, \psi) = -(\varphi_2, \psi) = 2, \quad (\psi_i, \psi) = 4. \end{aligned}$$

Solving (21), we obtain

$$\psi = -(\varphi_1 - \varphi_2) + (\psi_3 + \psi_4 + \psi_5), \tag{22}$$

so that

$$\begin{aligned} H'(\varphi_1 - \varphi_2) &= (6 + 6\alpha)(\varphi_1 - \varphi_2) \\ &\quad + 2(1 - 2\alpha)(\psi_3 + \psi_4 + \psi_5), \\ H'(\psi_3 + \psi_4 + \psi_5) &= 6(\varphi_1 - \varphi_2) + 4(\psi_3 + \psi_4 + \psi_5), \end{aligned} \tag{23}$$

and the secular equation is

$$\begin{vmatrix} 6+6\alpha-\lambda & 2(1-2\alpha) \\ 6 & 4-\lambda \end{vmatrix} = 0, \qquad (24)$$

with the roots

$$\lambda = (5+3\alpha) \pm (13-18\alpha+9\alpha^2)^{\frac{1}{2}}. \qquad (25)$$

We have therefore determined all the eigenvalues of $S=0$ levels. The eigenvalue of the spin $S=3$ levels is of course 0. Hence a spin-zero state crosses the ferromagnetic ground state when $\alpha = -0.25$. For the antiferromagnetic case a new possibility shows itself at $\alpha = \frac{1}{2}$. A spin-zero state $(\varphi_1 + \varphi_2)$, which was higher than the ground state at $\alpha = 0$, crosses the ground state and becomes the new ground state for $\alpha > \frac{1}{2}$. By the Wigner–von Neumann theorem, the two states must differ in symmetry, which we shall examine below.

We have also determined the exact eigenvalues for all states in the $S_z = 0$ subspace. There are 20 $S_z = 0$ states, and, rather than proceeding analytically as above, the matrix of the Hamiltonian can be easily diagonalized on a CDC 3600 computer. Figure 3 presents the eigenvalues of the 20 states. Their spins can be determined by comparing our eigenvalues with those given by Orbach[8] for 6 spins, or directly by analytic computation. The point to observe is the following. In the antiferromagnetic case the ground state is always of spin zero. In the ferromagnetic case ($J<0$) the ground state is either of maximum spin 3 or spin zero. States of lower spin do cross the spin-3 state, but they all do so at values of α more negative than that necessary for spin-zero-level crossing.

The symmetries of the 6-spin complex can be described by placing them at the six corners of a regular hexagon (Fig. 4). The next-nearest neighbors can be joined by dotted lines, which are replaced by full lines for $\alpha = 1$, and we get the square graph of the hexagon, which, unlike the 4-spin case, is not complete. The next-nearest-neighbor interactions are already fully developed. For $\alpha \neq 1$, the symmetry of the graph is obviously that of the dihedral group D_6 of the hexagon. Therefore the corresponding permutation group that leaves the Hamiltonian invariant has 6 irreducible representations: A_1', A_2', A_1'', A_2'', E', and E'' of dimensions 1, 1, 1, 1, 2, and 2, respectively. The maximum degeneracy of the states is 2, which is the case in Fig. 3 excepting a few points of accidental degeneracy. Since two spin-zero states cross for the ground state, it is necessary that they must have different symmetry, according to von Neumann and Wigner. It is easy to verify that the state $a(\varphi_1 - \varphi_2) + b(\psi_3 + \psi_4 + \psi_5)$ (with a and b constants) has the symmetry A_1'', while the state $(\varphi_1 + \varphi_2)$ transforms as the representation A_2''. For $\alpha = 1$, the graph obviously represents an octahedron. The automorphisms of the graph constitute the well-known group O_h with 48 elements. The corresponding permutation group for 6 spins can easily be written down and be verified to leave the Hamiltonian invariant. The irreducible representations of O_h have dimensions 1, 2, and 3. Figure 3, however, shows that at $\alpha = 1$, besides triply degenerate levels, we have two levels of degeneracy 6. No group bigger than O_h has been found so far. Hence we may say that there are accidental degeneracies present at $\alpha = 1$.

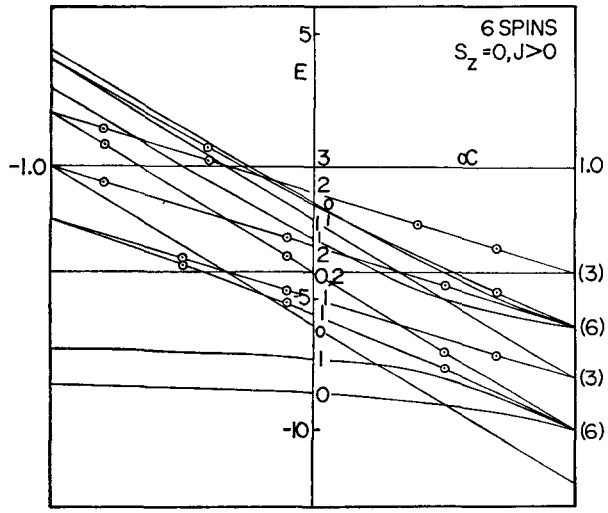

FIG. 3. Eigenvalues of $-H'$ for 6 spins. The small circle implies a degeneracy of 2. Level spins are indicated. The numbers on the right indicate the total number of levels of the cluster.

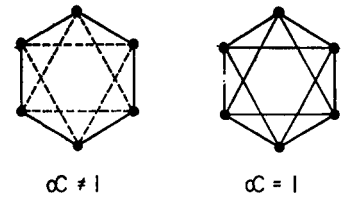

FIG. 4. Symmetries of the 6-spin complex. ($\alpha = 1$) represents an octahedron.

C. 8 Spins

For 8 spins there are 70 states in the $S_z = 0$ subspace and 14 of them have spin $S = 0$. Following Hulthén, we take the 14 basis functions as

$$\varphi_1 = [12]\,[34]\,[56]\,[78], \quad \varphi_2 = [23]\,[45]\,[67]\,[81],$$
$$\psi_1 = [23]\,[41]\,[56]\,[78], \quad \psi_2 = [34]\,[52]\,[67]\,[81],$$
$$\chi_1 = [23]\,[41]\,[67]\,[85],$$
$$\psi_3 = [12]\,[45]\,[63]\,[78], \quad \psi_4 = [23]\,[56]\,[74]\,[81],$$
$$\chi_3 = [81]\,[27]\,[45]\,[63],$$
$$\psi_5 = [12]\,[34]\,[67]\,[85], \quad \psi_6 = [23]\,[45]\,[78]\,[16],$$
$$\chi_2 = [78]\,[16]\,[34]\,[52],$$
$$\psi_7 = [34]\,[56]\,[81]\,[27], \quad \psi_8 = [45]\,[67]\,[12]\,[38],$$
$$\chi_4 = [12]\,[38]\,[56]\,[74]. \qquad (26)$$

The operation of the Hamiltonian

$$H' = -\frac{H - 4(1+\alpha)J}{J} = \sum_{i=1}^{8} \tfrac{1}{2}(1 - \boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_{i+1}) + \alpha \sum_{i=1}^{8} \tfrac{1}{2}(1 - \boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_{i+2}) \quad (27)$$

(with $\boldsymbol{\sigma}_9 \equiv \boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_{10} \equiv \boldsymbol{\sigma}_2$) gives

$$\begin{aligned}
H'\varphi_1 &= 8\varphi_1 + \psi_1 + \psi_3 + \psi_5 + \psi_7 \\
&\quad + \alpha(8\varphi_1 - 2\psi_1 - 2\psi_3 - 2\psi_5 - 2\psi_7), \\
H'\psi_1 &= 6\psi_1 + 2\varphi_1 + \psi_4 + \psi_6 + \chi_1 \\
&\quad + \alpha(8\psi_1 - 2\varphi_1 - \psi_4 - \psi_6 - 2\chi_1 - x_2 - x_1), \\
H'\psi_2 &= 6\psi_2 + 2\varphi_2 + \psi_5 + \psi_7 + \chi_2 \\
&\quad + \alpha(8\psi_2 - 2\varphi_2 - \psi_5 - \psi_7 - 2\chi_2 - x_5 - x_6), \\
H'\chi_1 &= 4\chi_1 + 2\varphi_2 + 2\psi_1 + 2\psi_5 \\
&\quad + \alpha(8\chi_1 - 2\psi_1 - 2\psi_5 - 2x_5 - 2x_7). \quad (28)
\end{aligned}$$

The structure of the remaining ten equations that we have not written down can be surmized from these. There are 8 extra states denoted by $x_1, \cdots, x_8$:

$$\begin{aligned}
x_1 &= [25]\,[36]\,[41]\,[78], & x_5 &= [14]\,[83]\,[52]\,[67], \\
x_2 &= [14]\,[83]\,[56]\,[27], & x_6 &= [47]\,[36]\,[52]\,[81], \\
x_3 &= [12]\,[63]\,[58]\,[47], & x_7 &= [23]\,[74]\,[61]\,[58], \\
x_4 &= [34]\,[85]\,[61]\,[72], & x_8 &= [83]\,[72]\,[45]\,[16].
\end{aligned} \quad (29)$$

The appearance of these states accentuates the difficulties of having the next-nearest-neighbor interactions. We have now to express these extra states in terms of our nonorthogonal basis (26). This is an extremely tedious but straightforward calculation along the line of Eqs. (20), (21), and (22). Hence, we get

$$\begin{aligned}
x_1 &= -\tfrac{5}{7}\varphi_1 - \tfrac{15}{28}\varphi_2 + \tfrac{6}{7}\psi_1 + \tfrac{4}{7}\psi_3 - \tfrac{1}{28}\psi_5 - \tfrac{3}{7}\psi_7 \\
&\quad + \tfrac{3}{14}\psi_2 + \tfrac{15}{56}\psi_4 - \tfrac{41}{56}\psi_6 + \tfrac{3}{14}\psi_8 + \tfrac{1}{56}\chi_1 + \tfrac{25}{28}\chi_2 \\
&\quad + \tfrac{9}{14}\chi_3 - \tfrac{3}{28}\chi_4, \\
x_2 &= -\varphi_1 + \psi_1 + \psi_7 - \psi_4 + \chi_4, \quad (30)
\end{aligned}$$

and six similar equations. Thus the operation of the Hamiltonian gives us, finally,

$$\begin{aligned}
H'\varphi_1 &= 8\varphi_1 + \psi_1 + \psi_3 + \psi_5 + \psi_7 \\
&\quad + \alpha[8\varphi_1 - 2\psi_1 - 2\psi_3 - 2\psi_5 - 2\psi_7], \\
H'\psi_1 &= 6\psi_1 + 2\varphi_1 + \psi_4 + \psi_6 + \chi_1 \\
&\quad + \alpha[-\tfrac{2}{7}\varphi_1 + \tfrac{15}{28}\varphi_2 + \tfrac{43}{7}\psi_1 - \tfrac{4}{7}\psi_3 + \tfrac{1}{28}\psi_5 \\
&\quad - \tfrac{4}{7}\psi_7 - \tfrac{3}{14}\psi_2 - \tfrac{15}{56}\psi_4 - \tfrac{15}{56}\psi_6 - \tfrac{3}{14}\psi_8 \\
&\quad - \tfrac{113}{56}\chi_1 - \tfrac{25}{28}\chi_2 - \tfrac{9}{14}\chi_3 - \tfrac{25}{28}\chi_4], \quad (31)
\end{aligned}$$

and 12 similar equations for the other functions. For $\alpha = 0$, Hulthén was able, by taking proper linear combinations, to decompose the 14×14 subspace into smaller subspaces. Because of the complicated nature of our coefficients, it is somewhat difficult to use the same technique; in fact the decomposition cannot be pushed as far as Hulthén did, while it was still feasible for the 6 spins. We have therefore diagonalized the 14×14 nonsymmetric matrix arising from Eq. (31) directly on a CDC 3600 computer, using the known values due to Orbach[8] at $\alpha = 0$ and the four analytically computed values $6 + 6\alpha \pm (2)^{\frac{1}{2}}$, $5 + 6\alpha \pm (5 + 4\alpha^2 - 4\alpha)^{\frac{1}{2}}$ as checks. Figure 5 represents the result. Only one new feature is present: in the case $J < 0$, $\alpha < 0$, the ferromagnetic ground state is first crossed by a spin-zero state which in turn is crossed by another spin-zero state. We have also analyzed the entire $S_z = 0$ subspace of 70 states and verified that the ferromagnetic ground state becomes unstable first with respect to a spin-zero state and that the ground state in all cases belong to either $S = 0$ or $S = 4$. The symmetry group is dihedral group D_8, with maximum allowed degeneracy 2. Nothing particularly striking happens at $\alpha = 1$.

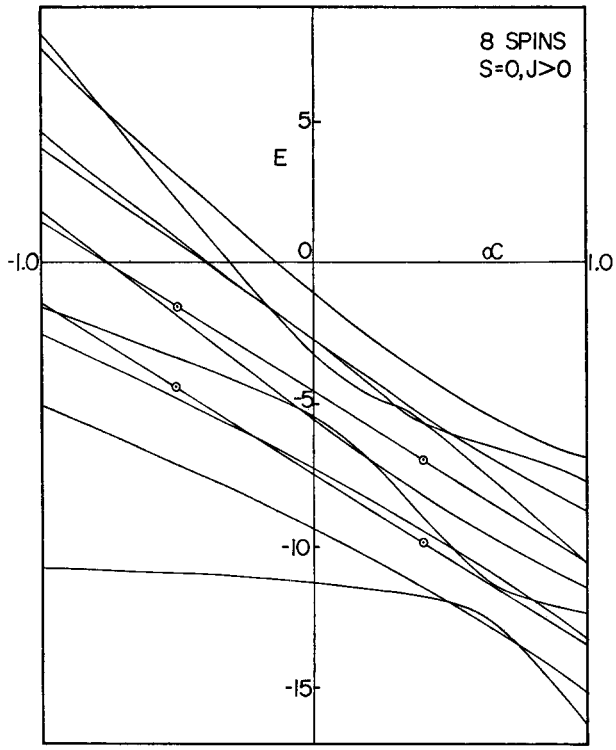

FIG. 5. Eigenvalues of $-H'$ for 8 spins. Two levels are indistinguishably close (the third excited state at $\alpha = 0$). The circles denote degeneracy of 2.

IV. ODD NUMBER OF SPINS (3, 5, 7)

We shall summarize the results for odd number of spins briefly. For 3 spins there is no question of next-nearest-neighbor interactions. For odd number of spins, the total spin can only be half integral, so we

TABLE I. Ground-state energy per spin for 5, 6, 7, and 8 particles.

No. of spins	5		6		7		8	
α	$J>0$	$J<0$	$J>0$	$J<0$	$J>0$	$J<0$	$J>0$	$J<0$
−1.0	−0.8944	−0.8944	−1.3875	−0.7280	−1.1093	−0.7674	−1.3496	−1.0120
−0.9	−0.8797	−0.8197	−1.3402	−0.6737	−1.0787	−0.6901	−1.3033	−0.9091
−0.8	−0.8650	−0.7450	−1.2930	−0.6263	−1.0483	−0.6137	−1.2574	−0.8185
−0.7	−0.8503	−0.6703	−1.2463	−0.5947	−1.0180	−0.5380	−1.2119	−0.7499
−0.6	−0.8355	−0.5955	−1.2000	−0.5333	−0.9878	−0.4639	−1.1669	−0.6943
−0.5	−0.8208	−0.5208	−1.1540	−0.4873	−0.9582	−0.4248	−1.1226	−0.6361
−0.4	−0.8061	−0.4461	−1.1087	−0.4420	−0.9287	−0.4026	−1.0788	−0.5784
−0.3	−0.7914	−0.3714	−1.0640	−0.3972	−0.8996	−0.3802	−1.0359	−0.5219
−0.2	−0.7765	−0.4000	−1.0200	−0.4000	−0.8710	−0.4000	−0.9938	−0.4600
−0.1	−0.7619	−0.4500	−0.9770	−0.4500	−0.8430	−0.4500	−0.9527	−0.4500
0	−0.7472	−0.5000	−0.9340	−0.5000	−0.8158	−0.5000	−0.9128	−0.5000
0.1	−0.7325	−0.5500	−0.8930	−0.5500	−0.7896	−0.5500	−0.8743	−0.5500
0.2	−0.7178	−0.6000	−0.8540	−0.6000	−0.7648	−0.6000	−0.8376	−0.6000
0.3	−0.7030	−0.6500	−0.8167	−0.6500	−0.7419	−0.6500	−0.8034	−0.6500
0.4	−0.6889	−0.7000	−0.7820	−0.7000	−0.7217	−0.7000	−0.7730	−0.7000
0.5	−0.6736	−0.7500	−0.7500	−0.7500	−0.7051	−0.7500	−0.7500	−0.7500
0.6	−0.6589	−0.8000	−0.8000	−0.8000	−0.6936	−0.8000	−0.7729	−0.8000
0.7	−0.6442	−0.8500	−0.8500	−0.8500	−0.7366	−0.8500	−0.7988	−0.8500
0.8	−0.6294	−0.9000	−0.9000	−0.9000	−0.8099	−0.9000	−0.8529	−0.9000
0.9	−0.6174	−0.9500	−0.9500	−0.9500	−0.8851	−0.9500	−0.9399	−0.9500
1.0	−0.6000	−1.0000	−1.0000	−1.0000	−0.9615	−1.0000	−1.0326	−1.0000

always have states with finite nonzero spin and the interesting question is whether the state of maximum or that of the minimum spin lies lowest. For 5 spins, we find that a state of spin $\frac{1}{2}$ has always the lowest energy for $J>0$, $-1\leq\alpha\leq 1$.

With $J<0$, $\alpha>0$, the state of maximum spin $\frac{5}{2}$ lies lowest; as the next-nearest-neighbor interaction becomes more and more antiferromagnetic, a state of spin $\frac{1}{2}$ becomes the lowest state and remains so. The symmetries of 5-spin system are analogous to those of 4-spin case, as the graph of 5 spins on a pentagon becomes a complete graph for $\alpha=1$. For $\alpha\neq 1$, the group of symmetries has 10 elements divided into 4 classes, with representations of order 1, 1, 2, and 2. For $\alpha=1$ we have the full symmetric group S_5.

The situation with 7 spins is similar. The ferromagnetic ground state becomes unstable at a value of $|\alpha|$ smaller than that for 5 spins.

In Table I we collected the ground-state energy per spin for all values of α for 5, 6, 7, and 8 spins. The cases of 3 and 4 spins are trivial and in any case can be determined from known results.

V. BOUNDS FOR THE GROUND-STATE ENERGY FOR LARGE N

In this section we shall present upper and lower bounds of the ground-state energy for arbitrarily large N, for the Hamiltonian of equation (3). Similar bounds can also be found for the square lattice, where the interactions between the nearest neighbors is J and that between next-nearest-neighbors is αJ.

Consider Eq. (3). Let $NzF(\alpha,J)$ be the ground-state energy, z is the coordination number or the number of nearest neighbors. Let ψ_0 be the exact ground-state wavefunction of the Hamiltonian (3). Then

$$\begin{aligned} NzF(\alpha,J) &= \langle\psi_0|\, H\,|\psi_0\rangle \\ &= \langle\psi_0|\,\tfrac{1}{2}J\sum_i \boldsymbol{\sigma}_i\cdot\boldsymbol{\sigma}_{i+1}\,|\psi_0\rangle \\ &\quad + \langle\psi_0|\,\tfrac{1}{2}J\alpha\sum_i \boldsymbol{\sigma}_i\cdot\boldsymbol{\sigma}_{i+2}\,\langle\psi_0\rangle \\ &= \langle\psi_0|\, H_1\,|\psi_0\rangle + \langle\psi_0|\, H_2\,\langle\psi_0\rangle, \end{aligned} \tag{32}$$

with

$$H_1 = \tfrac{1}{2}J\sum_i \boldsymbol{\sigma}_i\cdot\boldsymbol{\sigma}_{i+1}, \quad N+1\equiv 1, \tag{33}$$

$$H_2 = \tfrac{1}{2}J\alpha\sum_i \boldsymbol{\sigma}_i\cdot\boldsymbol{\sigma}_{i+2}, \quad N+1\equiv 1, \quad N+2\equiv 2. \tag{34}$$

Since ψ_0 is a variational function for H_1 as well as H_2, we have

$$NzF(\alpha,J)\geq \langle H_1\rangle_{\text{g.s.}} + \langle H_2\rangle_{\text{g.s.}}. \tag{35}$$

H_1 is precisely the Hamiltonian treated by Bethe and Hulthén and its ground-state energy is known. If J is negative, the ground state is ferromagnetic and has energy $-\frac{1}{4}Nz|J|$. If J is positive, the ground-state energy is the celebrated Bethe–Hulthén result, $-NJ(2\ln 2-\frac{1}{2}) = -\frac{1}{2}NzJ(0.88629)$. As for $\langle H_2\rangle_{\text{g.s.}}$, if αJ is negative, i.e., J and α have opposite sign. we have

$$\begin{aligned} \langle H_2\rangle_{\text{g.s.}} &= \left\langle -\tfrac{1}{2}|J|\,|\alpha|\sum_j \boldsymbol{\sigma}_j\cdot\boldsymbol{\sigma}_{j+2}\right\rangle_{\min} \\ &= -\tfrac{1}{2}|J|\,|\alpha|\left\langle\sum_j \boldsymbol{\sigma}_j\cdot\boldsymbol{\sigma}_{j+2}\right\rangle_{\max} \\ &= -\tfrac{1}{2}|J|\,|\alpha|\,\tfrac{1}{2}Nz. \end{aligned} \tag{36}$$

Hence, for J negative, α positive,

$$NzF(|\alpha|, -|J|) \geq -\tfrac{1}{4}Nz\,|J| - \tfrac{1}{4}\,|J|\,|\alpha|\,Nz$$

or

$$F(|\alpha|, -|J|) \geq -\tfrac{1}{4}\,|J| - \tfrac{1}{4}\,|J|\,|\alpha|. \tag{37}$$

For J positive, α negative,

$$NzF(-|\alpha|, |J|) \geq -\tfrac{1}{2}Nz\,|J|\,(0.88629) - \tfrac{1}{4}Nz\,|J|\,|\alpha|$$

or

$$F(-|\alpha|, |J|) \geq -\tfrac{1}{2}\,|J|\,(0.88629) - \tfrac{1}{4}\,|J|\,|\alpha|. \tag{38}$$

The case $\alpha J \geq 0$, i.e., α and J have the same sign, is more interesting. We have now to determine ground state of H_2 for $\alpha J \geq 0$. But taking N even, H_2 is the sum of two Hamiltonians of spins that interact only with nearest neighbors [odd-numbered spins interact only among themselves, so do the even-numbered ones, Fig. 6(a)], each sub-Hamiltonian containing $\frac{1}{2}N$ particles. As the two sub-Hamiltonians are not coupled, the ground-state energy follows directly from Hulthén's work:

$$\langle H_2\rangle_{\text{g.s.}} = -\tfrac{1}{2}Nz\,|J\alpha|\,(0.88629). \tag{39}$$

The result is also valid for odd N, as then the chain with next-nearest-neighbor interaction can be unfolded into a single chain with nearest-neighbor interactions (Fig. 6b). Thus, from (35) for $J < 0$, $\alpha < 0$,

$$NzF(-|\alpha|, -|J|) \geq -\tfrac{1}{4}Nz\,|J| - \tfrac{1}{2}Nz\,|J|\,|\alpha|\,(0.88629)$$

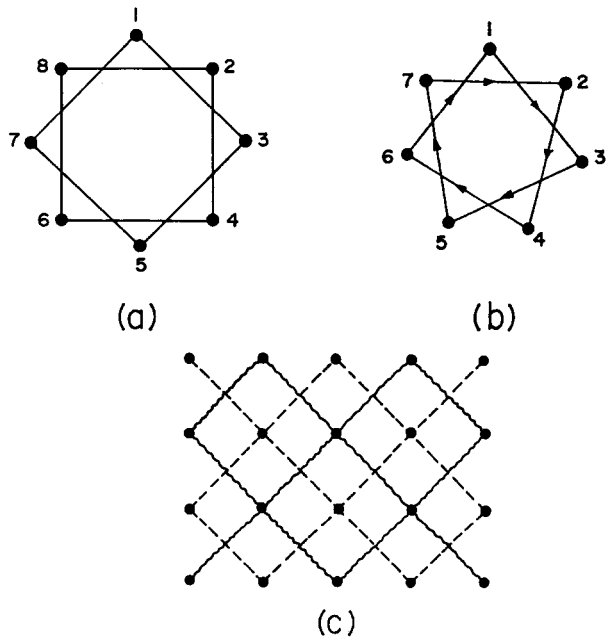

FIG. 6. (a) Decomposition of a linear chain into two independent chains (even N); (b) unfolding of a linear chain into a single chain by following the arrow (odd N); (c) decomposition of the square lattice.

or

$$F(-|\alpha|, -|J|) \geq -\tfrac{1}{4}\,|J| - \tfrac{1}{2}\,|J|\,|\alpha|\,(0.88629). \tag{40}$$

Similarly, for $J > 0$, $\alpha > 0$,

$$NzF(\alpha, J) \geq -\tfrac{1}{2}NzJ(0.88629) - \tfrac{1}{2}NzJ\alpha(0.88629)$$

or

$$F(\alpha, J) \geq -\tfrac{1}{2}J(0.88629)(1 + \alpha). \tag{41}$$

To obtain an upper bound, consider a state ψ_s completely symmetric in all spins:

$$\boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_{i+1}\psi_s = \psi_s,$$
$$\boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_{i+2}\psi_s = \psi_s. \tag{42}$$

Then for any J and α

$$NzF(\alpha, J) \leq \langle\psi_s|\, H\,|\psi_s\rangle = \tfrac{1}{4}JNz + \tfrac{1}{4}\alpha JNz$$

or

$$F(\alpha, J) \leq \tfrac{1}{4}J + \tfrac{1}{4}J\alpha. \tag{43}$$

Since for $J < 0$, $\alpha > 0$ the upper and the lower bounds coincide, we have an exact solution for the situation when all interactions are ferromagnetic. For $J > 0$, the upper bound (43) is trivial. To get an improved upper bound, we consider as a variational function the following alternating function (N even):

$$\psi_A = \alpha(1)\beta(2)\alpha(3)\beta(4)\cdots. \tag{44}$$

Then,

$$NzF(\alpha, |J|) \leq \langle\psi_A|\, H\,|\psi_A\rangle = -\tfrac{1}{4}Nz\,|J| + \tfrac{1}{4}Nz\,|J|\alpha$$

or

$$F(\alpha, |J|) \leq -\tfrac{1}{4}\,|J| + \tfrac{1}{4}\,|J|\,\alpha = -\tfrac{1}{4}J(1 - \alpha). \tag{45}$$

For large odd N, this result will also hold. Table II contains all the bounds for various cases.

The particular topological property of the linear chain which enabled us to determine the lower bound is shared by the square lattice [Fig. 6(c)]; that is, the square lattice with *purely* next-nearest-neighbor interaction can be decomposed into two mutually noninteracting square lattices with half as many particles and nearest-neighbor interactions only. The exact ground-state energy of the Hamiltonian with nearest-neighbor interaction for a square lattice is not known. However, an inequality of the type (35) still holds if we know the lower bounds to the ground-state energy. Such lower bounds can be obtained from the work of Yang and Yang.[9]

Write for a square lattice

$$H = \tfrac{1}{2}J\sum \boldsymbol{\sigma}\cdot\boldsymbol{\sigma}' + \tfrac{1}{2}J\alpha\sum{}' \boldsymbol{\sigma}\cdot\boldsymbol{\sigma}'; \tag{46}$$

the second sum goes over next-nearest-neighbor pairs only. Call $NzF^s(\alpha, J)$ the ground-state energy, with z the number of nearest- as well as next-nearest

TABLE II. Upper and lower bounds for the ground-state energy for large N. When comparing with Table I, note the difference that the coordination number z appears in the definition of F.

$J<0, \quad \alpha>0$	$J>0, \quad \alpha>0$
$F(\alpha,J) \geq -\frac{1}{4}\lvert J\rvert(1+\alpha)$.	$F(\alpha,J) \geq -\frac{1}{2}(0.88629)J(1+\alpha)$.
$F(\alpha,J) \leq -\frac{1}{4}\lvert J\rvert(1+\alpha)$.	$F(\alpha,J) \leq -\frac{1}{4}J+\frac{1}{4}J\alpha$.
$J<0, \quad \alpha<0$	$J>0, \quad \alpha<0$
$F(\alpha,J) \geq -\frac{1}{4}\lvert J\rvert-\frac{1}{2}\lvert J\rvert\lvert\alpha\rvert(0.88629)$.	$F(\alpha,J) \geq -\frac{1}{2}(0.88629)J-\frac{1}{4}J\lvert\alpha\rvert$.
$F(\alpha,J) \leq -\frac{1}{4}\lvert J\rvert+\frac{1}{4}\lvert J\rvert\lvert\alpha\rvert$.	$F(\alpha,J) \leq -\frac{1}{4}J-\frac{1}{4}\lvert\alpha\rvert J$.

neighbors and N even. Using the completely symmetric state, we get an upper bound immediately for any J and α:

$$NzF^s(\alpha,J) \leq \tfrac{1}{2}J\cdot\tfrac{1}{2}Nz+\tfrac{1}{2}J\alpha\cdot\tfrac{1}{2}Nz$$

or

$$F^s(\alpha,J) \leq \tfrac{1}{4}J(1+\alpha). \tag{47}$$

For $J<0$, $\alpha>0$, the lower bound is easily obtained:

$$\begin{aligned} NzF^s(\alpha,-\lvert J\rvert) &\geq \langle -\tfrac{1}{2}\lvert J\rvert\textstyle\sum\boldsymbol{\sigma}\cdot\boldsymbol{\sigma}'\rangle_{\text{g.s.}} \\ &\quad +\langle -\tfrac{1}{2}\alpha\lvert J\rvert\textstyle\sum'\boldsymbol{\sigma}\cdot\boldsymbol{\sigma}'\rangle_{\text{g.s.}} \\ &= -\tfrac{1}{2}\lvert J\rvert\tfrac{1}{2}Nz-\tfrac{1}{2}\alpha\lvert J\rvert\tfrac{1}{2}Nz \end{aligned}$$

or

$$F^s(\alpha,-\lvert J\rvert) \geq -\tfrac{1}{4}\lvert J\rvert(1+\alpha), \tag{48}$$

so that we have an exact solution as before.

For $J>0$, Yang and Yang give a lower bound to the ground-state energy of $\frac{1}{2}J\sum\boldsymbol{\sigma}\cdot\boldsymbol{\sigma}'$, for any fixed magnetization. It follows from the Lieb–Mattis theorem that the ground state has spin zero and no magnetization. So we obtain the inequality

$$\langle\tfrac{1}{2}\lvert J\rvert\textstyle\sum\boldsymbol{\sigma}\cdot\boldsymbol{\sigma}'\rangle_{\text{g.s.}} \geq \tfrac{3}{4}\lvert J\rvert Nz. \tag{49}$$

Now consider the lower bound for $J<0$, $\alpha<0$:

$$\begin{aligned} NzF^s(-\lvert\alpha\rvert,-\lvert J\rvert) &\geq \langle -\tfrac{1}{2}\lvert J\rvert\textstyle\sum\boldsymbol{\sigma}\cdot\boldsymbol{\sigma}'\rangle_{\text{g.s.}} \\ &\quad +\langle\tfrac{1}{2}\lvert\alpha\rvert\lvert J\rvert\textstyle\sum'\boldsymbol{\sigma}\cdot\boldsymbol{\sigma}'\rangle_{\text{g.s.}} \\ &\geq -\tfrac{1}{4}\lvert J\rvert Nz-\tfrac{3}{4}\lvert\alpha\rvert\lvert J\rvert Nz \end{aligned}$$

or

$$F^s(-\lvert\alpha\rvert,-\lvert J\rvert) \geq -\tfrac{1}{4}\lvert J\rvert-\tfrac{3}{4}\lvert J\rvert\lvert\alpha\rvert. \tag{50}$$

For $J>0$, $\alpha<0$, we have

$$\begin{aligned} NzF^s(-\lvert\alpha\rvert,J) &\geq \langle\tfrac{1}{2}J\textstyle\sum\boldsymbol{\sigma}\cdot\boldsymbol{\sigma}'\rangle_{\text{g.s.}} \\ &\quad +\langle -\tfrac{1}{2}J\rvert\alpha\lvert\textstyle\sum'\boldsymbol{\sigma}\cdot\boldsymbol{\sigma}'\rangle_{\text{g.s.}} \\ &\geq -\tfrac{3}{4}JNz-\tfrac{1}{4}\lvert\alpha\rvert JNz \end{aligned}$$

or

$$F^s(-\lvert\alpha\rvert,J) \geq -\tfrac{3}{4}J-\tfrac{1}{4}J\lvert\alpha\rvert. \tag{51}$$

For $J>0$, $\alpha>0$, we have

$$NzF^s(\lvert\alpha\rvert,\lvert J\rvert) \geq -\tfrac{3}{4}JNz-\tfrac{3}{4}J\alpha Nz$$

or

$$F^s(\lvert\alpha\rvert,\lvert J\rvert) \geq -\tfrac{3}{4}J(1+\alpha). \tag{52}$$

By using a variational function which has all spins up in one sublattice and all spins down in the other, we can improve the upper bound in the case $J>0$ to get

$$F^s(\alpha,\lvert J\rvert) \leq -\tfrac{1}{4}J+\tfrac{1}{4}J\alpha. \tag{53}$$

Finally, using the Yang and Yang result for one dimension, we obtain lower bounds for the linear chain, which are naturally not as good as those using the exact result:

$$\begin{aligned} F(\lvert\alpha\rvert,\lvert J\rvert) &\geq -\tfrac{3}{4}J(1+\alpha) = -0.75J(1+\alpha), \\ F(-\lvert\alpha\rvert,-\lvert J\rvert) &\geq -\tfrac{1}{4}\lvert J\rvert-\tfrac{3}{4}\lvert J\rvert\lvert\alpha\rvert \\ &= -0.25\lvert J\rvert-0.75\lvert J\rvert\lvert\alpha\rvert, \\ F(-\lvert\alpha\rvert,\lvert J\rvert) &\geq -\tfrac{3}{4}\lvert J\rvert-\tfrac{1}{4}\lvert J\rvert\lvert\alpha\rvert \\ &= -0.75\lvert J\rvert-0.25\lvert J\rvert\lvert\alpha\rvert. \end{aligned} \tag{54}$$

VI. DISCUSSION

The general trend of the results shows the usual nonmagnetic character of the ground states except when all interactions are ferromagnetic. The most interesting case is $J<0$, $\alpha<0$. Considering the limit $\alpha\to-\infty$ and applying the Lieb–Mattis theorem, we may easily see that, for sufficiently large and negative α, the ground state must belong to spin zero. The interesting trend of the results on short chains indicates that the ferromagnetic state becomes unstable with respect to a spin-zero state for the smallest absolute value of α. It is, of course, easy to show that the ferromagnetic state is unstable with respect to spin waves for $\alpha<-\frac{1}{4}$, but spin waves have very high

spin, $S = \frac{1}{2}N - 1$. The indications are that the spin-zero state crosses the ferromagnetic state at a very small value of α for large N. Quite possibly $\alpha = 0$ itself is a critical value; this situation is probably characteristic of one-dimensional system and is almost certainly untrue in three dimensions. Nevertheless, it emphasizes the rather tenuous nature of the ferromagnetic states.

ACKNOWLEDGMENTS

We would like to thank A. S. Anikhindi, B. K. Chanda, and Y. S. T. Rao for their generous help and advice in numerical computation.

APPENDIX

We give here the character tables of D_4, D_6, D_8, and D_5 for convenient reference.

D_4

Class	E	C_2	C_4	C_2	C_2'
Order of class	1	1	2	2	2
A_1	1	1	1	1	1
A_2	1	1	1	−1	−1
B_1	1	1	−1	1	−1
B_2	1	1	−1	−1	1
E	2	−2	0	0	0

D_6

Class	E	C_6	C_6^2	C_6^3	C_2	C_2'
Order of class	1	2	2	1	3	3
A_1'	1	1	1	1	1	1
A_2'	1	1	1	1	−1	−1
A_1''	1	−1	1	−1	−1	−1
A_2''	1	−1	1	−1	1	−1
E'	2	−1	−1	2	0	0
E''	2	1	−1	−2	0	0

D_8

Class	E	C_8	C_8^2	C_8^3	C_8^4	C_2	C_2'
Order of class	1	2	2	2	1	4	4
A_1	1	1	1	1	1	1	1
A_2	1	1	1	1	1	−1	−1
B_1	1	−1	1	−1	1	1	−1
B_2	1	−1	1	−1	1	−1	1
E_1	2	1	0	−1	−2	0	0
E_2	2	−1	0	1	−2	0	0
E_3	2	0	−2	0	2	0	0

D_5

Class	E	C_5	C_5^2	C_2
Order of class	1	2	2	5
A_1	1	1	1	1
A_2	1	1	1	−1
E_1	2	$\frac{1}{2}(\sqrt{5}-1)$	$-\frac{1}{2}(\sqrt{5}+1)$	0
E_2	2	$-\frac{1}{2}(\sqrt{5}+1)$	$\frac{1}{2}(\sqrt{5}-1)$	0

On Next-Nearest-Neighbor Interaction in Linear Chain. II

CHANCHAL K. MAJUMDAR AND DIPAN K. GHOSH
Tata Institute of Fundamental Research, Bombay, India

(Received 29 May 1968)

Continuing our work on the ground-state properties of the Hamiltonian

$$H = \tfrac{1}{2}J\sum_{i=1}^{N}\boldsymbol{\sigma}_i\cdot\boldsymbol{\sigma}_{i+1} + \tfrac{1}{2}J\alpha\sum_{i=1}^{N}\boldsymbol{\sigma}_i\cdot\boldsymbol{\sigma}_{i+2}, \quad -1 \le \alpha \le 1,$$

we have completed the study of 10 spins. The results of short-chain calculations provide better upper and lower bounds of the ground-state energy per particle as $N \to \infty$, but no simple formula can be fitted to the data to get this limit for all α. For $J > 0$ and $\alpha = \frac{1}{2}$, however, this is exactly found to be $-\frac{3}{4}J$. Some upper and lower bounds for the free energy are also derived.

In a previous work,[1] we studied the ground-state properties of the Hamiltonian

$$H = \tfrac{1}{2}J\sum_{i=1}^{N}\boldsymbol{\sigma}_i\cdot\boldsymbol{\sigma}_{i+1} + \tfrac{1}{2}J\alpha\sum_{i=1}^{N}\boldsymbol{\sigma}_i\cdot\boldsymbol{\sigma}_{i+2} \equiv H_0 + \alpha H' \tag{1}$$

($N+1 \equiv 1$, $N+2 \equiv 2$) for both positive and negative J and $-1 \le \alpha \le 1$. We obtained the upper and lower bounds of the ground-state energy per particle in the limit $N \to \infty$ for (1), as well as for its two-dimensional version with a square lattice. To obtain more specific information about the dependence of the ground-state energy on α, we studied short linear chains up to 8 particles. It has been shown[2] for linear chain with $J > 0$ and only nearest-neighbor interaction ($\alpha = 0$) that simple empirical formulas, such as

$$E_N = E_\infty + (a_2/N^2) + \cdots, \tag{2}$$

for the ground-state energy per particle E_N, exist and these can be used to get E_∞, the limiting value as $N \to \infty$, with great accuracy. In order to examine whether such a formula exists for all α, we have now pushed our calculations to a chain of 10 particles. The entire data are presented here in Figs. 1 and 2. Although one can visualize the behavior of the ground-state energy per particle E_∞ for all α fairly well, it is extremely hard to find any empirical formula of the type (2) for all α or even a limited but useful region of α. Nevertheless, the short-chain calculations appear to provide much better bounds for E_∞ than we have obtained previously.

An interesting by-product of these calculations is to show that the exact ground-state energy per particle E_∞, for $J > 0$, $\alpha = \frac{1}{2}$, is $-\frac{3}{4}J$. The ground state is of spin zero, but degenerate, and the corresponding eigenfunctions for finite and even N can be written down easily.

Some exact bounds for the free energy of the system (1) can be obtained. More detailed information on the finite temperature properties of system with nearest- as well as next-nearest-neighbor interactions is available from the work of Dalton.[3]

I. RESULTS OF SHORT-CHAIN CALCULATION

The method of computation for the 10-spin case is the same as employed in Ref. 1 and is originally due to Hulthén.[4] There are 42 spin-zero states and we take them as

$$\begin{aligned}
\varphi_1 &= [12]\,[34]\,[56]\,[78]\,[90],\\
\varphi_2 &= [23]\,[45]\,[67]\,[89]\,[01],\\
\psi_1 &= [23]\,[41]\,[56]\,[78]\,[90],\\
\chi_1 &= [23]\,[41]\,[67]\,[85]\,[90],\\
\Gamma_1 &= [23]\,[45]\,[16]\,[78]\,[90],\\
\Lambda_1 &= [23]\,[41]\,[67]\,[89]\,[50],\\
\omega_1 &= [23]\,[41]\,[78]\,[96]\,[50].
\end{aligned} \tag{3}$$

There are 10 states each of the type ψ, χ, and Λ, and five of each of the type Γ and ω. Here 0 stands for the tenth spin. The square bracket $[lm] = \alpha(l)\beta(m) - \beta(l)\alpha(m)$, where α and β are the up and down spin states. A simple way of generating these states is to start with φ_1 and φ_2 and operate with the nearest-neighbor-interaction part H_0 only. Note that these states are not orthogonal. Operating with H' on these states, we generate, besides these, other states of similar structure, for example,

$$H'\psi_1 = 10\psi_1 - 2\varphi_1 - 2\chi_1 - \Gamma_1 - \Gamma_2 - \theta_1 - \theta_2, \tag{4}$$

[1] C. K. Majumdar and D. K. Ghosh, J. Math. Phys. **10**, 1392 (1969) (preceding paper).
[2] J. Bonner and M. Fisher, Phys. Rev. **135**, A640 (1964).
[3] N. Dalton, Proc. Phys. Soc. (London) **88**, 659 (1964).
[4] L. Hulthén, Arkiv Mat. Astron. Fysik **26A**, No. 11 (1938).

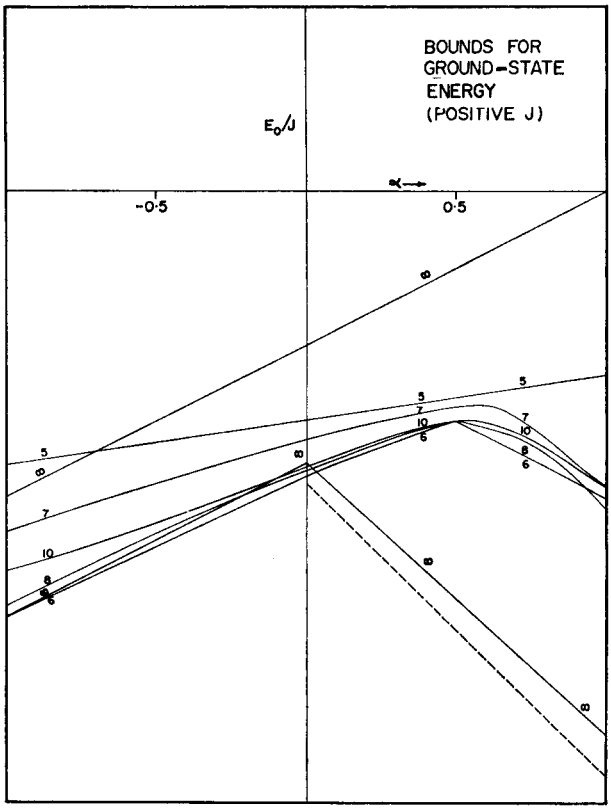

FIG. 1. Ground-state energy per spin for $J > 0$ (antiferromagnetic case). The curves marked ∞ are bounds obtained directly (Ref. 1) for H in Eq. (1), and the dashed curve is obtained from Eq. (23); the others are short-chain results. (The vertical scale may be fixed from Table I.)

where

$$\theta_1 = [36]\,[25]\,[41]\,[78]\,[90],$$
$$\theta_2 = [29]\,[30]\,[41]\,[56]\,[78]. \qquad (4')$$

The same difficulty was present in the 8-spin case, but now it is far worse; we have 45 such states which must be reexpressed in terms of the above nonorthogonal basis by a long, tedious calculation. The presence of these terms therefore destroys a basic simplifying feature that prevails in the problem with $\alpha = 0$—it is now impossible to reduce the matrix of the Hamiltonian to smaller submatrices by inspection. The matrix is nonsymmetric and has been diagonalized on a CDC 3600 computer. The spin-zero states known at $\alpha = 0$ from the work of Orbach[5] serve as a useful check. The ground-state energy per spin is given in Table I. For $J < 0$, we have assumed, on the basis of our previous experience, that the spin-zero state becomes the ground state when α is large and negative.

At $\alpha = 0$, $J > 0$, the limit E_∞ is approached from below for even number of particles and from above for odd number of particles in the chain. By continuity this would hold for small α. From Fig. 1 it appears that the even-chain (i.e., a chain containing even number of spins) results are converging fast to a limit for $0 \leq \alpha \leq \frac{1}{2}$. In the other regions nothing definite can be said about convergence. But it appears that the even chain results are below and the odd chain results are above the exact answer for $-1 \leq \alpha \leq 1$. Assuming this to be true, we get the lower and the upper bounds given by the results of 10 spins and seven spins, respectively. For $J < 0$, the odd-chain results are not useful and the even-chain results may be assumed to approach the right answer from above; we get its bounds by our previous calculation and the ten spin results.

II. EXACT RESULT AT $\alpha = \frac{1}{2}$ FOR $J > 0$

Since at $\alpha = \frac{1}{2}$ $(J > 0)$ all the even-chain results indicate the same value, it is apparent that we may have an exact answer here, $E_\infty/J = -0.75$. Actually, one can guess the eigenfunctions for the Hamiltonian here. Take N even and consider

$$\varphi_1 = [12]\,[34]\cdots[N-3, N-2][N-1, N]. \qquad (5)$$

Now

$$H = \tfrac{1}{2}NJ(1+\alpha) - J\sum_{i=1}^{N}\tfrac{1}{2}(1 - \boldsymbol{\sigma}_i\cdot\boldsymbol{\sigma}_{i+1}) - J\alpha\sum_{i=1}^{N}\tfrac{1}{2}(1 - \boldsymbol{\sigma}_i\cdot\boldsymbol{\sigma}_{i+2}) \equiv \tfrac{1}{2}NJ(1+\alpha) - J\tilde{H}. \qquad (6)$$

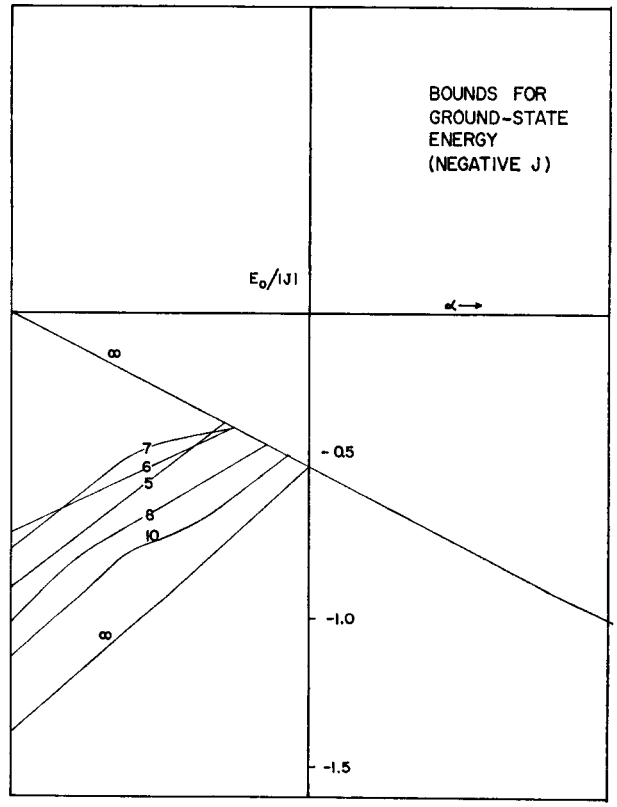

FIG. 2. Ground-state energy per spin for $J < 0$ (ferromagnetic case). The notation is the same as in Fig. 1.

[5] R. Orbach, Phys. Rev. **112**, 309 (1958); **115**, 1181 (1959).

TABLE I. Ground-state energy per spin for 10 particles (divided by $|J|$).

α	$J>0$	$J<0$	α	$J>0$	$J<0$
−1.0	−1.2429	−1.1310	0.1	−0.8679	−0.5500
−0.9	−1.2130	−1.0398	0.2	−0.8337	−0.6000
−0.8	−1.1816	−0.9496	0.3	−0.8013	−0.6500
−0.7	−1.1492	−0.8610	0.4	−0.7721	−0.7000
−0.6	−1.1157	−0.7746	0.5	−0.7500	−0.7500
−0.5	−1.0813	−0.7339	0.6	−0.7568	−0.8000
−0.4	−1.0463	−0.6992	0.7	−0.7850	−0.8500
−0.3	−1.0107	−0.6366	0.8	−0.8384	−0.9000
−0.2	−0.9748	−0.5642	0.9	−0.9019	−0.9500
−0.1	−0.9389	−0.4878	1.0	−0.9659	−1.0000
0.0	−0.9031	−0.5000			

Then,

$$\begin{aligned}\tilde{H}\varphi_1 = {} & N\varphi_1 + [23]\,[41]\,[56]\cdots[N-1,N] \\ & + [12]\,[45]\,[63]\,[78]\cdots[N-1,N] \\ & + \cdots + [34]\cdots[2,N-1][N1] \\ & - \alpha\{[13]\,[42]\,[56]\cdots[N-1,N] \\ & + [24]\,[31]\,[56]\cdots[N-1,N] + \cdots \\ & + [N-1,1]\,[2N]\,[34]\cdots[N-3,N-2] \\ & + [N2]\,[1,N-1]\,[34]\cdots[N-3,N-2]\}. \end{aligned} \tag{7}$$

The structure of the terms is obvious. The terms (second, third, etc.,on the right) are those we should denote by $\psi_1, \psi_3, \psi_5, \cdots$ in accordance with our previous notation. If we use the identity

$$[kl]\,[mn] + [kn]\,[lm] + [km]\,[nl] = 0, \tag{8}$$

termwise for the coefficient of α, we get

$$\begin{aligned}\tilde{H}\varphi_1 = {} & N(1+\alpha)\varphi_1 + (1-2\alpha)\{[23]\,[41]\cdots \\ & \times [N-1,N] + [12]\,[45]\,[63]\cdots[N-1,N] \\ & + \cdots + [N1]\,[34]\cdots[2,N-1]\}. \end{aligned} \tag{9}$$

For $\alpha = \frac{1}{2}$,

$$\tilde{H}\varphi_1 = \tfrac{3}{2}N\varphi_1, \tag{10}$$

so that

$$H\varphi_1 = (\tfrac{3}{4}NJ - \tfrac{3}{2}NJ)\varphi_1 = -\tfrac{3}{4}NJ\varphi_1. \tag{11}$$

Hence the energy per spin is $-\frac{3}{4}J$ for $N\to\infty$. Similarly, we can show that, for N even,

$$\varphi_2 = [23]\,[45]\,[67]\cdots[N1] \tag{12}$$

is another eigenfunction of H belonging to the same eigenvalue. For odd N, the states are magnetic, but as $N\to\infty$, the energy per spin can only differ by terms of $O(1/N)$. Hence at $\alpha = \frac{1}{2}$, $E_N = -\frac{3}{4}NJ$ for $N\to\infty$, irrespective of evenness or oddness of N. Strictly speaking, we have shown that (5) and (12) are just eigenfunctions of H and, therefore, the eigenvalue $-\frac{3}{4}NJ$ greater than or equal to the exact ground-state energy. But the numerical evidence of Fig. 1 leaves little doubt that it is the exact ground-state energy.

III. BOUNDS FOR THE FREE ENERGY

Here we shall present some bounds on the free energy for the Hamiltonian (1) in the linear-chain case.

The upper bound for the free energy

$$F = -\beta^{-1}\ln Z = -\beta^{-1}\ln \mathrm{Tr}\,[\exp(-\beta H)]$$

is obtained from Peierls's theorem.[6] If F_0 is the free energy calculated with only the diagonal elements of the Hamiltonian (1),

$$H_D = \tfrac{1}{2}J\sum_{i=1}^{N}\sigma_i^z\sigma_{i+1}^z + \tfrac{1}{2}J\alpha\sum_{i=1}^{N}\sigma_i^z\sigma_{i+2}^z, \tag{13}$$

we know $F \le F_0$.

But H_D is the Hamiltonian of the Ising model with second-neighbor interactions. This can be solved by the transfer-matrix method in one dimension (but not, as far as we know, in two or three dimensions). This gives an upper bound for the free energy: Put $F = Nf$; we have

$$\begin{aligned} f \le -\tfrac{1}{2}k_BT\ln\,[&\exp(-J\alpha/k_BT)\cosh(J/k_BT) \\ & + \exp(J\alpha/k_BT) + \{\exp(-2J\alpha/k_BT) \\ & \times \sinh^2(J/k_BT) + 4\cosh^2(J/2k_BT)\}^{\frac{1}{2}}]. \end{aligned} \tag{14}$$

In the case $J>0$, $\alpha>0$, we can also get a lower bound of the free energy by utilizing the convexity property.[7] The method was used by Griffiths[8] and utilized the solution of the X–Y model by Schultz, Lieb, and Mattis.[9] We have to use the fact that the linear chain with only the next-nearest-neighbor interaction can be decomposed into two noninteracting linear chains with half the number of spins each and with only the nearest-neighbor interaction (consider N even for convenience).

Write (1) as

$$\begin{aligned} H = {} & X + Y + Z + X_0 + Y_0 + Z_0 \\ & + X_e + Y_e + Z_e, \end{aligned} \tag{15}$$

where

$$X = \tfrac{1}{2}J\sum\sigma_i^x\sigma_{i+1}^x, \quad Y = \tfrac{1}{2}J\sum\sigma_i^y\sigma_{i+1}^y, \quad Z = \tfrac{1}{2}J\sum\sigma_i^z\sigma_{i+1}^z, \tag{16a}$$

$$X_o = \tfrac{1}{2}J\alpha\sum\sigma_i^x\sigma_{i+2}^x, \quad Y_o = \tfrac{1}{2}J\alpha\sum\sigma_i^y\sigma_{i+2}^y, \quad Z_o = \tfrac{1}{2}J\alpha\sum\sigma_i^z\sigma_{i+2}^z, \tag{16b}$$

$$X_e = \tfrac{1}{2}J\alpha\sum\sigma_i^x\sigma_{i+2}^x, \quad Y_e = \tfrac{1}{2}J\alpha\sum\sigma_i^y\sigma_{i+2}^y, \quad Z_e = \tfrac{1}{2}J\alpha\sum\sigma_i^z\sigma_{i+2}^z. \tag{16c}$$

In (16b) the sum is over the odd-numbered spins and in (16c) it is over the even-numbered spins.

The convexity of the free energy provides the

[6] R. E. Peierls, Phys. Rev. **54**, 918 (1938); for a recent proof, see D. Ruelle, Helv. Phys. Acta **36**, 789 (1963).
[7] D. Ruelle, Helv. Phys. Acta **36**, 789 (1963).
[8] R. B. Griffiths, Phys. Rev. **136**, A751 (1964).
[9] T. D. Schultz, E. Lieb, and D. Mattis, Ann. Phys. (N.Y.) **16**, 407 (1961).

relation

$$F\left(\sum_i \alpha_i A_i\right) \geq \sum_i \alpha_i F(A_i), \quad \sum_i \alpha_i = 1. \tag{17}$$

Rewriting (15) as

$$H = \tfrac{1}{9}[\tfrac{9}{2}(X + Y) + \tfrac{9}{2}(Y + Z) + \tfrac{9}{2}(Z + X) + \cdots + \tfrac{9}{2}(Z_e + X_e)], \tag{18}$$

and using the following relations,which hold because of symmetry,

$$F(T, \tfrac{9}{2}(X + Y)) = F(T, \tfrac{9}{2}(Y + Z)) = F(T, \tfrac{9}{2}(Z + X)), \tag{19}$$

etc., we obtain

$$F(T) \geq \tfrac{1}{3}F(T, \tfrac{9}{2}(X + Y)) + \tfrac{2}{3}F(T, \tfrac{9}{2}(X_e + Y_e)). \tag{20}$$

Now the Schultz, Lieb, and Mattis solution of the X–Y model gives

$$F(J, T) = -Nk_BT\left[\ln 2 + (2/\pi)\int_0^{\pi/2} dk \times \ln\{\cosh[(J/k_BT)\cos k]\}\right]. \tag{21}$$

Remember that the interaction in the second term on the right-hand side is $J\alpha$ and the number of particles is $\frac{1}{2}N$ in $X_e + Y_e$, so we get

$$F(T) \geq -\tfrac{1}{3}Nk_BT\left[\ln 2 + (2/\pi)\int_0^{\pi/2} dk \times \ln\{\cosh[(9J/2k_BT)\cos k]\}\right] - \tfrac{1}{3}Nk_BT\left[\ln 2 + (2/\pi)\int_0^{\pi/2} dk \times \ln\{\cosh[(9J\alpha/2k_BT)\cos k]\}\right]. \tag{22}$$

The lower limit for f, the free energy per particle, follows after division by N.

Taking the limit $T \to 0$, we get a lower bound for the ground-state energy per spin:

$$E_\infty \geq -0.955J(1 + \alpha). \tag{23}$$

We have obtained a lower bound better than this in our previous work.

Such results are difficult to obtain in higher dimensions, but there exact series expansion results are available.[3]

IV. DISCUSSION

In this study we have unravelled some of the difficulties of having second-neighbor interaction as apparent in the linear chain. For $J < 0$, the most interesting feature is the singularity at $\alpha = 0$ and that the ground state becomes one of spin zero for arbitrarily small second-neighbor interaction of opposite sign. Such instability is to be expected for the same reason as destroys the long-range order of the ferromagnetic ground state in one dimension at all finite temperatures. For $J > 0$, the existence of an exact result at $\alpha = \frac{1}{2}$ is surprising, but $\alpha = \frac{1}{2}$ must be a singular point. Here $\alpha = 0$ appears to be smooth, but may not really be analytic, thus explaining the failure of the existence of simple analytic extrapolation formula. While the singularity at $\alpha = 0$ may be removed in two or three dimensions, the existence of second-neighbor interaction brings into play complicated topologies; the trouble in the 8 and 10 spins of the nature of Eq. (4) arose precisely from these reasons.

Phase Transition in Zero Dimensions: A Remark on the Spherical Model

ELLIOTT H. LIEB* AND COLIN J. THOMPSON†
Mathematics Department, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received 31 January 1969)

It is shown that the spherical model consisting of N spins with nonzero interaction between two spins only has a phase transition in the limit $N \to \infty$. This is a counterexample to a suggestion of Kac which states that an Ising model will have a transition if the corresponding spherical model has a transition. Possible modifications of Kac's conjecture are suggested and discussed.

SECTION I

It has been suggested by Kac[1] that an Ising model will have a phase transition if the corresponding spherical model has a phase transition.

The purpose of this note is to present a counterexample and to suggest that the conjecture may well be true if the interaction potential is purely ferromagnetic and periodic.

The example we consider is in itself amusing: it has a phase transition and is in a sense zero-dimensional.

SECTION II

Consider a set of N spins s_i with interaction energy

$$E = -\sum_{1\le i<j\le N} J_{ij}s_is_j, \tag{1}$$

for the Ising model $s_i = \pm 1$ and for the spherical model[2] $-\infty < s_i < \infty$ with the constraint

$$\sum_{i=1}^{N} s_i^2 = N. \tag{2}$$

We take the particular case

$$\begin{aligned} J_{ij} &= 2J, \quad \text{when} \quad i = 1 \quad \text{and} \quad j = 2, \\ &= 0, \quad \text{otherwise}, \end{aligned} \tag{3}$$

i.e., only spins 1 and 2 interact.

We wish to emphasize that this interaction is not pathological in any sense. The above coupling constant is a perfectly finite and fixed number, e.g., π.

The Ising-model partition function is given by

$$\begin{aligned} Z_N^I &= 2^{-N}\sum_{\mu_1=\pm1}\sum_{\mu_2=\pm1}\cdots\sum_{\mu_N=\pm1} e^{2\nu s_1s_2} \\ &= \cosh 2\nu, \end{aligned} \tag{4}$$

where $\nu = J/kT$ (k is Boltzmann's constant and T is the absolute temperature) and, in the thermodynamic limit, the free energy per spin ψ is given by

$$-\frac{\psi}{kT} = \lim_{N\to\infty} N^{-1}\log Z_N^I = 0 \tag{5}$$

and ψ is clearly an analytic function of T, i.e., the Ising model with interaction (3) does not have a phase transition.

The spherical-model partition function with interaction (3) is given by[2]

$$Z_N^S = A_N^{-1}\int\cdots\int_{\sum_{i=1}^{N} s_i^2 = N} e^{2\nu s_1s_2}\, ds_1\, ds_2\cdots ds_N, \tag{6}$$

where

$$A_N = 2\pi^{N/2}N^{(N-1)/2}(\Gamma(N/2))^{-1} \tag{7}$$

is the surface area of the N-dimensional sphere.

We now integrate first over the variables $s_3, s_4, \cdots, s_N$ in Eq. (6) as follows. Since

$$\delta(r^2 - a^2) = \delta(r - a)/2a$$

(δ denotes the Dirac delta function),

$$\begin{aligned} &A_N^{-1}\int\cdots\int_{r^2=\sum_{i=1}^{N} s_i^2 = N} e^{2\nu s_1s_2}\, ds_1\, ds_2\cdots ds_N \\ &= A_N^{-1}\int_{-\infty}^{\infty}\cdots\int e^{2\nu s_1s_2}\delta(r - \sqrt{N})\, ds_1\, ds_2\cdots ds_N \\ &= 2\sqrt{N}\, A_N^{-1}\int_{-\infty}^{\infty}\cdots\int e^{2\nu s_1s_2}\delta(r^2 - N)\, ds_1\, ds_2\cdots ds_N \\ &= 2\sqrt{N}\, A_N^{-1}\int_{-\infty}^{\infty}\cdots\int e^{2\nu s_1s_2} \\ &\quad\times \delta\Big(\sum_{i=3}^{N} s_i^2 - (N - s_1^2 - s_2^2)\Big)\, ds_1\, ds_2\cdots ds_N \\ &= \sqrt{N}\, A_N^{-1}\iint_{s_1^2+s_2^2\le N} ds_1\, ds_2 e^{2\nu s_1s_2}(N - s_1^2 - s_2^2)^{-\frac{1}{2}} \\ &\quad\times\int_{-\infty}^{\infty}\cdots\int\delta\Big(\Big(\sum_{i=3}^{N} s_i^2\Big)^{\frac{1}{2}} - (N - s_1^2 - s_2^2)^{\frac{1}{2}}\Big)\, ds_3\cdots ds_N, \end{aligned}$$

---

* Work supported by National Science Foundation Grant No. GP-9414.

† Permanent address: Northwestern University, Evanston, Illinois.

[1] M. Kac, *Brandeis Lectures, 1966* (Gordon and Breach, Science Publishers, New York, 1968).

[2] T. H. Berlin and M. Kac, Phys. Rev. **86**, 821 (1952).

and since

$$\int_{-\infty}^{\infty}\cdots\int \delta\left(\left(\sum_{i=1}^{n}x_i^2\right)^{\frac{1}{2}}-a\right)dx_1\,dx_2\cdots dx_n=\frac{2\pi^{n/2}a^{n-1}}{\Gamma(n/2)},$$

we have that

$$Z_N^S=\left(\frac{N-2}{2\pi}\right)N^{(1-N/2)}\times\iint_{s_1^2+s_2^2\le N}e^{2\nu s_1 s_2}[N-s_1^2-s_2^2]^{(N-4)/2}\,ds_1\,ds_2. \tag{8}$$

Transforming to polar coordinates then gives immediately

$$Z_N^S=\left(\frac{N-2}{2}\right)\int_0^1 I_0(N\nu y)(1-y)^{(N-4)/2}\,dy, \tag{9}$$

where

$$I_0(z)=\frac{1}{2\pi}\int_0^{2\pi}e^{z\sin\theta}\,d\theta \tag{10}$$

is the zero-order Bessel function with imaginary argument.

The integral in Eq. (9) for large N can be evaluated either by the method of steepest descents or, if one wishes to be more precise, by obtaining upper and lower bounds and then proceeding to the limit $N\to\infty$. We adopt the latter approach.

To obtain an upper bound for Z_N^S we use the elementary result

$$I_0(z)\le e^z, \quad \text{for} \quad z\ge 0, \tag{11}$$

which follows directly from (10). Substitution into (9) then gives

$$Z_N^S\le\frac{N-2}{2}\int_0^1 e^{N\nu y}(1-y)^{(N-4)/2}\,dy \le\left[\frac{N-2}{2}\right]\max_{0\le y\le 1}\{e^{N\nu y}(1-y)^{(N-4)/2}\} \tag{12}$$

from which it follows that

$$-\frac{\psi}{kT}=\lim_{N\to\infty}N^{-1}\log Z_N^S\le f(\nu), \tag{13}$$

where

$$f(\nu)=\tfrac{1}{2}(2\nu-1-\log 2\nu), \quad 2\nu>1, \qquad =0, \quad 2\nu\le 1. \tag{14}$$

To obtain a lower bound for Z_N^S we use the result

$$I_0(z)\ge(1/2z)(e^z-1), \quad \text{for} \quad z\ge 0, \tag{15}$$

which follows from (10) by reducing the range of integration to $(0,\pi/2)$, expanding the exponential, integrating term by term, and using

$$\frac{2}{\pi}\int_0^{\pi/2}\sin^p\theta\,d\theta\ge(p+1)^{-1}, \quad \text{for} \quad p\ge 0. \tag{16}$$

Substituting (15) into (9) gives

$$Z_N^S\ge\left(\frac{N-2}{2}\right)\int_0^1(1-y)^{(N-4)/2}\left[\frac{e^{N\nu y}-1}{2N\nu y}\right]dy \ge\left(\frac{N-2}{4N\nu}\right)\int_0^1(1-y)^{(N-4)/2}[e^{N\nu y}-1]\,dy =\left(\frac{N-2}{4N\nu}\right)\int_0^1(1-y)^{(N-4)/2}e^{N\nu y}dy-\frac{1}{2N\nu}. \tag{17}$$

Now, assuming for simplicity that N is even,

$$\int_0^1(1-y)^{(N-4)/2}e^{N\nu y}\,dy =\frac{[(N-4)/2]!\,e^{N\nu}}{(N\nu)^{(N/2)-1}}-\sum_{l=0}^{[(N-4)/2]}\frac{[(N-4)/2]!}{[(N-4)/2-l]!\,(N\nu)^{l+1}} \ge\frac{[(N-4)/2]!\,e^{N\nu}}{(N\nu)^{(N/2)-1}}-\frac{2}{N(2\nu-1)}, \quad \text{provided} \quad 2\nu>1, \tag{18}$$

and it follows from (17) and (18) that

$$-\psi/kT\ge f(\nu), \quad \text{provided} \quad 2\nu>1, \tag{19}$$

where $f(\nu)$ is defined by Eq. (14).

Also, since $I_0(z)\ge 1$,

$$Z_N^S\ge\left(\frac{N-2}{2}\right)\int_0^1(1-y)^{(N-4)/2}\,dy=1$$

and thus

$$-\psi/kT\ge 0, \quad \text{for all } \nu. \tag{20}$$

Equation (19) is true, therefore, for all ν. Combining this result with Eq. (13) then gives

$$-\psi/kT=f(\nu), \tag{21}$$

with $f(\nu)$ defined by Eq. (14), and it is obvious that $f(\nu)$, and hence ψ, is a nonanalytic function of temperature, i.e., the spherical model with interaction (3) has a phase transition at temperature T_c given by

$$2\nu_c=2J/kT_c=1. \tag{22}$$

SECTION III

It is perhaps of interest to note that the free energy of the "Curie–Weiss spherical model," i.e., with interaction energy

$$E=-\frac{J}{N}\left(\sum_{j=1}^{N}s_j\right)^2, \tag{23}$$

is precisely the same as the above Eqs. (21) and (14)! Thus, for the interaction (23),

$$Z_N=A_N^{-1}\int_{\sum_{i=1}^{N}s_i^2=N}\cdots\int\exp\left[\nu\left(N^{-\frac{1}{2}}\sum_{i=1}^{N}s_i\right)^2\right]ds_1\cdots ds_N. \tag{24}$$

Making the orthogonal change of variables to $y_1, y_2, \cdots, y_N$ with

$$y_1 = N^{-\frac{1}{2}} \sum_{i=1}^{N} s_i,$$

we have that

$$Z_N = A_N^{-1} \int \cdots \int_{\sum_{i=1}^{N} y_i^2 = N} e^{\nu y_1^2} \, dy_1 \cdots dy_N \tag{25}$$

and integration over $y_2, \cdots, y_N$ as before gives

$$\begin{aligned} Z_N &= \frac{2\Gamma(N/2)}{\sqrt{\pi}\, N^{(N/2)-\frac{1}{2}} \Gamma((N-1)/2)} \\ &\quad \times \int_{y_1^2 \le N} e^{\nu y_1^2} (N - y_1^2)^{(N-3)/2} \, dy_1 \\ &= \frac{2\Gamma(N/2)}{\sqrt{\pi}\, \Gamma((N-1)/2)} \int_0^1 e^{N\nu x^2} (1 - x^2)^{(N-3)/2} \, dx \end{aligned} \tag{26}$$

and it is obvious from the above analysis that

$$-\psi/kT = f(\nu) \quad \text{(Curie–Weiss, spherical).} \tag{27}$$

The difference now of course is that the Curie–Weiss Ising model has a phase transition. In particular, for the interaction (23) with $s_i = \pm 1$,

$$-\psi/kT = -\nu\eta^2 + \log\,[2 \cosh\,(2\nu\eta)] \quad \text{(Curie–Weiss, Ising),} \tag{28}$$

where η is a solution of

$$\eta = \tanh\,(2\nu\eta). \tag{29}$$

So, at least for this case, the Kac conjecture is true!

SECTION IV

It seems likely to us that, in general, the conjecture may be true if the interaction potential J_{ij} in Eq. (3) is periodic and ferromagnetic, i.e., if

$$J_{ij} = J(\mathbf{r}_i - \mathbf{r}_j) > 0, \tag{30}$$

where $\mathbf{r}_i$ is the position vector of the ith spin. In support of this "modified Kac conjecture" we have, in addition to the Curie–Weiss models, the following:

(i) the three-dimensional nearest neighbor Ising and spherical models both have phase transitions, and

(ii) the one-dimensional Ising[3] and spherical models[4] with interaction

$$J_{ij} = |i - j|^{-(p+1)}$$

have transitions when $0 < p < 1$ and do not have transitions when $p > 1$ (when $p = 1$, the spherical model does not have a transition and, at the moment, the Ising case is undecided).

We remark also that for the Curie–Weiss models, the magnetization M is given by

$$\begin{aligned} M &= \lim_{N\to\infty} \left\langle \left(\frac{1}{N} \sum_{i=1}^{N} s_i \right)^2 \right\rangle \\ &= \eta^2, \qquad \text{for Ising,} \\ &= 1 - 1/2\nu, \quad \text{for spherical} \quad (2\nu \ge 1), \end{aligned}$$

where η is defined by Eq. (29) and, since the slope of the line $y = \eta$ is greater than the slope of $y = \tanh 2\nu\eta$ where they intersect,

$$\eta^2 \gtrsim 1 - 1/2\nu, \quad \text{when} \quad 2\nu \ge 1. \tag{31}$$

(When $2\nu < 1$, both magnetizations are zero.)

It is then tempting to suggest that, in general, if

$$\begin{aligned} J_{ij} &= J(\mathbf{r}_i - \mathbf{r}_j) > 0, \\ M_{\text{Ising}} &\ge M_{\text{spherical}}, \end{aligned} \tag{32}$$

with the proviso that the inequality may only be true for sufficiently low temperatures.

In addition to being true for the Curie–Weiss models [Eq. (31)], it is true for the one-dimensional nearest-neighbor models.

In conclusion we remark that

$$M_{\text{Ising}} \ge M_{\text{continuum}}, \tag{33}$$

where "continuum" describes the model with $|s_i| \le 1$. To prove (33) write

$$\begin{aligned} \langle s_k s_l \rangle_{\text{continuum}} &= \int_{-1}^{+1} \cdots \int s_k s_l \exp\left(\nu \sum_{i<j} J_{ij} s_i s_j \right) ds_1 \cdots ds_N \Big/ \int_{-1}^{+1} \cdots \int \exp\left(\nu \sum_{i<j} J_{ij} s_i s_j \right) ds_1 \cdots ds_N \\ &= \int_0^1 \cdots \int s_k s_l \sum_{\mu_1=\pm 1} \cdots \sum_{\mu_N=\pm 1} \mu_k \mu_l \exp\left(\nu \sum_{i<j} (J_{ij} s_i s_j) \mu_i \mu_j \right) ds_1 \cdots ds_N \Big/ \\ &\qquad \int_0^1 \cdots \int \sum_{\mu_1=\pm 1} \cdots \sum_{\mu_N=\pm 1} \exp\left(\nu \sum_{i<j} (s_i J_{ij} s_j) \mu_i \mu_j \right) ds_1 \cdots ds_N \\ &\le \max_{0 \le s_i \le 1} \langle \mu_k \mu_l \rangle_{\text{Ising}}^{(s_1 s_2 \cdots s_N)}, \end{aligned}$$

[3] F. J. Dyson, "Existence of a Phase-Transition in a One-Dimensional Ising Ferromagnet" and "Non-Existence of Spontaneous Magnetization in a One-Dimensional Ising Ferromagnet," Commun. Math. Phys. (to be published).

[4] G. S. Joyce, Phys. Rev. **146**, 349 (1966).

where

$$\langle \mu_k \mu_l \rangle_{\text{Ising}}^{(s_1, s_2, \cdots, s_N)} = \sum_{\mu_1=\pm 1} \cdots \sum_{\mu_N=\pm 1} \mu_k \mu_l \exp\left[\nu \sum_{i<j} (s_i J_{ij} s_j) \mu_i \mu_j\right] \Big/ \sum_{\mu_1=\pm 1} \cdots \sum_{\mu_N=\pm 1} \exp\left[\nu \sum_{i<j} (s_i J_{ij} s_j) \mu_i \mu_j\right].$$

If we now appeal to a theorem of Griffiths,[5] which states that ferromagnetic Ising-model correlation functions $\langle \mu_i \mu_j \rangle$ are not decreased if the interaction between any two spins is increased, we see that the maximum of $\langle \mu_k \mu_l \rangle_{\text{Ising}}^{(s_1, s_2, \cdots, S_N)}$ is achieved when all $s_i = +1$. This is sufficient to prove (33) or, for that matter, the stronger result

$$\langle s_k s_l \rangle_{\text{continuum}} \leq \langle s_k s_l \rangle_{\text{Ising}}.$$

Notice that periodicity is *not* required in this case; the only requirement being that the interaction J_{ij} be ferromagnetic.

[5] R. B. Griffiths, J. Math. Phys. **8**, 478 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Dynamics in the Diagonal Coherent-State Representation*

MARVIN M. MILLER
Purdue University, Lafayette, Indiana

(Received 6 January 1969)

The problem of following the dynamical behavior of a quantum-mechanical system in the diagonal coherent-state representation is examined for those systems whose time evolution is specified by equations of motion for the coherent-state weight functional which resemble Fokker–Planck equations but have non-positive-definite diffusion matrices. A particular equation of this type describing a linear parametric process is considered in detail and several proposed generalizations of the diagonal representation, which include dynamical effects of the atomic system coupled to the electromagnetic field in simple models of a unimodal laser, are also briefly discussed.

I. INTRODUCTION

It has been noted recently by a number of authors[1–4] that the problem of following the time evolution of such quantum-mechanical systems as models of a unimodal laser and optical parametric amplifier in the diagonal representation[5,6] of the density operator with respect to the "coherent" states $|z\rangle$,[7]

$$\rho(t) = (1/\pi) \int P(z, t) \, |z\rangle\langle z| \, d^2 z, \quad z = (x + iy)/(2)^{\frac{1}{2}}, \tag{1}$$

can lead to mathematical difficulty because the weight function $P(z, t)$ is, in general, not a well-behaved classical function but a generalized function in a subspace of $Z'(R_2)$.[8] This difficulty manifests itself in the fact that the equation of motion for $P(z, t)$ in these cases resembles the Fokker–Planck equation familiar from the classical theory of stochastic processes,[9] but has a non-positive-definite diffusion

* Supported by the NASA Multidisciplinary Research Program under Grant No. 5522-52-12859.

[1] B. R. Mollow and R. J. Glauber, Phys. Rev. **160**, 1097 (1967); B. R. Mollow, *ibid.* **162**, 1256 (1967).
[2] J. R. Klauder, private communication (January, 1968).
[3] J. P. Gordon, Phys. Rev. **161**, 367 (1967).
[4] R. Graham, Z. Physik **210**, 319 (1968).
[5] E. C. G. Sudarshan, Phys. Rev. Letters **10**, 277 (1963).
[6] R. J. Glauber, Phys. Rev. **131**, 2766 (1963).
[7] R. J. Glauber, Phys. Rev. Letters **10**, 84 (1963).

[8] M. M. Miller and E. A. Mishkin, Phys. Rev. **164**, 1610 (1967). The fact that the diagonal weight function associated with the density operator actually belongs to a small subspace of $Z'(R_2)$ is stressed in Sec. V of this paper, and a characterization of this subspace $Z'_\rho(R_2)$ was given, based on the fact that the Fourier transform of the diagonal weight is a continuous function with a quadratic exponential bound. In terms of the classical generalized function spaces, the space $Z'_\rho(R_2)$ most closely corresponds to the space $S'^{\frac{1}{2}}(R_2)$. {For the definition of S spaces, see I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic Press Inc., New York, 1968), Vol. II; see also J. R. Klauder and E. C. G. Sudarshan [*Fundamentals of Quantum Optics* (W. A. Benjamin, Inc., New York, 1968), Chap. 8] for a comprehensive discussion of the diagonal coherent-state representation.} The fact that the diagonal weight function lies in a subspace of $Z'(R_2)$ has also been noted by K. E. Cahill, "Regularization of the *P*-Representation," Phys. Rev. (to be published). Cahill gives an alternate characterization of this subspace based upon the decomposition of the Fourier transform of the diagonal weight into the sum of a square-integrable function and an infinitely differentiable function. [The terminology "*P*-Representation," introduced by Glauber (Ref. 6), is frequently used in the literature to designate the diagonal coherent-state representation.]

[9] M. C. Wang and G. E. Uhlenbeck, Rev. Mod. Phys. **17**, 323 (1945).

where

$$\langle \mu_k \mu_l \rangle_{\text{Ising}}^{(s_1, s_2, \cdots, s_N)} = \sum_{\mu_1=\pm 1} \cdots \sum_{\mu_N=\pm 1} \mu_k \mu_l \exp\left[\nu \sum_{i<j} (s_i J_{ij} s_j) \mu_i \mu_j\right] \Big/ \sum_{\mu_1=\pm 1} \cdots \sum_{\mu_N=\pm 1} \exp\left[\nu \sum_{i<j} (s_i J_{ij} s_j) \mu_i \mu_j\right].$$

If we now appeal to a theorem of Griffiths,[5] which states that ferromagnetic Ising-model correlation functions $\langle \mu_i \mu_j \rangle$ are not decreased if the interaction between any two spins is increased, we see that the maximum of $\langle \mu_k \mu_l \rangle_{\text{Ising}}^{(s_1, s_2, \cdots, S_N)}$ is achieved when all $s_i = +1$. This is sufficient to prove (33) or, for that matter, the stronger result

$$\langle s_k s_l \rangle_{\text{continuum}} \leq \langle s_k s_l \rangle_{\text{Ising}}.$$

Notice that periodicity is *not* required in this case; the only requirement being that the interaction J_{ij} be ferromagnetic.

[5] R. B. Griffiths, J. Math. Phys. **8**, 478 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Dynamics in the Diagonal Coherent-State Representation*

MARVIN M. MILLER
Purdue University, Lafayette, Indiana

(Received 6 January 1969)

The problem of following the dynamical behavior of a quantum-mechanical system in the diagonal coherent-state representation is examined for those systems whose time evolution is specified by equations of motion for the coherent-state weight functional which resemble Fokker–Planck equations but have non-positive-definite diffusion matrices. A particular equation of this type describing a linear parametric process is considered in detail and several proposed generalizations of the diagonal representation, which include dynamical effects of the atomic system coupled to the electromagnetic field in simple models of a unimodal laser, are also briefly discussed.

I. INTRODUCTION

It has been noted recently by a number of authors[1–4] that the problem of following the time evolution of such quantum-mechanical systems as models of a unimodal laser and optical parametric amplifier in the diagonal representation[5,6] of the density operator with respect to the "coherent" states $|z\rangle$,[7]

$$\rho(t) = (1/\pi) \int P(z, t) \, |z\rangle\langle z| \, d^2 z, \quad z = (x + iy)/(2)^{\frac{1}{2}}, \tag{1}$$

can lead to mathematical difficulty because the weight function $P(z, t)$ is, in general, not a well-behaved classical function but a generalized function in a subspace of $Z'(R_2)$.[8] This difficulty manifests itself in the fact that the equation of motion for $P(z, t)$ in these cases resembles the Fokker–Planck equation familiar from the classical theory of stochastic processes,[9] but has a non-positive-definite diffusion

* Supported by the NASA Multidisciplinary Research Program under Grant No. 5522-52-12859.

[1] B. R. Mollow and R. J. Glauber, Phys. Rev. **160**, 1097 (1967); B. R. Mollow, *ibid.* **162**, 1256 (1967).
[2] J. R. Klauder, private communication (January, 1968).
[3] J. P. Gordon, Phys. Rev. **161**, 367 (1967).
[4] R. Graham, Z. Physik **210**, 319 (1968).
[5] E. C. G. Sudarshan, Phys. Rev. Letters **10**, 277 (1963).
[6] R. J. Glauber, Phys. Rev. **131**, 2766 (1963).
[7] R. J. Glauber, Phys. Rev. Letters **10**, 84 (1963).

[8] M. M. Miller and E. A. Mishkin, Phys. Rev. **164**, 1610 (1967). The fact that the diagonal weight function associated with the density operator actually belongs to a small subspace of $Z'(R_2)$ is stressed in Sec. V of this paper, and a characterization of this subspace $Z'_\rho(R_2)$ was given, based on the fact that the Fourier transform of the diagonal weight is a continuous function with a quadratic exponential bound. In terms of the classical generalized function spaces, the space $Z'_\rho(R_2)$ most closely corresponds to the space $S'^{\frac{1}{2}}(R_2)$. {For the definition of S spaces, see I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic Press Inc., New York, 1968), Vol. II; see also J. R. Klauder and E. C. G. Sudarshan [*Fundamentals of Quantum Optics* (W. A. Benjamin, Inc., New York, 1968), Chap. 8] for a comprehensive discussion of the diagonal coherent-state representation.} The fact that the diagonal weight function lies in a subspace of $Z'(R_2)$ has also been noted by K. E. Cahill, "Regularization of the *P*-Representation," Phys. Rev. (to be published). Cahill gives an alternate characterization of this subspace based upon the decomposition of the Fourier transform of the diagonal weight into the sum of a square-integrable function and an infinitely differentiable function. [The terminology "*P*-Representation," introduced by Glauber (Ref. 6), is frequently used in the literature to designate the diagonal coherent-state representation.]

[9] M. C. Wang and G. E. Uhlenbeck, Rev. Mod. Phys. **17**, 323 (1945).

matrix. They are, therefore, not of the parabolic type[10] and the general question of the existence and uniqueness of solutions to such equations when the drift and diffusion coefficients are variable has not been resolved.[11] If, on the other hand, the drift and diffusion coefficients are constant, then the "Fokker–Planck-like," non-positive-definite diffusion equations —hereafter referred to as FPL equations—belong to the class of "incorrectly posed" partial differential equations which have been studied intensively during the past ten years.[10] A characteristic feature of such equations is the existence of a time interval τ during which the solution is a well-behaved classical function, but after which the solution exists only in the sense of a generalized function. (A simple example of such an equation is the heat equation for negative times.[12]) For "incorrectly posed" equations to have classical solutions for a nonzero time interval, the initial data must be an entire analytic function for complex values of the spatial variables; then the magnitude of τ depends on the growth of the initial data in the complex plane.[13] Although an analogous theory has not yet been developed for FPL equations with variable coefficients, the simple, variable coefficient equation considered in Sec. II does exhibit such behavior. However, even this example illustrates that variable coefficient PFL equations are much more subtle and it is presently not known how to prove existence and uniqueness and obtain an estimate for τ in the general case.[11]

In this context, it is important to note that in the case of the diagonal weight functional $P(z, t)$ it can be proved directly from the definition (1) that $P(z, t)$ exists and is unique.[8] Hence, proofs of existence and uniqueness derived directly from the differential equation for $P(z, t)$, although interesting from a mathematical viewpoint, are really "frosting on the cake." From a physical point of view, what is needed is a solution for $P(z, t)$ which would enable one to calculate the normally ordered correlation functions characterizing the coherence properties of the electromagnetic field[14] using simple c-number operations. In the sense that no general prescription is presently available for calculating such averages when the time evolution of $P(z, t)$ is specified by a FPL equation, several authors[15] have concluded that no "reasonable" or "useful" diagonal representation exists in these cases, and have advocated following the time evolution of the density operator $\rho(t)$ using other "quasiprobability" weight functions which are "better behaved" mathematically such as the Wigner distribution[16] or the diagonal coherent-state matrix element of the density operator.[17,18] While these alternate procedures for calculating correlation functions are, of course, valid, we believe that it is still of interest from a physical as well as a mathematical point of view to study the time evolution of the density operator in the diagonal coherent-state representation even in those cases in which $P(z, t)$ is a generalized function and bears only a formal relationship to a classical probability density function. In Sec. II, we discuss in detail the model of a linear parametric process previously considered by Mollow[1] and Klauder,[2] and exhibit solutions for $P(z, t)$ for all times. In Sec. III, we consider the problem of following the time evolution of the interaction between an atomic source and its radiation field by means of a generalization of the diagonal weight function which includes atomic as well as field variables.

II. TIME EVOLUTION OF A PARAMETRIC PROCESS

We consider a single-mode field governed by the parametric Schrödinger Hamiltonian

$$H = \omega a^\dagger a + (a^\dagger e^{-i\omega t}\sigma e^{-i\omega t} a^\dagger + \text{H.c.}), \qquad (2)$$

where a, $a^\dagger$ are the usual boson annihilation and creation operators, and ω, σ are c numbers which specify the angular frequency of the mode and the strength of the parametric interaction.[19] The interaction Hamiltonian in the interaction representation is, therefore,

$$H_I(t) = H_I(0) = e^{i\omega a^\dagger a t}\{a^\dagger e^{-i\omega t}\sigma e^{-i\omega t} a^\dagger + \text{H.c.}\}e^{-i\omega a^\dagger a t}$$
$$= \sigma a^{\dagger 2} + \sigma^* a^2 \qquad (3)$$

[10] I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic Press Inc., New York, 1967), Vol. III, Chap. III.

[11] This question is currently being investigated by F. Treves and S. Steinberg (private communication, November, 1968). See F. Treves, *Ovciannikov Theorem and Hyperdifferential Operators* (Instituto Matetica Pura ae Aplicada, Rio de Janeiro, 1969).

[12] F. John, *Partial Differential Equations* (Interscience Publishers, Inc., New York, 1964), p. 102.

[13] Reference 10, p. 163, Theorem 3.

[14] R. J. Glauber, Phys. Rev. **130**, 2529 (1963).

[15] See, for example, R. J. Glauber, *Proceedings of the Symposium on Modern Optics* (Interscience Publishers, Inc., New York, 1967); R. Graham *et al.*, Z. Physik **213**, 21 (1968).

[16] E. P. Wigner, Phys. Rev. **40**, 749 (1932).

[17] C. L. Mehta and E. C. G. Sudarshan, Phys. Rev. **138**, B274 (1965).

[18] R. J. Glauber, *Quantum Optics and Electronics* (*Les Houches, 1964*); (Gordon and Breach, Science Publishers, New York, 1965).

[19] Since an actual parametric process involves a coupling between two (or more) harmonic-oscillator modes, a and $a^\dagger$ in Eq. (2) should be interpreted as n-component row and column vectors and ω and σ as real symmetric $n \times n$ matrices in the general case of an interaction between n-coupled oscillators. However, the resulting Hamiltonian can be converted into a sum of n terms of the form of Eq. (1) by diagonalizing the matrix σ (the matrix ω is already diagonal), so that there is no loss in generality in considering the single-mode case. (The n-mode problem is discussed in detail in the paper by Mollow in Ref. 1.)

and the time evolution of the interaction-representation density operator $\rho_I(t)$ is specified by the equation

$$i\frac{\partial\rho_I(t)}{\partial t} = [H_I(t), \rho_I(t)]. \tag{4}$$

We can transcribe this equation for the density operator into an equation of motion for the diagonal weight function $P(z,t)$ by combining Eqs. (1), (3), and (4),

$$\begin{aligned} i\frac{\partial\rho_I(t)}{\partial t} &= \frac{i}{\pi}\int\frac{\partial P(z,t)}{\partial t}\,|z\rangle\langle z|\,d^2z \\ &= \frac{1}{\pi}\int P(z,t)[H_I(t), |z\rangle\langle z|]\,d^2z \\ &= \frac{i}{\pi}\int P(z,t)[(\sigma a^{\dagger 2} + \sigma^* a^2), |z\rangle\langle z|]\,d^2z, \end{aligned} \tag{5}$$

and making use of the differential-operator representation of the creation and annihilation operators[20] and two partial integrations to rewrite Eq. (5) in the form

$$\begin{aligned} &\frac{1}{\pi}\int\frac{\partial P(z,t)}{\partial t}\,|z\rangle\langle z|\,d^2z \\ &\quad = \frac{1}{\pi}\int\left\{\left(\sigma\frac{\partial^2}{\partial z^2} - 2\sigma z^*\frac{\partial}{\partial z}\right) - \text{c.c.}\right\}P(z,t)\,|z\rangle\langle z|\,d^2z. \end{aligned} \tag{6}$$

In terms of the real variables x, y and with the choice $\sigma = 1/2i$, the resulting equation for $P(z,t)$ is

$$\begin{aligned} &\frac{\partial P(x,y,t)}{\partial t} \\ &\quad = \left(x\frac{\partial}{\partial x} - \frac{1}{2}\frac{\partial^2}{\partial x^2} - y\frac{\partial}{\partial y} + \frac{1}{2}\frac{\partial^2}{\partial y^2}\right)P(x,y,t), \end{aligned} \tag{7}$$

which has the formal solution

$$\begin{aligned} &P(x,y,t) \\ &= \exp\left[t\left(x\frac{\partial}{\partial x} - \frac{1}{2}\frac{\partial^2}{\partial x^2} - y\frac{\partial}{\partial y} + \frac{1}{2}\frac{\partial^2}{\partial y^2}\right)\right]P(x,y,0). \end{aligned} \tag{8}$$

If we assume that the mode is initially in a chaotic state, then[6]

$$P(x,y,0) = \frac{1}{2\langle N\rangle}\exp\left\{-\frac{x^2+y^2}{2\langle N\rangle}\right\}, \tag{9}$$

where $\langle N\rangle$ is the average number of quanta at $t = 0$. With this choice for $P(x,y,0)$ the solution can be written in the form

$$P(x,y,t) = f(x,t)g(y,t),$$

where

$$f(x,t) = \exp\left[t\left(x\frac{\partial}{\partial x} - \frac{1}{2}\frac{\partial^2}{\partial x^2}\right)\right]\frac{\exp\{-x^2/2\langle N\rangle\}}{[2\langle N\rangle]^{\frac{1}{2}}}, \tag{10a}$$

$$g(y,t) = \exp\left[t\left(-y\frac{\partial}{\partial y} + \frac{1}{2}\frac{\partial^2}{\partial y^2}\right)\right]\frac{\exp\{-y^2/2\langle N\rangle\}}{[2\langle N\rangle]^{\frac{1}{2}}}. \tag{10b}$$

To solve these equations, we note that the exponential in Eq. (10a) is of the form e^{A+B} where

$$A = tx\frac{\partial}{\partial x}, \quad B = -\frac{t}{2}\frac{\partial^2}{\partial x^2}. \tag{11}$$

Since the commutator product $[A,B] \doteq AB - BA$ is proportional to B,

$$[A,B] = +\frac{\partial^2}{\partial x^2}t^2 = -2tB, \tag{12}$$

the Zassenhaus formula[21] reduces to

$$e^{A+B} = e^A e^{f(t)B}, \quad f(t) = \frac{(e^{2t}-1)}{2t}. \tag{13}$$

Hence,

$$\begin{aligned} f(x,t) &= \exp\left\{tx\frac{\partial}{\partial x}\right\} \\ &\quad\times\exp\left\{\frac{(1-e^{2t})}{4}\frac{\partial^2}{\partial x^2}\right\}\frac{\exp\{-x^2/2\langle N\rangle\}}{[2\langle N\rangle]^{\frac{1}{2}}}. \end{aligned} \tag{14}$$

The right-hand side of Eq. (14) may be evaluated by first utilizing the identity

$$\begin{aligned} &\exp\left\{\frac{(1-e^{2t})}{4}\frac{\partial^2}{\partial x^2}\right\}\frac{\exp\{-x^2/2\langle N\rangle\}}{[2\langle N\rangle]^{\frac{1}{2}}} \\ &\quad = \exp\left\{\frac{(1-e^{2t})}{2}\frac{\partial}{\partial\langle N\rangle}\right\}\frac{\exp\{-x^2/2\langle N\rangle\}}{[2\langle N\rangle]^{\frac{1}{2}}} \end{aligned} \tag{15}$$

and the Taylor-shift formula to rewrite Eq. (14) in the form

$$\begin{aligned} f(x,t) &= \exp\left\{tx\frac{\partial}{\partial x}\right\} \\ &\quad\times\frac{\left[\exp\left(\dfrac{-x^2}{[2\langle N\rangle + (1-e^{2t})]}\right)\right]}{[2\langle N\rangle + (1-e^{2t})]^{\frac{1}{2}}}. \end{aligned} \tag{16}$$

By applying the Taylor multiplication formula[22]

$$\exp\left(tx\frac{\partial}{\partial x}\right)\psi(x) = \psi(xe^t), \tag{17}$$

[20] See, for example, p. 127 of the book by Klauder and Sudarshan in Ref. 8.

[21] W. Magnus, Commun. Pure Appl. Math. 7, 649 (1954).

[22] The function $\psi(x)$ in Eq. (17) represents an arbitrary function of x.

we obtain the solution

$$f(x,t) = \frac{1}{e^{t}[(2\langle N\rangle + 1)e^{-2t} - 1]^{\frac{1}{2}}} \times \exp\left\{\frac{-x^2}{[(2\langle N\rangle + 1)e^{-2t} - 1]}\right\}. \quad (18)$$

The right-hand side of Eq. (10b) may be evaluated in a similar fashion with the result

$$g(y,t) = \frac{1}{e^{-t}[(2\langle N\rangle + 1)e^{+2t} - 1]^{\frac{1}{2}}} \times \exp\left\{\frac{-y^2}{[(2\langle N\rangle + 1)e^{2t} - 1]}\right\}. \quad (19)$$

Therefore,

$$P(x,y,t) = \frac{1}{\{[(2\langle N\rangle + 1)e^{-2t} - 1][(2\langle N\rangle + 1)e^{+2t} - 1]\}^{\frac{1}{2}}} \times \exp\left\{\frac{-x^2}{[(2\langle N\rangle + 1)e^{-2t} - 1]} + \frac{-y^2}{[(2\langle N\rangle + 1)e^{+2t} - 1]}\right\}. \quad (20)$$

We note the following important points about the solution for $P(x,y,t)$ given by Eq. (20):

(1) Since the density operator $\rho_I(t)$ must satisfy the constraint

$$\mathrm{Tr}\,\rho_I(t) = \frac{1}{\pi}\int P(z,t)\,d^2z = 1, \quad (21)$$

this solution is only valid for times t such that

$$(2\langle N\rangle + 1)e^{-2t} > 1 \quad (22)$$

or $0 < t < \ln(2N+1)^{\frac{1}{2}} = \tau$ because, for times $t > \tau$, $f(x,t)$ diverges as x becomes infinite.

(2) The solution for $t < \tau$ can also be obtained[1,2] by solving the equation of motion for the Fourier transform of $P(z,t)$, the normally ordered characteristic function $X_N(\gamma,t)$, defined by[17,23]

$$X_N(\gamma,t) = \mathrm{Tr}\,\rho e^{\gamma a^\dagger(t)}e^{-\gamma^\dagger a(t)}, \quad (23)$$

and then taking the inverse transform. Since the Heisenberg equations of motion for $a(t)$, $a^\dagger(t)$ are linear, the right-hand side of Eq. (23) may be easily evaluated with the result

$$X_N(u,v,t) = \exp\left[-\tfrac{1}{4}\{u^2[(2\langle N\rangle + 1)e^{-2t} - 1] + v^2[(2\langle N\rangle + 1)e^{2t} - 1]\}\right], \quad \gamma = (u+iv)/(2)^{\frac{1}{2}}. \quad (24)$$

In contrast to Eq. (20), this solution for $X_N(u,v,t)$ is valid for *all* times. Indeed it is easily shown that $X_N(u,v,t)$ is a continuous *function* with the exponential bound

$$|X_N(u,v,t)| \le \exp\left[\tfrac{1}{4}(u^2+v^2)\right]. \quad (25)$$

Thus, for $t > \tau$, the solution for $X_N(u,v,t)$ is

$$X_N(u,v,t) = \exp\{+\tfrac{1}{4}u^2[1 - (2\langle N\rangle + 1)e^{-2t}] - \tfrac{1}{4}v^2[(2\langle N\rangle + 1)e^{+2t} - 1]\}, \quad (26)$$

which does not have a Fourier transform in the classical sense because of the exponentially increasing functional form of u.

(3) The time interval during which the solution for $P(z,t)$ exists as a classical function depends strongly on the initial data, i.e., we have seen that with the choice

$$P(x,y,0) = \{\exp[-(x^2+y^2)/2\langle N\rangle]\}/2\langle N\rangle$$

a classical solution exists for the finite time interval τ. If, on the other hand, $P(x,y,0) = \pi\delta(x)\delta(y)$, corresponding to the mode initially in the vacuum state, a classical solution does not exist for any nonzero time interval, while if $P(x,y,0)$ decreases faster than the exponential of any negative-definite quadratic form in x as $x \to \infty$, a classical solution exists for all time.[1] In this regard, it is important to note that not all rapidly decreasing functions are admissible as valid initial data for $P(z,t)$ or $X_N(\gamma,t)$. For example, the function $\exp(-|\gamma|^4)$ is not an admissible normally-ordered characteristic function since it characterizes a state for which $\langle a^\dagger a\rangle = 0$, and this state, the vacuum, is uniquely specified by $X_N(\gamma,t) = 1$.[24] (This restriction on the class of admissible characteristic functions is well known in the context of classical probability theory. See, for example, Ref. 25.)

(4) The technique used to solve Eq. (10) may also be used to obtain an estimate of the time interval during which a classical solution exists even in the case of more general FPL equations for $P(z,t)$,

$$\frac{\partial P(z,t)}{\partial t} = L\left(z, \frac{\partial}{\partial z}\right)P(z,t), \quad (27)$$

when it is impossible to obtain an explicit solution because of the form of the differential operator $L(z, \partial/\partial z)$ and/or the initial data $P(z,0)$. That is,

[23] R. J. Glauber, *Physics of Quantum Electronics* (McGraw-Hill Book Co., Inc., 1966), p. 788.

[24] I wish to thank B. Mollow for a discussion concerning this question.

[25] E. Lukacs, *Characteristic Functions* (Charles Griffin, London, (1960), p. 59.

since the formal solution of Eq. (27) has the infinite-series representation

$$P(z,t) = \exp tL\left(z, \frac{\partial}{\partial z}\right) P(z,0)$$
$$= \sum_0^\infty \frac{t^n}{n!} L^{(n)}\left(z, \frac{\partial}{\partial z}\right) P(z,0), \qquad (28)$$

an estimate of the "classical time interval" τ can be obtained from the convergence of the infinite series if it is possible to estimate the effect of the nth-order differential operator $L^{(n)}(z, \partial/\partial z)$ on the analytic initial data $P(z,0)$.

We next consider the solution of Eq. (7) in the time interval $t > \tau$. This solution follows immediately from the form of Eq. (26) for $X_N(u,v,t)$ and the fact that the Fourier transform of $e^{u^2/4a}$, $a > 0$, is the analytic functional corresponding to the function $i(2a)^{\frac{1}{2}}e^{as^2}$ and the contour $(-i\infty, +i\infty)$ in the complex s plane.[26] Hence, the solution of Eq. (7) for $t > \tau$ with initial data given by Eq. (8) is the analytic functional of two complex variables s_1, s_2, corresponding to the function

$$P(s_1, s_2, t)$$
$$= i\{[1 - (2\langle N\rangle + 1)e^{-2t}][(2\langle N\rangle + 1)e^{2t} - 1]\}^{-\frac{1}{2}}$$
$$\times \exp\left\{\frac{s_1^2}{[1 - (2\langle N\rangle + 1)e^{-2t}]} + \frac{-s_2^2}{[(2\langle N\rangle + 1)e^{2t} - 1]}\right\} \qquad (29)$$

and the hypersurface Γ defined by

$$S_1{:}(-i\infty, +i\infty), \quad S_2{:}(-\infty, +\infty). \qquad (30)$$

It is easy to verify that this representation leads to the same normally ordered moments as are obtained, for example, by differentiating $X_N(u,v,t)$. For example, the mean number of quanta $\langle a^\dagger(t)a(t)\rangle$ for times $t > \tau$ is given by

$$\langle a^\dagger(t)a(t)\rangle = \frac{1}{\pi}\iint_\Gamma P(s_1, s_2, t)\psi(s_1, s_2)\, ds_1\, ds_2, \qquad (31)$$

where the test function $\psi(s_1, s_2)$ is obtained via the prescription

$$s_1 \to ix/(2)^{\frac{1}{2}}, \quad s_2 \to y/(2)^{\frac{1}{2}}, \quad \psi(s_1, s_2) \to \psi(x, y).$$

In this case,

$$\psi(x,y) = \langle z|\, a^\dagger a\, |z\rangle = \tfrac{1}{2}(x^2 + y^2); \qquad (32)$$

hence,

$$\psi(s_1, s_2) = (-s_1^2 + s_2^2). \qquad (33)$$

Substituting (33) and (29) into (31), we obtain

$$\langle a^\dagger(t)a(t)\rangle = \langle N\rangle \cosh 2t + \sinh^2 t. \qquad (34)$$

Thus, if it is possible to obtain an analytic functional representation for the diagonal weight, it can be used to obtain expectation values of interest in a straightforward manner. However, aside from the type of FPL equation considered in this section, i.e., those with linear drift and constant diffusion coefficients, it is presently not known how to obtain explicit solutions in this form. To overcome this difficulty, an alternate procedure for evaluating expectation values has been advocated by Klauder.[27] This method, which obviates the necessity of working with generalized functions inherent in the direct approach, is based on the observation that the derivative of a generalized function is usually defined by its reciprocal derivative action on an appropriate test function space. Therefore, to find the mean value of an operator $O(a, a^\dagger, t)$ at time t, we may either follow the time evolution of the density operator via the equation of motion for $P(z,t)$, Eq. (27), or follow the time evolution of the function $O(z,t) = \langle z|\, O\, |z\rangle$ via the equation

$$\frac{\partial O(z,t)}{\partial t} = L^\dagger\left(z, \frac{\partial}{\partial z}\right) O(z,t), \qquad (35)$$

where $L^\dagger$ is the Hermitian adjoint of L. That is,

$$\langle O(a, a^\dagger, t)\rangle = \frac{1}{\pi}\int P(z,t)\langle z|\, O(a^\dagger, a, 0)\, |z\rangle\, d^2z$$
$$= \int e^{tL}P(z,0)O(z,0)\, d^2z$$
$$= \int P(z,0)e^{tL^\dagger}O(z,0)\, d^2z$$
$$= \int P(z,0)O(z,t)\, d^2z. \qquad (36)$$

Although conceptually it may be more satisfying to calculate $P(z,t)$ and then use it to compute the mean value of *any* operator (Schrödinger picture), the alternate technique of following the time evolution of a particular mean value (Heisenberg picture) via an equation of the form of Eq. (35) has the virtue of avoiding the generalized function $P(z,t)$ in favor of the well-behaved classical function $O(z,t)$. Applications of this technique are currently under investigation.

[26] I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic Press Inc., New York, 1965), Vol. I, p. 188.

[27] J. R. Klauder (private communication, October, 1968).

III. GENERALIZATIONS OF THE DIAGONAL COHERENT-STATE REPRESENTATION

Recently several authors[3,28,29] have considered generalizations of the diagonal coherent-state representation to include dynamical effects of the atomic system coupled to the electromagnetic field in simple models of a unimodal laser. In previous work,[30] Fokker–Planck equations describing the statistical state of the field alone were derived by adiabatically eliminating the atomic variables from the stochastically equivalent Langevin equations.[31] Both Haken[28] and Lax[29] derive a stochastic differential equation, i.e., one containing derivatives higher than the second order, for a generalized weight function including atomic variables and also obtain the Fokker–Planck approximation to the exact equation. Their weight function is defined as the Fourier transform of a characteristic function which is an atoms–field generalization of the normally ordered characteristic function defined by Eq. (23). For example, for a two-level laser described by the five variables: a, $a^\dagger$ (field), M, $M^\dagger$ (atomic polarization), and $[N_2 - N_1]$ (population inversion), it is natural to define the generalized diagonal weight function P as the Fourier transform of the characteristic function[28]

$$F(\gamma, \lambda, \varphi, t) = \mathrm{Tr}\,\{\rho(t)e^{i\lambda^* M}e^{i\varphi[N_2-N_1]}e^{i\lambda M^\dagger}e^{i\gamma a^\dagger}e^{i\gamma^* a}\}. \quad (37)$$

If we let

$$A = e^{L\lambda^* M}e^{i\varphi[N_2-N_1]}e^{i\lambda M^\dagger}, \quad (38)$$

$$B = e^{i\gamma a^\dagger}e^{i\gamma^* a} = e^{|\gamma|^2/2}e^{i(\gamma a^\dagger+\gamma^* a)}, \quad (39)$$

it follows from the inequality

$$|\mathrm{Tr}\,(\rho AB)|^2 \leq \mathrm{Tr}\,\rho(A^\dagger A)\,\mathrm{Tr}\,\rho(B^\dagger B) \quad (40)$$

and the fact that the operator B is traceable (since it is defined in the finite-dimensional Hilbert space of the atomic system) that the characteristic function exists. Because of the q-number nature of the atomic communators,[29] we have been unable to establish a bound on F in terms of the atomic variables analogous to the exponential bound and therefore cannot characterize precisely the function space to which F belongs. However, in general, it is clear that the inverse of F, the diagonal weight P, is a generalized function in a subspace of Z'.

In contrast to the above procedure, Gordon's generalization[3] of the diagonal coherent-state representation is a direct extension of the definition Eq. (1) that makes explicit use of the differential operator representation of the atomic variables. Using the laser model of Lax,[30] he obtains a FPL equation for his diagonal weight. In addition to the mathematical difficulty due to non-positive-definite diffusion matrix, some of the drift coefficients are quadratic, and some of the diffusion coefficients are linear in the atom–field variables. For these reasons, no explicit solutions have been obtained and the adjoint approach (see Sec. II) seems to be the only method presently available for distilling information from this equation. In this regard, it is important to note that this method does not lead to an increase in the order of the differential operator, i.e., $L^\dagger$ in Eq. (35) is of the same order as L, whereas exact equations for such well-behaved distribution functions as the generalized Wigner density[3] are of infinite order and can only be handled in a truncated diffusion approximation.

ACKNOWLEDGMENTS

The author is grateful to Dr. J. R. Klauder for stimulating his interest in this problem, and to him and Professor B. R. Mollow for helpful discussions.

[28] H. Haken, H. Risken, and W. Weidlich, Z. Physik **206**, 355 (1967).

[29] M. Lax and H. Yuen, Phys. Rev. **172**, 363 (1968).

[30] See, for example, M. Lax, Phys. Rev. **145**, 110 (1966).

[31] The "stochastic equivalence" between the Langevin equations and the Fokker–Planck equation is discussed, for example, by R. L. Stratonovitch, *Topics in the Theory of Random Noise* (Gordon and Breach, Science Publishers, New York, 1963), Vol. I, p. 102. Stratonovitch's proof that it is always possible to go from the Fokker–Planck equation to a set of equivalent Langevin equations is based on the fact that the diffusion matrix of a true Fokker–Planck equation is positive definite. Hence, the equivalence is open to question in the case of the FPL equations we have been discussing. In this regard, it is interesting to note that Gordon adiabatically eliminates atomic variables directly from his basic FPL equation [see Ref. 3, Eq. (3.19)] in contrast to the usual procedure.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Periodic Small-Amplitude Solutions to Volterra's Problem of Two Conflicting Populations and Their Application to the Plasma Continuity Equations

J. REECE ROTH
Lewis Research Center, National Aeronautics and Space Administration, Cleveland, Ohio

(Received 7 March 1969)

The coupled set of first-order nonlinear differential equations describing a generalized form of Volterra's problem of two conflicting populations

$$\dot{x} = C_0 + C_1x + C_2y + C_3xy + C_4x^2 + C_5y^2,$$
$$\dot{y} = A_0 + A_1x + A_2y + A_3xy + A_4x^2 + A_5y^2$$

are solved by an approximate method which gives $y(t)$ for the particular case in which the variables x and y vary periodically, the coefficients C_i and A_i are real, and the peak-to-peak amplitude of x is small compared with the mean value of x. The peak-to-peak amplitude of y, however, is not necessarily small compared with the mean value of y. When these conditions are satisfied, the functional form of $y(t)$ is approximated by Jacobian elliptic functions. The solutions obtained in this analysis are relevant to special cases of the classical problem of predator and prey, and also to certain low-frequency oscillations in partially ionized plasmas that arise from periodic solutions to the neutral and charged-particle continuity equations.

INTRODUCTION

This paper will be devoted to the following coupled set of first-order nonlinear differential equations:

$$\dot{x} = C_0 + C_1x + C_2y + C_3xy + C_4x^2 + C_5y^2, \quad (1)$$

$$\dot{y} = A_0 + A_1x + A_2y + A_3xy + A_4x^2 + A_5y^2, \quad (2)$$

which are a generalized form of Volterra's problem of two conflicting populations.[1–9] A less general form of these equations was originally analyzed by Lotka in an attempt to formulate a mathematical theory of the behavior of two conflicting species in a state of nature[1,2] and of certain oscillatory chemical reactions.[3] This work was later refined by Volterra,[4,5] whose work is summarized by Davis.[6]

Most of the literature on this problem is concerned with determining the conditions under which the population of one or both species will remain bounded, or under which periodic fluctuations of the populations will occur.[6–9] Little emphasis has been placed on determining the peak-to-peak amplitude, period, and waveform of the solutions to Eqs. (1) and (2) because of the difficulty of the exact nonlinear problem. Analytical solutions have been obtained for the linearized case in which the peak-to-peak amplitude of *both* x and y are small by comparison with their mean values.[3,5] The waveforms are sinusoidal in this case. In the present analysis, analytical solutions to Eqs. (1) and (2) were obtained under the assumption that the fluctuations of only *one* of the two variables are small by comparison with its mean value.

Equations (1) and (2) can be derived in a plasma physics context by writing the continuity equations for each of the three components (ions, electrons, and neutrals) of a partially ionized gas and assuming that spatial variations are small over the region of interest. One of the three equations may be eliminated by assuming that the Debye distance is small compared with the apparatus dimensions, so that the ion and electron densities are approximately equal at each point. There will remain (in the absence of three-body processes) two equations similar in form to Eqs. (1) and (2).[10–12] The variable x might represent the number density of neutral particles and y the number density of electrons, or vice versa. It has been demonstrated experimentally[11,12] that the periodic solutions to Eqs. (1) and (2) can be identified with a previously unrecognized mode of plasma oscillation.

ANALYSIS

It is an easy matter to put Eqs. (1) and (2) on a computer and obtain numerical solutions for a

[1] A. J. Lotka, Proc. Natl. Acad. Sci., US **6**, 410 (1920), pp. 410–415
[2] A. J. Lotka, *Elements of Physical Biology* (Williams and Wilkins Co., Baltimore, Md., 1925), Chap. 6.
[3] A. J. Lotka, J. Am. Chem. Soc. **42**, 1595 (1920).
[4] V. Volterra, *Lecons sur la theorie mathematique de la lutte pour la vie* (Gauthier-Villars et Cie., Paris, 1931).
[5] V. Volterra, J. Conseil Perm. Intern. Exploration Mer **3**, 1 (1928).
[6] H. T. Davis, *Introduction to Nonlinear Differential and Integral Equations* (Dover Publications, Inc., New York, 1962), Chap. 6.
[7] M. Frommer, Math. Ann. **109**, 395 (1934).
[8] R. D. Pittle and Thomas J. Higgins, J. Franklin Inst. **282**, 291 (1966).
[9] R. D. Pittle, "A Detailed Study of the Nonlinear Systems Characterized by Volterra's Differential Equations of Growth," MS thesis, Dept. of Electrical Eng., University of Wisconsin, 1965.
[10] J. R. Roth, Phys. Fluids **10**, 2712 (1967).
[11] J. R. Roth, Phys. Fluids **12**, 260 (1969).
[12] J. R. Roth, "Experimental Observation of Oscillations Described by The Continuity Equations of Slightly Ionized Deuterium, Neon, and Helium Gas," NASA Technical Note TND-4950, 1968.

particular set of C_i and A_i. However, it is desirable to have a closed-form solution for $x(t)$ and $y(t)$ in terms of the coefficients C_i and A_i of the differential equations, in order to compare experiment with the predictions of the mathematical model. The most obvious approach to obtaining a closed-form solution is to linearize these equations through a "double-perturbation" analysis, in which both x and y are assumed to have constant mean values x_1 and y_{00}. The time-varying portions of x and y [$x_2(t)$ and $y_{01}(t)$] are assumed to be much smaller than their respective mean values, so that $x_1 \gg x_2(t)$ and $y_{00} \gg y_{01}(t)$. This method was used by Lotka[3] and by Volterra [5] in their analyses of a less general form of Eqs. (1) and (2), and yields exponential and sinusoidal solutions.

The small-amplitude approach used in this analysis starts with the assumption that the peak-to-peak amplitude of only one of the two variables is small by comparison with its mean value. Removal of this constraint from the second variable is a desirable generalization: It permits the amplitude of the second variable to be calculated as a function of the coefficients C_i and A_i. The amplitude and waveform of the first variable (whose peak-to-peak amplitude is assumed small) cannot, in general, be obtained in closed form. The period of both variables is, of course, the same.

The approximate closed-form solution of this paper is obtained by assuming that the peak-to-peak amplitude of the fluctuations in x is small compared with the mean value of x:

$$\Delta x/x_1 \ll 1. \tag{3}$$

In a plasma physics context, this is equivalent to assuming that the gas is lightly ionized and the neutral gas density only slightly perturbed. In the following analysis, it is assumed in all cases that the function y is periodic in time and that the boundary conditions at time $t = 0$ are given by Eq. (4):

$$y\,(t = 0) = y_0, \quad \dot{y}\,(t = 0) = 0. \tag{4}$$

It is further assumed that y_0 is positive-definite and may be either the maximum value or minimum value of $y(t)$. No assumption was made about the relative magnitude of Δy and y, so that the assumption $\Delta x/x_1 \ll 1$ could also cover the case in which the relative amplitude Δy is either comparable to or small in comparison with the maximum value of y. In addition, the mathematical development is sufficiently general so that x_1 may be either larger *or* smaller than the mean value of y.

Immediate introduction of the small-amplitude approximation into Eqs. (1) and (2), setting $x(t) = x_1 + \Delta x(t)$,and then obtaining a solution for $y(t)$ by ignoring terms containing $\Delta x(t)$ would be premature since it would merely decouple Eq. (2) from the variation of $x(t)$ and lead to a family of trivial solutions for $y(t)$. If Eq. (2) is differentiated with respect to time, one obtains

$$\ddot{y} = A_1\dot{x} + A_2\dot{y} + A_3\dot{x}y + A_3x\dot{y} + 2A_4x\dot{x} + 2A_5y\dot{y}. \tag{5}$$

The first-order time derivatives in Eq. (5) may be replaced by substituting Eqs. (1) and (2) for them to obtain an equation of the form

$$\ddot{y} = l_0 + l_1y + l_2y^2 + l_3y^3, \tag{6}$$

where the coefficients l_i are polynomials in x, algebraic in the starting coefficients C_i and K_i. If it is assumed that the peak-to-peak variation of $x(t)$ is sufficiently small that $x(t)$ in the coefficients l_i may be approximated by its mean value x_1,

$$x(t) = x_1 + \Delta x(t) \approx x_1, \tag{7}$$

the coefficients l_i in Eq. (6) become constants. Equation (6) with these coefficients may then be considered an approximation to $y(t)$, a necessary condition for the validity of which is that the small-amplitude approximation of Eq. (3) be satisfied.

Equation (6) is recognizible as a differential equation whose periodic solutions are given in terms of Jacobian elliptic functions.[6,13] The integral tables of Byrd and Friedman[14] are extremely helpful in particular cases. The solutions to Eq. (6) for which $l_3 = 0$ and the other l_i are constant have been worked out and are available elsewhere.[15]

DISCUSSION

The utility of the above approximation may be assessed by considering a particular case given by the pair of equations

$$\dot{x} = -649.5 + 6.02x + 145y - 0.95xy - 0.00025x^2, \tag{8}$$

$$\dot{y} = 105 + 0.15x - 100y + 0.95xy - 0.0115x^2, \tag{9}$$

with the initial conditions $x(0) = 100.494$ and $y(0) = 0.881$. Note that $C_5 = A_5 = 0$. In Fig. 1 is shown the x–y phase trajectory of the exact solutions to Eqs. (8) and (9), and the exact solution for $y(t)$ is shown in Fig. 2(a).

[13] G. M. Murphy, *Ordinary Differential Equations and Their Solutions* (D. Van Nostrand Co., Inc., New York, 1960), p. 160.

[14] P. F. Byrd and M. D. Friedman, *Handbook of Elliptic Integrals for Engineers and Physicists* (Springer-Verlag, Berlin, 1954).

[15] J. R. Roth, "Periodic, Small-Amplitude Solutions to the Spatially Uniform Plasma Continuity Equations," NASA Technical Note TND-4472, 1968.

If a mean value $x_1 = 100$ is adopted, Eq. (6) becomes

$$\ddot{y} = 83.8653 - 131.2249y + 47.0542y^2. \quad (10)$$

A solution to this equation may be obtained by using expression 233:00 of Ref. 14, to obtain

$$y(t) = 0.881 + 0.2329\,\mathrm{sn}^2\,(3.2017t), \quad (11)$$

for which the elliptic modulus is

$$k^2 = 0.1782 \quad (12)$$

and sn is the Jacobian elliptic sine. Equation (11) is plotted in Fig. 2(b). A comparison of Figs. 2(a) and 2(b) shows that $y(t)$ as given by Eq. (11) differs from the exact solution only by a few percent in peak-to-peak amplitude and period.

Equation (6) will not provide a satisfactory approximation to $y(t)$ when the time-dependent terms in the coefficients l_i containing $x(t)$ are nonnegligible. An example of the failure of Eq. (6) to provide a satisfactory small-amplitude approximation is provided by Eq. (13) of Ref. 10, which was found experimentally to hold over only a portion of the range of interest.[11,12]

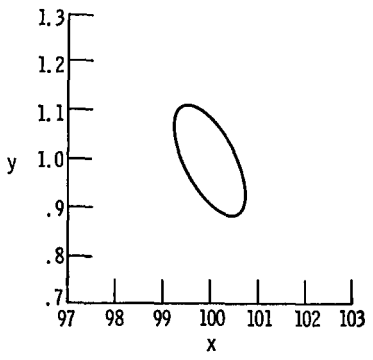

FIG. 1. Plot of the trajectory of the solutions to Eqs. (8) and (9) on the x–y phase plane.

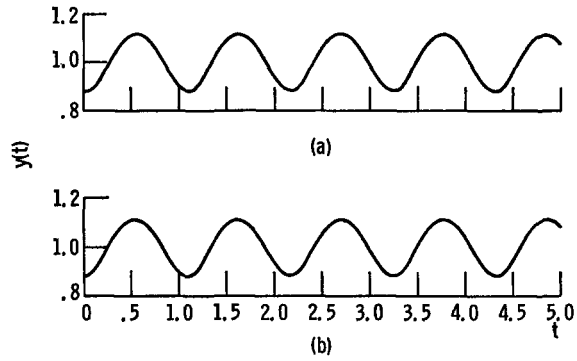

FIG. 2. (a) Plot of the exact solution of Eqs. (8) and (9) for $y(t)$ as a function of t. (b) Plot of the approximation to $y(t)$, given by Eq. (11), as a function of t. Note the very good agreement in amplitude and period between the exact solution and the present approximation.

CONCLUSIONS

The analysis presented herein extends the approximate solutions available in the literature (Refs. 3 and 5) to the case in which the peak-to-peak amplitude of $y(t)$ may be comparable to its mean value. It can be shown that the results obtained in this analysis reduce to those given in Refs. 3 and 5, in the limit of small peak-to-peak fluctuations in $y(t)$. The present analysis, therefore, contains the previously obtained closed-form solution to Eqs. (1) and (2) as a special case.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

On the Growth of the Ground-State Binding Energy with Increase in Potential Strength*

BARRY SIMON†
Princeton University, Princeton, New Jersey

(Received 7 February 1969)

We study the asymptotic behavior of the ground-state binding energy $G(\lambda)$ of $-\Delta + \lambda V$ as $\lambda \to \infty$. Unlike the number of bound states, $G(\lambda)$ does not have a universal power growth as $\lambda \to \infty$. It is shown, however, that as $\lambda \to \infty$ for Kato potentials

$$A\lambda < G(\lambda) < B\lambda^4.$$

Examples are presented for which $G \sim \lambda^\beta$ for any $1 < \beta < 4$. Other examples are presented which obey no power growth. We also prove theorems which reflect the close connection between the large λ behavior of G and the small r behavior of V for potentials with a single attractive singularity at $r = 0$. These can be roughly phrased as follows: If $V \sim -r^{-\alpha}$ for $\mathbf{r} \to 0$, then $G(\lambda) \sim \lambda^\beta$ with $\beta = 2/(2 - \alpha)$ as $\lambda \to \infty$.

1. INTRODUCTION

For large classes of potentials it has been shown that the total number of bound states $[N(\lambda V)]$, the number of states in any l channel $[n_l(\lambda V)]$, and the largest l-channel with bound states $[l_{\max}(\lambda V)]$, for the Hamiltonian $-\Delta + \lambda V$, all have power growth in λ as $\lambda \to \infty$. Specifically,

$$A\lambda^{\frac{3}{2}} < N(\lambda V) < B\lambda^{\frac{3}{2}} \quad \text{(see Ref. 1)},$$
$$a\lambda^{\frac{1}{2}} < l_{\max}(\lambda V) < b\lambda^{\frac{1}{2}} \quad \text{(see Ref. 1)},$$
$$C_l\lambda^{\frac{1}{2}} < n_l(\lambda V) < D_l\lambda^{\frac{1}{2}} \quad \text{(see Ref. 2)},$$

where each formula holds for λ sufficiently large, and the constants are V-dependent. However, the powers $\frac{1}{2}$ or $\frac{3}{2}$ are not V-dependent.

In this paper, we examine the analogous question for $G(\lambda V)$, the binding energy of the ground state of $-\Delta + \lambda V$. We define G so it is positive, i.e., G is the negative of the energy of the bound state. We only deal with potentials which are "negative somewhere" (see Ref. 1, Corollary 1). In this case if $\lambda_0 = \inf\{\lambda > 0 \mid N(\lambda V) > 0\}$, $G(\lambda V)$ only makes sense if $\lambda > \lambda_0$, and so we henceforth suppose $\lambda > \lambda_0$.

We see that, unlike N, n_l, and $l_{\max}$, G does not necessarily have power growth and, when it does, the power can vary between 1 and 4; i.e., there is not a universal power growth. We also show that, when V is a (not necessarily central) potential with a single attractive singularity $\mathbf{r} = 0$, that power growth of V at $\mathbf{r} = 0$ leads to power growth of G as $\lambda \to \infty$. Thus, the large coupling-constant behavior of G "mirrors" the small $\mathbf{r}$ behavior of V, which is an intuitively expected result.

For convenience, we restrict ourselves to potentials V which are the sum of an L^2 and an L^∞ function. This class was first considered by Kato,[3] who showed that $-\Delta + V$ is self-adjoint on $D(\Delta)$, the domain of Δ. We call such potentials Kato potentials.

Let us summarize the remainder of the paper. In Sec. 2, we discuss various types of power growth. In Sec. 3, we present a class of Kato potentials for which $G(\lambda) \sim \lambda^\beta$ for any $1 < \beta < 4$. In Sec. 4, we derive several general properties of G, including the fact[4] that any power growth must have $1 \leq \beta \leq 4$. In Sec. 5, we show that,if V has power growth $V \sim -r^\alpha$ for r small $(0 < \alpha < \frac{3}{2})$, then $G \sim \lambda^\beta$ with $\beta = 2/(2 - \alpha)$. Finally, in Sec. 6, we present a V with $G(\lambda)$ not possessing power growth as $\lambda \to \infty$.

2. TYPES OF POWER GROWTH

For a function $F(\lambda)$, there are at least five natural interpretations for the expression $F(\lambda) \sim \lambda^\alpha$:

(a) There is a C such that, for any ϵ, there is an Ω_0, with

$$(C - \epsilon)\lambda^\alpha < F(\lambda) < (C + \epsilon)\lambda^\alpha, \quad \text{for} \quad \lambda > \Omega_0.$$

(b) There exist C, C', and Ω_0, such that

$$C\lambda^\alpha < F(\lambda) < C'\lambda^\alpha, \quad \text{for} \quad \lambda > \Omega_0.$$

(c) For every ϵ, there is an Ω_0 such that

$$\lambda^{\alpha-\epsilon} < F(\lambda) < \lambda^{\alpha+\epsilon}, \quad \text{for} \quad \lambda > \Omega_0.$$

(d) $\lim_{\lambda\to\infty} \left(\frac{\lambda}{F(\lambda)} \frac{\partial F}{\partial \lambda}\right) = \alpha.$

(e) $\lim_{\lambda\to\infty} \frac{F'(\lambda)}{\lambda^{\alpha-1}}$ exists.

* This research partially sponsored under Air Force Office of Scientific Research under Contract AF49(638)-1545.

† N.S.F. pre-doctoral fellow.

[1] B. Simon, J. Math. Phys. **10**, 1123 (1969); F. Calogero and G. Cosenza, Nuovo Cimento **45A**, 867 (1966).

[2] F. Calogero, Commun. Math. Phys. **1**, 80 (1965); *Variable Phase Approach to Potential Scattering* (Academic Press Inc., New York, 1967), Chap. 23; J. H. E. Cohn, J. London Math. Soc. **40**, 523 (1965); **41**, 469 (1966); W. Frank, J. Math. Phys. **8**, 466 (1967).

[3] T. Kato, Trans. Am. Math. Soc. **70**, 195 (1951).

[4] $\beta < 4$ depends essentially on the fact that V is Kato. If we are less restrictive and allow $V \sim r^{-\alpha}$ with $2 > \alpha > \frac{3}{2}$, we get $G \sim \lambda\beta$ with $\beta > 4$.

(a), (b), and (c) can be rewritten in terms of limits:

(a') $\lim_{\lambda\to\infty} [F(\lambda)/\lambda^\alpha]$ exists and is finite and nonzero.

(b') $\overline{\lim}\, [F(\lambda)/\lambda^\alpha]$ and $\underline{\lim}\, [F(\lambda)/\lambda^\alpha]$ are finite and nonzero.

(c') $\lim_{\lambda\to\infty} \log F(\lambda)/\log \lambda = \alpha$.

We also remark that (c) is equivalent to a form with C, C' as in (b). The five types of growth are related by:

Theorem 2.1:

$$\begin{array}{ccccc} & & (a) & \to & (b) \\ & \nearrow & & & \searrow \\ (e) & \to & (d) & \to & (c) \end{array}$$

and (d) + (a) $\to$ (e). Moreover, no additional implications (single or multiple) hold.

Proof: (a) $\to$ (b) $\to$ (c) is immediate, as is (d) + (a') $\to$ (e).

To see that (d) $\to$ (c), suppose (d) holds and ϵ is given. Then find Ω_0 so that $\lambda > \Omega_0$ implies

$$\frac{\alpha-\epsilon}{\lambda} < \frac{1}{F(\lambda)}\frac{\partial F}{\partial\lambda} < \frac{\alpha+\epsilon}{\lambda},$$

which, integrated from Ω_0 to λ and then exponentiated, says

$$\left(\frac{\lambda}{\Omega_0}\right)^{\alpha-\epsilon} < \frac{F(\lambda)}{F(\Omega_0)} < \left(\frac{\lambda}{\Omega_0}\right)^{\alpha+\epsilon},$$

which is (c) (in the C, C' form).

(e) $\to$ (a) requires the integration of an inequality as in (d) $\to$ (c). This integration also shows that when (e) holds,

$$\lim \frac{F'(\lambda)}{\lambda^{\alpha-1}} = \alpha \lim \frac{F(\lambda)}{\lambda^\alpha},$$

which is L'Hôpital's rule. This means that (e) $\to$ (d).

To see that no other implications hold, we remark that $F(\lambda) = \lambda(2 + \sin\lambda^{-1})$ obeys (b) and (d) but not (a) or (e); $F(\lambda) = \lambda \log \lambda$ obeys (c) but not (a), (b), or (e); and $F(\lambda) = [1 + (1+\lambda)^{-1}\sin\lambda^4]$ obeys (a), (b), (c) but not (d) or (e). Q.E.D.

We write $F(\lambda) \sim \lambda^\alpha(-)$, to indicate growth of type $(-)$. We remark that similar power growth for $V(r)$ at $r \to 0$ is also meaningful. In this notation, we have the following:

Theorem 2.2: For a large class of potentials, $N(\lambda V) \sim \lambda^{\frac{3}{2}}$ (b); $n_l(\lambda V) \sim \lambda^{\frac{1}{2}}$(b); $l_{\max}(\lambda V) \sim \lambda^{\frac{1}{2}}$(b). (See Ref. 5.) For proof, see Refs. 1 and 2.

[5] K. Chadan [Nuovo Cimento **58A**, 191 (1968)] has shown that $n_l(\lambda V) \sim \lambda^{\frac{1}{2}}$ (a) for a restricted set of V.

3. A CLASS OF EXAMPLES WITH DIFFERENT POWER GROWTH

Consider the class of potentials, $V_\alpha(r) = -r^{-\alpha}$. For V to be Kato, we must[6] have $0 < \alpha < \frac{3}{2}$. We first remark that $-\Delta + \lambda V_\alpha$ always has bound states; in fact, because V falls at ∞ more slowly than r^{-2}, $-\Delta + \lambda V$ always has infinitely many bound states[7]; we do, in fact, prove this below.

Theorem 3.1: There is a unitary transformation $U(\lambda, \alpha)$ so that

$$U(\lambda,\alpha)^{-1}(-\Delta + \lambda V_\alpha)U(\lambda,\alpha) = \lambda^\beta(-\Delta + V_\alpha),$$

where $\beta = 2/(2-\alpha)$.

Proof: The formal scaling $\mathbf{r} \to \lambda^{\beta/2}\mathbf{r}$ takes $-\Delta + \lambda V_\alpha$ to $\lambda^\beta(-\Delta + V_\alpha)$. Thus we let

$$[U(\lambda,\alpha)\psi](\mathbf{r}) = \lambda^{3\beta/4}\psi(\lambda^{\beta/2}\mathbf{r}).$$

Then U is unitary and $U^{-1}(\Delta)U = \lambda^\beta\Delta$, $U^{-1}VU = \lambda^{\alpha\beta/2}V$. But

$$\tfrac{1}{2}\alpha\beta + 1 = 1 + [\alpha/(2-\alpha)] = 2/(2-\alpha) = \beta,$$

so the theorem is proven.

Corollary 3.2: For any λ_1, $-\Delta + \lambda_1 V_\alpha$ has infinitely many bound states.

Proof: A simple variational-principle argument as in Simon,[1] Corollary 5, shows that $-\Delta + \lambda V_\alpha$ has arbitrarily many bound states for λ sufficiently large [i.e., $N(\lambda V) \to \infty$]. But by Theorem 3.1, $-\Delta + \lambda V_\alpha$ and $-\Delta + \lambda_1 V_\alpha$ are unitarily equivalent up to a factor and so have the same number of bound states.

Corollary 3.3: $G(\lambda V_\alpha) = \lambda^\beta G(V_\alpha)$, and thus

$$G(\lambda V_\alpha) \sim \lambda^\beta(\mathrm{e}).$$

Thus, in particular, any power growth $1 < \beta < 4$ is possible for Kato potentials. We can see that $\beta = 1$ and $\beta = 4$ are also possible. [For $\beta = 1$, see Theorem 5.1; for $\beta = 4$, we remark that Theorem 5.3(iii) implies that $V(r) = r^{-\frac{3}{2}}(1 + |\log r|)^{-1}$, which is Kato, has $G(\lambda V) \sim \lambda^4$(c).]

4. GENERAL PROPERTIES OF $G(V)$

Let us write $G(\lambda)$ for $G(\lambda V)$ when V is held fixed.

Theorem 4.1: There is an Ω and a $B > 0$ such that $G(\lambda) > B\lambda$ if $\lambda > \Omega$.

[6] Physically we expect no trouble if $\alpha < 2$, but the mathematics is more complicated for $\alpha > \frac{3}{2}$, so we restrict ourselves.

[7] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience Publishers, Inc., New York, 1953), Vol. I, p. 447.

Proof: Let λ_0 be as in Sec. 1 and let ψ be the ground state of $-\Delta + (\lambda_0 + 1)V$. Then $b = \langle \psi, V\psi \rangle < 0$, so that $\langle \psi, (-\Delta + \lambda V)\psi \rangle = a + \lambda b$ with $b < 0$. By the variational principle, $G(\lambda) > -a - \lambda b$. Let $\Omega = -2a/b$, $B = -b/2$, and the theorem follows.

Theorem 4.2: There is a $c > 0$ such that $G(\lambda) < C\lambda^4$ for λ sufficiently large.[4]

Proof[8]*:* First suppose $V \in L^2$. Let ψ_λ represent the ground state, and write $k^2 = G(\lambda)$, with $k > 0$. By Kato's theorem, $\psi_\lambda \in D(\Delta)$ so that $(-\Delta + \lambda V)\psi_\lambda = -k^2\psi_\lambda$ implies $\psi_\lambda = (\Delta - k^2)^{-1}(\lambda V)\psi_\lambda$. Thus ψ_λ obeys the integral equation

$$\psi_\lambda(x) = -\frac{\lambda}{4\pi}\int \frac{\exp(-k|x-y|)}{|x-y|} V(y)\psi_\lambda(y)\, d^3y.$$

Thus

$$\begin{aligned}\int d^3x\, |\psi_\lambda(x)|^2 &< \frac{\lambda^2}{(4\pi)^2}\int d^3x \int d^3y \int d^3y'\, |\psi_\lambda(y)|\,|\psi_\lambda(y')| \\ &\quad \times \frac{|V(y)|\,|V(y')| \exp[-k(|x-y| + |x-y'|)]}{|x-y|\,|x-y'|} \\ &= \frac{\lambda^2}{(4\pi)^2}\int d^3y \int d^3y'\, |V(y)|\,|V(y')|\,|\psi_\lambda(y)|\,|\psi_\lambda(y')| \\ &\quad \times \int d^3x \frac{\exp[-k(|x-y| + |x-y'|)]}{|x-y|\,|x'-y|}.\end{aligned}$$

The x integral is $(2\pi/k)\exp(-k|y-y'|) \leq 2\pi/k$.[9] Thus,

$$\begin{aligned}\|\psi_\lambda\|^2 &\leq \frac{\lambda^2}{8\pi k}\left[\int d^3y\, |V(y)|\,|\psi_\lambda(y)|\right]^2 \\ &\leq \frac{\lambda^2}{8\pi k}\|\psi_\lambda\|^2 \int d^3y\, |V(y)|^2,\end{aligned}$$

where this last inequality follows from the Cauchy–Schwartz theorem. Therefore, $k < C^{\frac{1}{2}}\lambda^2$ or $G(\lambda) = k^2 < C\lambda^4$.

Now suppose that $V = V_1 + V_2$ with $V_1 \in L^2$ and $V_2 \in L^\infty$ and let $\|V_2\|_\infty = D$. Then $V > W$, where $W = V_1 - D$, so $G(\lambda V) < G(\lambda W) = G(\lambda V_1) + \lambda D < \lambda^4 C + \lambda D$, and so the theorem holds true in this case. Q.E.D.

Let us henceforth restrict ourselves to V's which lead to nondegenerate ground states for $-\Delta + \lambda V$; in particular, V can be any central potential and any but the most pathological noncentral potential.

Theorem 4.3: $G(\lambda)$ is a real analytic function for $\lambda > \lambda_0$ and $G'(\lambda) = -\langle \psi_\lambda, V\psi_\lambda \rangle$, where ψ_λ is the ground state.

Proof: V is a small perturbation of $-\Delta + \lambda V$ in the technical sense.[10] Since the ground state is nondegenerate, Rayleigh–Schrödinger perturbation theory for $G(\lambda + \lambda_0)$ (expanded in λ) converges,[11] and so G is real analytic. Moreover, $G'(\lambda)$ is given by the perturbation theoretic result as $-\langle \psi_\lambda, V\psi_\lambda \rangle$. Q.E.D.

The condition $-G'(\lambda) = \langle \psi_\lambda, V\psi_\lambda \rangle$ is the famous Feynman–Hellman theorem,[12] whose proof is seen to be completely rigorous for Kato potentials.

Theorem 4.4:
(i) $G(\lambda) > 0$;
(ii) $G'(\lambda) > \lambda^{-1}G(\lambda)$;
(iii) $G''(\lambda) > 0$.

Proof:
(i) is immediate.
(ii) follows from the fact that $-\Delta$ is positive for

$$G(\lambda) = \langle \psi_\lambda, \Delta\psi_\lambda \rangle - \lambda\langle \psi_\lambda, V\psi_\lambda \rangle < -\lambda G'(\lambda).$$

(iii) follows from the fact that the second-order perturbation term for $-G(\lambda)$ is negative. Alternately, we can prove that $G'(\lambda)$ is increasing directly, as follows: Let $\lambda_1 > \lambda_2$ and write $\psi_i = \psi_{\lambda_i}$. By the variational principle,

$$\begin{aligned}G(\lambda_1) = \langle \psi_1, \Delta\psi_1 \rangle - \lambda_1\langle \psi_1, V\psi_1 \rangle &> \langle \psi_2, \Delta\psi_2 \rangle \\ &\quad - \lambda_1\langle \psi_2, V\psi_2 \rangle, \\ G(\lambda_2) = \langle \psi_2, \Delta\psi_2 \rangle - \lambda_2\langle \psi_2, V\psi_2 \rangle &> \langle \psi_1, \Delta\psi_1 \rangle \\ &\quad - \lambda_2\langle \psi_1, V\psi_1 \rangle.\end{aligned}$$

Adding, we see that

$$(\lambda_1 - \lambda_2)\langle \psi_2, V\psi_2 \rangle > (\lambda_1 - \lambda_2)\langle \psi_1, V\psi_1 \rangle$$

or

$$-G'(\lambda_2) > -G'(\lambda_1). \qquad \text{Q.E.D.}$$

(ii) can be used to give an alternate proof of Theorem 4.1. (iii) tells us that G is convex. We remark that the reasonable conjecture that a smooth convex function has power growth is false, as our example in Sec. 6 shows.

[8] This is a modification of an argument, due to G. Tiktopoulos, private communication.

[9] This can be done most easily in prolate spheroidal coordinates or by using the Green's-function equation for $\exp(-k|x-y|)/|x-y|$.

[10] T. Kato, *Perturbation Theory for Linear Operators* (Springer-Verlag, Berlin, 1966), pp. 375–377.

[11] Reference 10, pp. 381–382.

[12] R. P. Feynman, Phys. Rev. **56**, 340 (1939); H. Hellman, *Einführung in die Quantenchemie* (Franz Deuticke, Leipzig, 1937), p. 285.

Theorem 4.5:
(i) $G(\lambda) < -\langle\psi_\lambda, (\lambda V + 4r^{-2})\psi_\lambda\rangle$.
(ii) If $\partial V/\partial r$ exists and the Virial theorem is obeyed, then

$$G(\lambda) = -\lambda\langle\psi_\lambda, \tfrac{1}{2}r(\partial V/\partial r) + V\psi_\lambda\rangle.$$

Proof:
(i) follows from the simple fact that, for any $\psi \in D(\Delta)$,

$$\langle\psi, -\Delta\psi\rangle > \left\langle\psi, \frac{1}{4r^2}\psi\right\rangle$$

(see Ref. 13 for a proof when ψ is C^∞ of compact support and then use the fact that Δ is the closure of its restriction to these functions).
(ii) follows from the Virial theorem

$$2\langle\psi_\lambda, -\Delta\psi_\lambda\rangle = \langle\psi_\lambda, r(\partial V/\partial r)\psi_\lambda\rangle. \quad \text{Q.E.D.}$$

For a C^1 central Kato potential, the standard proof of the Virial theorem[14] should go through. For the noncentral case, Weidmann[15] has given simple conditions for the Virial theorem to hold as a rigorous result.

Theorem 4.6: If V_1 and V_2 are Kato and $V_1(r) < V_2(r)$ for all r, then $G(\lambda V_1) > G(\lambda V_2)$ (even if the V_i do not go to 0 at $r = \infty$).

Proof: This is an immediate consequence of the variational principle. We have, in fact, already used this argument in the proof of Theorem 4.2.

5. POWER GROWTH AT $r = 0$

Let us first eliminate the trivial case of potentials that are bounded below.

Theorem 5.1: If $V(r) > -C$ for some constant $C > 0$, then $G(\lambda) \sim \lambda$(e).

Proof: [Notice that (b) growth is trivial since we have Theorem 4.1 and $G(\lambda) < \lambda C$.] We have $G'(\lambda) = -\langle\psi_\lambda, V\psi_\lambda\rangle < C$. Thus G' is bounded above, but G' is increasing by Theorem 4.4(iii). Thus $\lim_{\lambda\to\infty} G'(\lambda)$ exists and is finite, i.e., $G(\lambda) \sim \lambda$(e). Q.E.D.

We deal in the remainder of this section with the more interesting case of potentials with attractive centers. We could deal with the case of finitely many centers, but we restrict ourselves to the following class.

Definition: We say a potential V is attractive if
(i) $\lim_{r\to 0} V(\mathbf{r}) = -\infty$,
(ii) for every $R > 0$, V is bounded in the region $\{\mathbf{r} \mid r > R\}$. (We say that V is "bounded outside spheres.")

Definition: We say V has attractive growth α(a) [or α(b), α(c)] if
(i) V is attractive,
(ii) $V \sim -r^{-\alpha}$(a) as $r \to 0$ [or (b) or (c)].
We reserve α(d) and α(e) for a slightly stronger condition.

We are heading towards proving that whenever V has attractive growth α of some type, $G(\lambda) \sim \lambda^\beta$ with the same type of growth where $\beta = 2/(2 - \alpha)$. For growth of type (a), (b), or (c), the crucial element of the proof is the following:

Lemma 5.2: If $V(r) < -Ar^{-\gamma} + B$, then $G(\lambda) > AR_\gamma\lambda^\delta - B\lambda$; and if $V(r) > -Cr^{-\gamma} - D$, then $G(\lambda) < CR_\gamma\lambda^\delta + D\lambda$, where $\delta = 2/(2 - \gamma)$, and $R_\gamma = G(V_\gamma)$ as in Sec. 3.

Proof: This follows immediately from Theorems 3.3 and 4.6.

Theorem 5.3: Let $0 < \alpha < \frac{3}{2}$. Then
(i) If V has attractive growth α(a), then $G(\lambda) \sim \lambda^\beta$(a).
(ii) If V has attractive growth α(b), then $G(\lambda) \sim \lambda^\beta$(b).
(iii) If V has attractive growth α(c), then $G(\lambda) \sim \lambda^\beta$(c).
In each case, $\beta = 2/(2 - \alpha)$. In case (i),

$$\lim_{\lambda\to\infty} \frac{G(\lambda)}{\lambda^\beta} = R_\alpha \lim_{r\to 0} -\frac{V(r)}{r^\alpha}.$$

Proof: (i) Since V is bounded outside spheres and $V \sim -r^{-\alpha}$(a), there is a $C = \lim_{r\to\infty} -V(r)/r^\alpha$, so that for every ϵ, there is a B with

$$-(C + \epsilon)r^{-\alpha} - B < V(r) < -(C - \epsilon)r^{-\alpha} + B.$$

Thus Lemma 5.2 implies

$$(C - \epsilon)R_\alpha\lambda^\beta - B\lambda < G(\lambda) < (C + \epsilon)R_\alpha\lambda^\beta + B\lambda.$$

Since $\beta > 1$, for every ϵ we have

$$(C - \epsilon)R_\alpha < \underline{\lim} \frac{G(\lambda)}{\lambda^\beta} < \overline{\lim} \frac{G(\lambda)}{\lambda^\beta} < (C + \epsilon)R_\alpha.$$

Since ϵ is arbitrary,

$$\lim_{\lambda\to\infty} \frac{G(\lambda)}{\lambda^\beta} = R_\alpha C.$$

[13] Reference 7, p. 446.
[14] The normal proof really supposes $\psi_\lambda \in D[(\partial^2/\partial r^2)r(\partial/\partial r)]$. For negative energy eigenfunctions in a C^1 Kato potential, one should be able to prove that ψ_λ decreases exponentially from its integral equation. For the radial Schrödinger equation, it follows that ψ_λ is C^3.
[15] J. Weidmann, Bull. Am. Math. Soc. 73, 452 (1967).

(ii) Since V is bounded outside spheres and $V \sim -r^{-\alpha}$(b), there is a B, C, C' so that $-Cr^{-\alpha} - B < V(r) < -C'r^{-\alpha} + B$. Thus Lemma 5.2 implies

$$R_\alpha C'\lambda^\beta - B\lambda < G(\lambda) < R_\alpha C\lambda^\beta + B\lambda.$$

Thus for λ sufficiently large,

$$\tfrac{1}{2}R_\alpha C'\lambda^\beta < G(\lambda) < 2R_\alpha C\lambda^\beta.$$

(iii) Since V is bounded outside spheres and $V \sim -r^{-\alpha}$(c), for every ϵ, there is a B so that

$$-r^{\alpha+\epsilon} - B \leq V(r) < -r^{\alpha-\epsilon} + B.$$

Thus Lemma 5.2 implies

$$-B\lambda + R_{\alpha-\epsilon}\lambda^{\beta-\epsilon'} \leq G(\lambda) \leq R_{\alpha+\epsilon}\lambda^{\beta+\epsilon''} + B\lambda,$$

so that

$$\beta - \epsilon' < \underline{\lim}\frac{\log G(\lambda)}{\log \lambda} < \overline{\lim}\frac{\log G(\lambda)}{\log \lambda} < \beta + \epsilon''.$$

Since ϵ is arbitrary and ϵ', $\epsilon'' \to 0$ as $\epsilon \to 0$, we have

$$\lim_{\lambda\to\infty}\frac{\log G(\lambda)}{\log \lambda} = \beta. \qquad \text{Q.E.D.}$$

We finally get to the interesting cases of Type (d) or (e) growth. The Virial theorem is crucial and so we restrict ourselves to the following:

Definition: We say V has attractive growth α(d) or α(e) if

(i) V is attractive,
(ii) $V \sim -r^{-\alpha}$(d) [or (e)],
(iii) V obeys a Virial theorem,
(iv) $r(\partial V/\partial r)$ is bounded outside spheres.

Condition (iv) is not essential in its strong form. Using the fact that ψ_λ falls off exponentially, we could probably survive with $e^{-\alpha r}[r(\partial V/\partial r)]$ bounded for some α. However, only the most pathological V's fail to obey (iv) and so we do not consider weakening it in detail.

Lemma 5.4: Let $f_i(\lambda)$, $g_i(\lambda)$ $(i = 1, 2)$ be positive functions with $g_2 \to \infty$ as $\lambda \to \infty$. Suppose that

(i) $\overline{\lim}_{\lambda\to\infty}\dfrac{g_1}{g_2} = C < \infty$,

(ii) $|f_i(\lambda) - g_i(\lambda)| \leq B$ for some B and all λ.

Then

$$\underline{\lim}\frac{f_1(\lambda)}{f_2(\lambda)} = \underline{\lim}\frac{g_1(\lambda)}{g_2(\lambda)}; \quad \overline{\lim}\frac{f_1(\lambda)}{f_2(\lambda)} = \overline{\lim}\frac{g_1(\lambda)}{g_2(\lambda)}.$$

Proof: We first remark that $f_2 \to \infty$ as $\lambda \to \infty$ by (i) and (ii). Moreover,

$$\frac{f_1}{f_2} - \frac{g_1}{g_2} = \frac{f_1g_2 - f_2g_1}{f_2g_2} = \frac{(f_1 - g_1)g_2 - g_1(f_2 - g_2)}{g_2f_2}$$
$$= \frac{(f_1 - g_1)}{f_2} - \frac{g_1}{g_2}\frac{(f_2 - g_2)}{f_2}.$$

But for λ sufficiently large, $(g_1/g_2) \leq 2C$, so

$$\left|\frac{f_1}{f_2} - \frac{g_1}{g_2}\right| \leq \frac{B}{f_2} + \frac{2CB}{f_2} \to 0.$$

Thus

$$\lim_{\lambda\to\infty}\left|\frac{f_1}{f_2} - \frac{g_1}{g_2}\right| = 0,$$

so that the theorem follows.

Theorem 5.5: Let $0 < \alpha < \frac{3}{2}$. Then:

(i) If V has attractive growth α(d), then $G(\lambda) \sim \lambda^\beta$(d).

(ii) If V has attractive growth α(e), then $G(\lambda) \sim \lambda^\beta$(e).

In either case, $\beta = 2/(2 - \alpha)$. In case (ii),

$$\lim_{\lambda\to\infty}\frac{G'(\lambda)}{\lambda^{\beta-1}} = \frac{\beta R_\alpha}{\alpha}\lim_{r\to 0} -\frac{V'(r)}{r^{\alpha-1}}.$$

Proof: (i) By the Virial and Feynman–Hellman theorems,

$$\frac{G(\lambda)}{\lambda G'(\lambda)} = \frac{\langle\psi_\lambda, (\frac{1}{2}r(\partial V/\partial r) + V)\psi_\lambda\rangle}{\langle\psi_\lambda, V\psi_\lambda\rangle}.$$

Let us fix some R and let

$$\langle\psi, \phi\rangle_R = \int_{|\mathbf{r}|<R} d^3r\bar{\psi}(\mathbf{r})\phi(\mathbf{r}).$$

Then $|\langle\psi_\lambda, V\psi_\lambda\rangle - \langle\psi_\lambda, V\psi_\lambda\rangle_R|$ and

$$|\langle\psi_\lambda, (\tfrac{1}{2}r(\partial V/\partial r) + V)\psi_\lambda\rangle - \langle\psi_\lambda, (\tfrac{1}{2}r(\partial V/\partial r) + V)\psi_\lambda\rangle_R|$$

are bounded,

$$\overline{\lim}\frac{G(\lambda)}{\lambda G'(\lambda)} \leq 1$$

[by Theorem 4.4(ii)] and $\langle\psi_\lambda, V\psi_\lambda\rangle \to \infty$. Thus, by Lemma 5.4, the $\overline{\lim}$ and $\underline{\lim}$ of $G(\lambda)/\lambda G'(\lambda)$ are the same as that of

$$\frac{\langle\psi_\lambda, (\frac{1}{2}r(\partial V/\partial r) + V)\psi_\lambda\rangle_R}{\langle\psi_\lambda, V\psi_\lambda\rangle_R}$$

for any fixed R. Since $V(r) \to -\infty$, pick R_0, so that $V(\mathbf{r}) \leq -1$ for $|\mathbf{r}| \leq R_0$. Given ϵ, choose R so that $R \leq R_0$ and

$$\left|\frac{r}{V}\frac{\partial V}{\partial r} + \alpha\right| \leq \epsilon \quad \text{for} \quad |r| \leq R.$$

Thus

$$\left|\left\langle \psi_\lambda, \left(\frac{r}{2}\frac{\partial V}{\partial r} + V\right)\psi_\lambda\right\rangle_R + \left(\frac{\alpha}{2} - 1\right)\langle \psi_\lambda, V\psi_\lambda\rangle_R\right|$$
$$= \left|\left\langle \psi_\lambda, V\left(\frac{r}{2V}\frac{\partial V}{\partial r} + \frac{\alpha}{2}\right)\psi_\lambda\right\rangle_R\right| \leq \frac{\epsilon}{2}|\langle \psi_\lambda, V\psi_\lambda\rangle_R|.$$

Thus

$$1 - \frac{\alpha}{2} - \frac{\epsilon}{2} \leq \frac{\langle \psi_\lambda, (\frac{1}{2}r(\partial V/\partial r) + V)\psi_\lambda\rangle_R}{\langle \psi_\lambda, V\psi_\lambda\rangle_R} < 1 - \frac{\alpha}{2} + \frac{\epsilon}{2},$$

so that

$$1 - \frac{\alpha}{2} - \frac{\epsilon}{2} \leq \underline{\lim}\, \frac{G(\lambda)}{\lambda G'(\lambda)} \leq \overline{\lim}\, \frac{G(\lambda)}{\lambda G'(\lambda)} \leq 1 - \frac{\alpha}{2} + \frac{\epsilon}{2}.$$

Since ϵ is arbitrary,

$$\lim \frac{G(\lambda)}{\lambda G'(\lambda)} = 1 - \frac{\alpha}{2} = \frac{1}{\beta}$$

so (i) is proved.

(ii) By Theorem 2.1, (e) implies (d) and (a) growth. Thus, by Theorems 5.3(a) and 5.5(a), if V has attractive growth α(e), then

$$G(\lambda) \sim \lambda^\beta(\mathrm{d}) + (\mathrm{a}).$$

But (d) + (a) $\to$ (e) so $G(\lambda) \sim \lambda^\beta$(e). Moreover,

$$\lim_{\lambda\to\infty} \frac{G'(\lambda)}{\lambda^{\beta-1}} = \lim_{\lambda\to\infty} \frac{\lambda G'(\lambda)}{G(\lambda)} \lim_{\lambda\to\infty} \frac{G(\lambda)}{\lambda^\beta}$$
$$= \beta R_\alpha \lim_{r\to\infty} \frac{V(r)}{r^\alpha} \qquad \text{(by Theorem 5.3)}$$
$$= \beta R_\alpha \left(\lim_{r\to 0} \frac{V'(r)}{r^{\alpha-1}}\right)\left(\lim_{r\to 0} \frac{V(r)}{rV'(r)}\right)$$
$$= \frac{\beta R_\alpha}{\alpha} \lim_{r\to 0} \frac{V'(r)}{r^{\alpha-1}}.$$

6. A CLASS OF EXAMPLES WITHOUT POWER GROWTH

In this final section we construct potentials for which $G(\lambda)$ does not have power growth. On the basis of Sec. 5, we should try to construct them from potentials V which do not have power growth at $\mathbf{r} = 0$. In fact, we use central potentials which alternately grow as $r^{-\alpha_1}$ and $r^{-\alpha_2}$. The key to handling $G(\lambda)$ in this case is the following lemma:

Lemma 6.1: Let V be a central potential which monotonically decreases as r decreases. There is a positive constant C, *independent of* V, so that for any fixed R and λ

$$-\left(\frac{C}{R^2} + \lambda V(R)\right) \leq G(\lambda) \leq -\min_{r>0}\left[\frac{1}{4r^2} + \lambda V(r)\right].$$

Proof: The upper bound follows from Theorem 4.5(i). To get the lower bound, pick (independently of V) some fixed normalized C^∞ function ϕ, with support inside the ball $\{\mathbf{r} \mid |\mathbf{r}| < 1\}$ and let

$$C = \langle \phi, -\Delta\phi\rangle > 0.$$

Let $\phi_R(\mathbf{r}) = R^{-\frac{3}{2}}\phi(\mathbf{r}/R)$, so that

$$\int |\phi_R(r)|^2\, d^3r = 1.$$

Then, by the variational principle, for any R,

$$-G(\lambda) \leq \langle \phi_R, (-\Delta + \lambda V)\phi_R\rangle.$$

We have $\langle \phi_R, -\Delta\phi_R\rangle = C/R^2$. Moreover, since supp $(\phi_R) \subset \{\mathbf{r} \mid |\mathbf{r}| < R\}$ and V is decreasing,

$$\langle \phi_R, V\phi_R\rangle \leq V(R).$$

Thus $-G(\lambda) \leq (C/R^2) + \lambda V(R)$, which yields the lower bound. Q.E.D.

We could have proven Theorem 5.3 using a strengthened form of Lemma 6.1.

We construct our examples by using the following basic fact about potentials which grow as a "variable power":

Lemma 6.2: Let γ_n be a sequence of numbers with $0 < \alpha_1 < \gamma_n \leq \alpha_2 < \frac{3}{2}$. Let A_n be a sequence of numbers with

$$0 < \cdots < A_n < A_{n-1} < \cdots < A_1 = 1 < A_0 = \infty.$$

Let $V(r)$ be given by

$$V(r) = -r^{-\gamma_n} + B_n,$$

if $A_{n+1} < r < A_n$, where the B_n's are defined by $B_0 = 0$, and by the requirement that V be continuous. Let

$$\delta_n = 2/(2 - \gamma_n).$$

Then there exist constants C_1 and C_2 dependent only on α_1 and α_2, so that

$$-\lambda B_n + C_1\lambda^{\delta_n} < G(\lambda) < -\lambda B_n + C_2\lambda^{\delta_n}$$

whenever

$$\tfrac{1}{2}\alpha_1^{-1}(A_n)^{\alpha_1-2} < \lambda < \{4[(A_{n+1})^{2-\gamma_n} - B_n(A_{n+1})^2]\}^{-1}.$$

Proof: We first show that under the conditions on λ, that $\lambda V + 4r^{-2}$ takes its minimum value when $r \in (A_{n+1}, A_n)$. For suppose that $r > A_n$. Then, for any m,

$$\gamma_m\lambda > \alpha_1\lambda > \tfrac{1}{2}(A_n)^{(\gamma_1-2)} > \tfrac{1}{2}(A_n)^{\alpha_m-2} > \tfrac{1}{2}r^{\gamma_m-2}$$

(we have used the facts, $\alpha_1 \leq \gamma_m < 2$ and $A_n < 1$). Thus

$$(d/dr)(-\lambda r^{-\gamma_m} + \tfrac{1}{4}r^{-2}) = (\lambda\gamma_m - \tfrac{1}{2}r^{\gamma_m-2})r^{-\gamma_m-1} \geq 0,$$

so that $\lambda V + 4r^{-2}$ decreases monotonically in the region $r > A_n$, and so the minimum is taken for $r \leq A_n$.

On the other hand, the upper bound for λ implies that

$$\frac{-\lambda}{(A_{n+1})^{\gamma_n}} + \lambda B_n + \frac{1}{4(A_{n+1})^2}$$
$$= \frac{1}{4(A_{n+1})^2}\{1 - 4\lambda[(A_{n+1})^{2-\gamma_n} - B_n(A_{n+1})^2]\} \geq 0,$$

so that $+\lambda V + \frac{1}{4}r^{-2} \geq 0$ at $r = A_{n+1}$. But $|r^2V|$ is monotone decreasing as $r \to 0$, so $\lambda V + \frac{1}{4}r^{-2} > 0$ if $r < A_{n+1}$. Thus, we have shown that $\lambda V + 4r^{-2}$ takes its minimum value in (A_{n+1}, A_n) and hence at the point

$$r = R_n(\lambda) \equiv (\gamma_n\lambda)^{1/\gamma_n-2}.$$

At this point

$$\lambda V(R_n) = \lambda B_n - (\gamma_n)^{1/(2-\gamma_n)}\lambda^{\delta_n}$$

and

$$R_n^{-2} = (\gamma_n)^{2/(2-\gamma_n)}\lambda^{\delta_n}.$$

Thus, using Lemma 6.1 with $R = R_n$, we see that

$$-\lambda B_n + C_1\lambda^{\delta_n} < G(\lambda) < -\lambda B_n + C_2\lambda^{\delta_n},$$

where

$$C_2 = \max_{\alpha_1<\gamma<\alpha_2}\{(\gamma)^{1/(2-\gamma)} - \tfrac{1}{4}(\gamma)^{2/(2-\gamma)}\},$$
$$C_1 = \min_{\alpha_1<\gamma<\alpha_2}\{(\gamma)^{1/(2-r)} - C(\gamma)^{2/(2-\gamma)}\}. \quad \text{Q.E.D.}$$

To assure that $C_1 > 0$, we restrict ourselves to $\alpha_2 < \min(1, C^{-1})$. This restriction is not essential; if we were to work harder, we could probably remove it. However, we only want examples of nonpower growth—we do not wish to examine it in detail—so we take the lazy way out.

Theorem 6.3: Let α_1, α_2 be two numbers with $0 < \alpha_1 < \alpha_2 < \min(1, C^{-1})$, and let $\beta_i = 2/(2 - \alpha_i)$. Then there is a Kato potential V with

$$\underline{\lim}\,\frac{\log G(\lambda V)}{\log\lambda} \leq \beta_1; \quad \overline{\lim}\,\frac{\log G(\lambda V)}{\log\lambda} \geq \beta_2.$$

In particular, G fails to have power growth at ∞.

Proof: Pick the sequence γ_n as in Lemma 6.2 by $\gamma_{2m} = \alpha_2$, $\gamma_{2m+1} = \alpha_1$. We shortly pick the A_n's and B_n's recursively. Once we have done this, we define V as in Lemma 6.2. For all r, $V > -r^{-\alpha_2}$. Thus V is Kato. Let $A_1 = 1$ and suppose we have picked $A_2, \cdots, A_n$, and thus also $B_1, \cdots, B_n$. If we had

$$-\lambda B_n + C_1\lambda^{\delta_n} < G(\lambda) < -\lambda B_n + C_2\lambda^{\delta_n}$$

for all $\lambda > (2\alpha_1)^{-1}A_n^{\alpha_1-2}$, then it would follow that

$$\lim_{\lambda\to\infty}\frac{\log G(\lambda)}{\log\lambda} = \delta_n;$$

so we pick $\lambda_n > (2\alpha_1)A_n^{\alpha_1-2}$, so that

$$\left|\frac{\log(-\lambda B_n + C_1\lambda^{\delta_n})}{\log\lambda} - \delta_n\right| < \frac{1}{n}, \quad \text{at } \lambda = \lambda_n,$$

and

$$\left|\frac{\log(-\lambda B_n + C_2\lambda^{\delta_n})}{\log\lambda} - \delta_n\right| < \frac{1}{n}, \quad \text{at } \lambda = \lambda_n$$

($C_1 > 0$ is crucial here). Now pick A_{n+1}, so that

$$\{4[(A_{n+1})^{2-\gamma_n} - B_n(A_{n+1})^2]\}^{-1} > \lambda_n$$

and

$$A_{n+1} < A_n.$$

This is always possible since

$$4[(A_{n+1})^{2-\gamma_n} - B_n(A_{n+1})^2]^{-1} \to \infty$$

as $A_{n+1} \to 0$. Determine B_{n+1} and proceed to pick $A_{n+2}, \cdots$. The V so constructed has the property that $\log G(\lambda_{2n})/\log\lambda_{2n} \to \beta_2$ and

$$\log G(\lambda_{2n+1})/\log\lambda_{2n+1} \to \beta_1.$$

ACKNOWLEDGMENT

The author would like to thank M. Reed for an enlightening conversation on the nonpower growth of convex functions.

Summation over Feynman Histories in Polar Coordinates

D. PEAK AND A. INOMATA
Department of Physics, State University of New York, Albany, New York

(Received 17 April 1968)

Use of polar coordinates is examined in performing summation over all Feynman histories. Several relationships for the Lagrangian path integral and the Hamiltonian path integral are derived in the central-force problem. Applications are made for a harmonic oscillator, a charged particle in a uniform magnetic field, a particle in an inverse-square potential, and a rigid rotator. Transformations from Cartesian to polar coordinates in path integrals are rather different from those in ordinary calculus and this complicates evaluation of path integrals in polars. However, it is observed that for systems of central symmetry use of polars is often advantageous over Cartesians.

I. INTRODUCTION

Of fundamental importance to quantum mechanics is the Schrödinger equation

$$-i\partial_t\psi(\mathbf{r}, t) = H\psi(\mathbf{r}, t) \tag{1}$$

containing H, the Hamiltonian of the system, as a differential operator. This differential equation can be replaced by an integral equation

$$\psi(\mathbf{r}'', \tau) = \int K(\mathbf{r}'', \mathbf{r}'; \tau)\psi(\mathbf{r}', 0)\, d\mathbf{r}', \tag{2}$$

if the initial condition $\psi(\mathbf{r}'', 0) = \psi(\mathbf{r}', 0)$ is satisfied. The kernel of Eq. (2) corresponds to the propagator of the wavefunction $\psi(\mathbf{r}, t)$ from the point $\mathbf{r}'$ to $\mathbf{r}''$ in time τ.

In Feynman's Lagrangian formulation,[1] it is asserted that the kernel is given by a path integral

$$K(\mathbf{r}'', \mathbf{r}'; \tau) = \int \exp\,[iS(\mathbf{r}'', \mathbf{r}')]\mathfrak{D}\mathbf{r}(t). \tag{3}$$

Here, integrations are over all possible paths, or histories, starting at $\mathbf{r}' = \mathbf{r}(0)$ and terminating at $\mathbf{r}'' = \mathbf{r}(\tau)$. The function $S(\mathbf{r}'', \mathbf{r}')$ in the integrand is the classical action

$$S(\mathbf{r}'', \mathbf{r}') = \int_0^\tau L(\dot{\mathbf{r}}, \mathbf{r})\, dt, \tag{4}$$

$L(\dot{\mathbf{r}}, \mathbf{r})$ being the Lagrangian of the system in question.

As an alternative approach to quantization, Feynman's formalism has attracted much attention.[2] However, this approach is applicable only to a limited class of problems.[3] Certainly any effort to extend it beyond its present limits would be worthwhile. In most applications available so far, calculations are done in Cartesian coordinates. It has been suggested that the integral over all paths may be performed in polar coordinates as well.[4] It is the purpose of the present paper to demonstrate the usefulness of polar coordinates in evaluating the path integral for specific particle systems. Indeed, it is observed that most solvable examples in Cartesians are equally well treated in polars. Use of polars seems of better advantage for certain systems of central symmetry, although the applications considered are all essentially of the harmonic-oscillator type.

In Sec. II, we derive several general expressions for the path integral in the central-force problem. The Hamiltonian path integral equivalent to Feynman's Lagrangian path integral is also discussed in polars. Section III is devoted to applications. The propagator of the harmonic oscillator is the first example, a limiting case of which includes the free particle. A slight modification of the procedure of computing the propagator for the harmonic oscillator in polars leads to the result of Sondheimer and Wilson for charged particles in a uniform magnetic field.[5] The third example is the rigid rotator, for which the Hamiltonian path integral is utilized. The final calculation, concerned with a particle in an inverse-square potential, could hardly be completed in Cartesians but is found trivial in polars. In an appendix, derivations of the formulas used in the text are given. Throughout this paper we employ natural units, i.e., $\hbar = c = 1$.

II. THE CENTRAL-FORCE PROBLEM

The Lagrangian Path Integral

It is customary to define the summation over Feynman histories (3) by[3]

$$K(\mathbf{r}'', \mathbf{r}'; \tau) = \lim_{N\to\infty} A_N \int \exp\left[i\sum_{j=1}^{N} S(\mathbf{r}_j, \mathbf{r}_{j-1})\right] d\mathbf{r}_1\, d\mathbf{r}_2 \cdots d\mathbf{r}_{N-1}, \tag{5}$$

where $\mathbf{r}_j = \mathbf{r}(t_j)$, $\mathbf{r}_0 = \mathbf{r}'$, $\mathbf{r}_N = \mathbf{r}''$, $t_j - t_{j-1} = \tau/N = \epsilon$, and A_N is the normalization factor in the Nth

[1] R. P. Feynman, Rev. Mod. Phys. **20**, 367 (1948).
[2] S. G. Brush, Rev. Mod. Phys. **33**, 79 (1961).
[3] See, e.g., R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill Book Co., Inc., New York, 1965).
[4] S. F. Edwards and Y. V. Gulyaev, Proc. Roy. Soc. (London) **A279**, 229 (1964); S. Ozaki, Lectures at Kyushu University, 1955 (unpublished).
[5] E. H. Sondheimer and A. H. Wilson, Proc. Roy. Soc. (London), **A210**, 173 (1951). For derivation by the path-integral method, see M. L. Glasser, Phys. Rev. **113**, B831 (1964); A. Inomata, Benét Laboratories, U.S. Army, Technical Report WVT-6718, 1967.

approximation. The partial action in a small time interval $\Delta t_j = t_j - t_{j-1}$ may be approximated by

$$S(\mathbf{r}_j, \mathbf{r}_{j-1}) \approx \epsilon L(\Delta \mathbf{r}_j/\epsilon, \mathbf{r}_j), \tag{6}$$

where $\Delta \mathbf{r}_j = \mathbf{r}_j - \mathbf{r}_{j-1}$ and $\Delta t_j = \epsilon$. This approximation reflects the situation that the important contributions to the path integral are only from the paths close to the classical one.

In polar coordinates, the squared distance between two points $\mathbf{r}_j(r_j, \theta_j, \phi_j)$ and $\mathbf{r}_{j-1}(r_{j-1}, \theta_{j-1}, \phi_{j-1})$ is

$$(\Delta \mathbf{r}_j)^2 = r_j^2 + r_{j-1}^2 - 2r_j r_{j-1} \cos \Theta_j, \tag{7}$$

where

$$\cos \Theta_j = \cos \theta_j \cos \theta_{j-1} + \sin \theta_j \sin \theta_{j-1} \cos (\phi_j - \phi_{j-1}). \tag{8}$$

For a particle of mass m in a central potential, the partial action is given by

$$S(\mathbf{r}_j, \mathbf{r}_{j-1}) = \tfrac{1}{2}m(r_j^2 + r_{j-1}^2)/\epsilon - (m/\epsilon) r_j r_{j-1} \cos \Theta_j - \epsilon V(r_j). \tag{9}$$

If use is made of the expansion formula

$$\exp (u \cos \Theta) = \left(\frac{\pi}{2u}\right)^{\frac{1}{2}} \sum_{l=0}^{\infty} (2l + 1) P_l(\cos \Theta) I_{l+\frac{1}{2}}(u) \tag{10}$$

in terms of $P_l(\cos \Theta)$, the Legendre function, and $I_{l+\frac{1}{2}}(u)$, the modified Bessel function, the integrand of Eq. (5) can be written as

$$\exp\left[i \sum_{j=1}^{N} S(\mathbf{r}_j, \mathbf{r}_{j-1})\right] = \prod_{j=1}^{N} \left[\sum_{l_j=0}^{\infty} (2l_j + 1) P_{l_j}(\cos \Theta_j) R_{l_j}(r_j, r_{j-1})\right], \tag{11}$$

where

$$R_l(r_j, r_{j-1}) = \left\{\frac{i\pi\epsilon}{2m r_j r_{j-1}}\right\}^{\frac{1}{2}} \times \exp\left[\frac{im}{2\epsilon}(r_j^2 + r_{j-1}^2) - i\epsilon V(r_j)\right] I_{l+\frac{1}{2}}\left(\frac{m}{i\epsilon} r_j r_{j-1}\right). \tag{12}$$

After interchanging multiplications and summations, the right-hand side of Eq. (11) becomes

$$\sum_{l_1 l_2 \cdots l_N} \left\{\prod_{j=1}^{N} [(2l_j + 1) P_{l_j}(\cos \Theta_j) R_{l_j}(r_j, r_{j-1})]\right\}.$$

Substitution of this result into Eq. (5) yields

$$K(\mathbf{r}'', \mathbf{r}'; \tau) = \lim_{N \to \infty} A_N \sum_{l_1 l_2 \cdots l_N} \times \int \prod_{j=1}^{N} \left\{(2l_j + 1) P_{l_j}(\cos \Theta_j) R_{l_j}(r_j, r_{j-1})\right\} \times \prod^{N-1} (r^2 \sin \theta \, dr \, d\theta \, d\phi). \tag{13}$$

Here,

$$\prod^{N-1} (r^2 \sin \theta \, dr \, d\theta \, d\phi) = \prod_{j=1}^{N-1} r_j^2 \sin \theta_j \, dr_j \, d\theta_j \, d\phi_j;$$

this convention will be adapted hereafter. The angular integrations in Eq. (13) can easily be carried out. First, expand $P_l(\cos \Theta)$ in terms of the spherical harmonics

$$P_l(\cos \Theta_j) = \frac{4\pi}{2l + 1} \sum_{n=-l}^{l} Y_l^{n*}(\theta_j, \phi_j) Y_l^n(\theta_{j-1}, \phi_{j-1}). \tag{14}$$

Then use the orthogonality relation

$$\iint Y_l^{n*}(\theta, \phi) Y_{l'}^{n'}(\theta, \phi) \sin \theta \, d\theta \, d\phi = \delta_{ll'} \delta_{nn'} \tag{15}$$

to obtain

$$\iint \prod_{j=1}^{N} \{(2l_j + 1) P_{l_j}(\cos \Theta_j)\} \prod^{N-1} (\sin \theta \, d\theta \, d\phi) = (4\pi)^N \delta_{l l_N} \prod_{j=1}^{N-1} \delta_{l_{j+1} l_N} \sum_{n=-l}^{l} Y_l^{n*}(\theta'', \phi'') Y_l^n(\theta', \phi'). \tag{16}$$

As a result, for each quantum number l, the radial and angular contributions to the propagator are separable; that is,

$$K(r'', \theta'', \phi''; r', \theta', \phi'; \tau) = \sum_{l=0}^{\infty} \sum_{n=-l}^{l} K_l(r'', r'; \tau) Y_l^{n*}(\theta'', \phi'') Y_l^n(\theta', \phi'), \tag{17}$$

with the radial propagator of the l wave

$$K_l(r'', r'; \tau) = \lim_{N \to \infty} (4\pi)^N A_N \int \prod_{j=1}^{N} \{R_l(r_j, r_{j-1})\} \prod^{N-1} (r^2 \, dr) \tag{18}$$

remaining to be evaluated, contingent on specification of the potential. The normalization factor, so chosen that the total propagator (17) may be unitary, is

$$A_N = (2\pi i\epsilon/m)^{-\frac{3}{2}N}. \tag{19}$$

The Hamiltonian Path Integral

It has been shown[6] that in Cartesian coordinates

$$A_N \int \exp\left[i \int L \, dt\right] \prod^{N-1} (d\mathbf{r}) = (2\pi)^{-3N} \iint \exp\left[i \int (\mathbf{p} \cdot \dot{\mathbf{r}} - H) \, dt\right] \prod^{N} (d\mathbf{p}) \prod^{N-1} (d)\mathbf{r}, \tag{20}$$

where $\mathbf{p}$ is the momentum conjugate to $\mathbf{r}$. This implies that the Hamiltonian path integral in phase space is

[6] H. Davies, Proc. Cambridge Phil. Soc. **59**, 147 (1963); C. Garrod, Rev. Mod. Phys. **38**, 483 (1966).

identical to Feynman's Lagrangian path integral as far as particle systems described on the Cartesian basis are concerned. Since we are interested in the approximation (6), corrections higher than the first order in ϵ are unimportant. If the approximation

$$\cos\delta \approx 1 - \tfrac{1}{2}\delta^2 \tag{21}$$

is valid for angular changes δ in the time interval ϵ, then one can express (20) in polars. However, the approximation (21) is not relevant, as Edwards and Gulyaev have pointed out.[4] This may be compared with the situation that the simple procedure of replacing p by $-i(\partial/\partial q)$ is not reliable in polars. The irrelevance arises from the fact that even if the changes in Cartesian variables are of the order of ϵ, the corresponding changes in angular variables are not.

In order to take all contributions up to first order in ϵ into account, we utilize the asymptotic form of $I_\nu(u/\epsilon)$ for small ϵ,

$$I_\nu\left(\frac{u}{\epsilon}\right) \sim \left(\frac{2\pi u}{\epsilon}\right)^{-\frac{1}{2}} \exp\left[\frac{u}{\epsilon} - \tfrac{1}{2}(\nu^2 - \tfrac{1}{4})\frac{\epsilon}{u} + O(\epsilon^2)\right], \tag{22}$$

and replace Eq. (10) by

$$\exp\left[\frac{u}{\epsilon}\cos\delta\right] \approx \frac{\epsilon}{2u}\sum_{\nu=-\infty}^{\infty} \exp\left[i\nu\delta + \frac{u}{\epsilon} - \frac{(\nu^2 - \tfrac{1}{4})\epsilon}{2u}\right]. \tag{23}$$

Use of this approximation formula and the identity

$$\tfrac{1}{2}m(\Delta r)^2 = p\Delta r - \tfrac{1}{2}\epsilon p^2/m + \tfrac{1}{2}\epsilon(p - m\Delta r/\epsilon)^2/m \tag{24}$$

enable us to derive

$$\begin{aligned}\exp&\left[iS(\mathbf{r}_j, \mathbf{r}_{j-1}) - \frac{i\epsilon}{2m}\left(p_j - \frac{m}{\epsilon}\Delta r_j\right)^2\right] \\ &= \frac{i\epsilon}{2\pi m r_j r_{j-1}}(\sin\theta_j \sin\theta_{j-1})^{\frac{1}{2}} \\ &\quad\times \sum_{\mu,\nu} \exp\left[ip_j\Delta r_j + i\mu\Delta\theta_j + i\nu\Delta\phi_j - \frac{i\epsilon(\mu^2 - \frac{1}{4})}{2mr_jr_{j-1}}\right. \\ &\quad\left. - \frac{i\epsilon(\nu^2 - \frac{1}{4})}{2mr_jr_{j-1}\sin\theta_j\sin\theta_{j-1}} - i\epsilon V(r_j)\right].\end{aligned} \tag{25}$$

Integrating both sides of (25) over the entire range of p_j and dividing by the constant factor resulting from the Fresnel integral on the left-hand side yield

$$\begin{aligned}\exp&[iS(\mathbf{r}_j, \mathbf{r}_{j-1})] \\ &= \left(\frac{i\epsilon}{2\pi m}\right)^{\frac{3}{2}}(r_jr_{j-1}\sin\theta_j\sin\theta_{j-1})^{-\frac{1}{2}} \\ &\quad\times \sum_{\mu,\nu}\int \exp\left[ip_j\Delta r_j + i\mu\Delta\theta_j + i\nu\Delta\phi_j - \frac{i\epsilon(\mu^2 - \frac{1}{4})}{2mr_jr_{j-1}}\right. \\ &\quad\left. - \frac{i\epsilon(\nu^2 - \frac{1}{4})}{2mr_jr_{j-1}\sin\theta_j\sin\theta_{j-1}} - i\epsilon V(r_j)\right] dp_j.\end{aligned} \tag{26}$$

On substitution of (26), the path integral turns out to be of the Hamiltonian form, analogous to that in Eq. (20); namely, for N large,

$$\begin{aligned}A_N\int \exp\left[i\int L\,dt\right]\prod^{N-1}(r^2\sin\theta\,dr\,d\theta\,d\phi) \\ = (2\pi)^{-3N}\int \exp\left[i\int(p\dot{r} + p_\theta\dot{\theta} + p_\phi\dot{\phi} - H)\,dt\right] \\ \times \prod^{N}(dp\,dp_\theta\,dp_\phi)\prod^{N-1}(dr\,d\theta\,d\phi),\end{aligned} \tag{27}$$

where we have made formal replacements

$$\mu \to p_\theta, \quad \nu \to p_\phi, \tag{28}$$

$$p^2 + (\mu^2 - \tfrac{1}{4})/r^2 + (\nu^2 - \tfrac{1}{4})/(r^2\sin^2\theta) + 2mV(r) \to H, \tag{29}$$

and

$$\sum_{\mu_1\mu_2\cdots\mu_N}\sum_{\nu_1\nu_2\cdots\nu_N} \to (r'^2r''^2\sin\theta'\sin\theta'')^{\frac{1}{2}} \times \int\int\prod^{N}(dp_\theta\,dp_\phi). \tag{30}$$

There is an essential feature of the representation in polars due to the premise that the system has rotational symmetry. Because of the periodicity associated with rotation, the angular momentum assumes only discrete values, so that the propagator may remain single-valued. In this regard, the replacements (28)–(30) are literally formal. It may be worth noting that if the system is bounded by a finite cubic box, the representation in Cartesians also requires each component of the linear momentum to take discrete values. Then integrations over the momentum variables in Eq. (20) must be treated as summations over possible discrete values. The difference of symmetries assumed for the system is the main source of the difference between the features of the representations in Cartesians and in polars.

In fact, the angular motion is solely determined by the rotational symmetry, and much involved calculations are unnecessary. What remains to be determined is only the radial motion. It is therefore more practical to develop the Hamiltonian path integral for the radial propagator than to handle the formal expression (27). In the following, we shall derive the radial propagator for the l-wave in the Hamiltonian form. With the approximation formula (22), the radial function (12) is given by

$$\begin{aligned}R_l(r_j, r_{j-1}) \\ = \frac{i\epsilon}{2mr_jr_{j-1}}\exp\left[\frac{im(\Delta r_j)^2}{2\epsilon} - \frac{i\epsilon l(l+1)}{2mr_jr_{j-1}} - i\epsilon V(r_j)\right].\end{aligned} \tag{31}$$

In the same fashion as Eq. (23), we write the radial function as

$$R_l(r_j, r_{j-1}) = \frac{1}{4}\left(\frac{2\pi\epsilon i}{m}\right)^{\frac{1}{2}}(r_j r_{j-1})^{-1}\int_{-\infty}^{\infty}\exp\left[ip_j\Delta r_j - \frac{i\epsilon}{2m}p_j^2 - \frac{i\epsilon l(l+1)}{2mr_j r_{j-1}} - i\epsilon V(r_j)\right]dp_j. \quad (32)$$

Equation (31) shows that the radial propagator for the l wave is

$$K_l(r'', r'; \tau) = (r'r'')^{-1}\lim_{N\to\infty}\left(\frac{m}{2\pi i\epsilon}\right)^{\frac{1}{2}N} \times\int\exp\left\{i\int\left[\tfrac{1}{2}m\dot{r}^2 - \frac{l(l+1)}{r^2} - V(r)\right]dt\right\}\prod^{N-1}(dr), \quad (33)$$

which coincides with the results of Ozaki and of Edwards and Gulyaev[4] for $V(r) = 0$. On the other hand, use of expression (32) leads to the Hamiltonian path integral

$$K_l(r'', r'; \tau) = (r'r'')^{-1}\lim_{N\to\infty}\left(\frac{m}{2i\epsilon}\right)^{N} \times\int\int\exp\left[i\int(p\dot{r} - H_l)\,dt\right]\prod^{N}(dp)\prod^{N-1}(dr), \quad (34)$$

where

$$H_l = \frac{1}{2m}\left[p^2 + \frac{l(l+1)}{r^2}\right] + V(r). \quad (35)$$

III. APPLICATIONS

The Harmonic Oscillator

For the harmonic oscillator having spring constant $k = m\omega^2$, the Lagrangian is

$$L = \tfrac{1}{2}m(\dot{\mathbf{r}}^2 - \omega^2 r^2) \quad (36)$$

and, hence, the partial action in the time interval Δt_j is given by

$$S(\mathbf{r}_j, \mathbf{r}_{j-1}) = \tfrac{1}{2}(m/\epsilon)(r_j^2 + r_{j-1}^2) - (m/\epsilon)r_j r_{j-1}\cos\Theta_j - \tfrac{1}{2}\epsilon m\omega^2 r_j^2. \quad (37)$$

The corresponding radial function reads

$$R_l(r_j, r_{j-1}) = \left[\frac{i\pi\epsilon}{2mr_j r_{j-1}}\right]^{\frac{1}{2}} \times\exp\left[\frac{im}{2\epsilon}(r_j^2 + r_{j-1}^2) + \tfrac{1}{2}i\epsilon m\omega^2 r_j^2\right] \times I_{l+\frac{1}{2}}\left(\frac{m}{i\epsilon}r_j r_{j-1}\right), \quad (38)$$

with which the radial propagator of the l wave can be put in the form

$$K_l(r'', r'; \tau) = (r'r'')^{-\frac{1}{2}}\lim_{N\to\infty}(-i\beta)^N\exp\left[\tfrac{1}{2}i\beta(r'^2 + r''^2)\right] \times\int\exp\left[i\alpha(r_1^2 + r_2^2 + \cdots + r_{N-1}^2)\right] \times I_{l+\frac{1}{2}}(-i\beta r_0 r_1)\cdots I_{l+\frac{1}{2}}(-i\beta r_{N-1}r_N)\prod^{N-1}(r\,dr), \quad (39)$$

where

$$\beta = m/\epsilon, \quad \alpha = \beta(1 - \tfrac{1}{2}\omega^2\epsilon^2). \quad (40)$$

As is shown in Appendix A, the formula

$$\int_0^{\infty}\exp(i\alpha r^2)I_\nu(-iar)I_\nu(-ibr)\,r\,dr = \frac{i}{2\alpha}\exp\left[\frac{-i(a^2+b^2)}{4\alpha}\right]I_\nu\left(-i\frac{ab}{2\alpha}\right) \quad (41)$$

is valid for $\mathrm{Re}\,(\nu) > -1$ and $\mathrm{Re}\,(\alpha) > 0$. Repeated use of the above formula yields

$$\int\exp\left[i\alpha(r_1^2 + \cdots + r_{N-1}^2)\right]I_\nu(-i\beta r_0 r_1)\cdots I_\nu(-i\beta r_{N-1}r_N)\prod^{N-1}(r\,dr) = \prod_{j=1}^{N-1}\left(\frac{i}{2\alpha_j}\right) \times\exp\left\{-i\left[r'^2\sum_{j=1}^{N-1}\frac{\beta_j^2}{4\alpha_j} + r''^2\frac{\beta^2}{\alpha_N}\right]\right\}I_\nu(-i\beta_N r_0 r_1), \quad (42)$$

where α_j and β_j are coefficients to be determined by solving the following algebraic equations:

$$\alpha_1 = \alpha, \quad \alpha_{j+1} = \alpha - \frac{\beta^2}{4\alpha_j}, \quad \text{for } j \geq 1, \quad (43)$$

$$\beta_1 = \beta, \quad \beta_{j+1} = \beta\prod_{k=1}^{j}\frac{\beta}{2\alpha_k}, \quad \text{for } j \geq 1. \quad (44)$$

The multi-integral formula (42) enables us to complete the radial integrations in Eq. (39); i.e.,

$$K_l(r'', r'; \tau) = -i(r'r'')^{-\frac{1}{2}}\lim_{N\to\infty}a_N \times\exp(if_N r'^2 + ig_N r''^2)I_{l+\frac{1}{2}}(-ia_N r'r''). \quad (45)$$

Our problem reduces to determining the factors

$$a_N = \prod_{j=1}^{N-1}\frac{\beta}{2\alpha_j}, \quad (46)$$

$$f_N = \tfrac{1}{2}\beta - \frac{1}{4}\sum_{j=1}^{N-1}\frac{\beta_j^2}{\alpha_j}, \quad (47)$$

$$g_N = \tfrac{1}{2}\beta - \frac{1}{4}\frac{\beta^2}{\alpha_N}. \quad (48)$$

As is seen in Appendix B, the coefficient α_j satisfying Eq. (43) can be given in terms of a polynomial so that

the factors a_N, f_N, and g_N defined above are expressible in series form. However, what we are interested in is the limiting value of each factor for $N \to \infty$. In Appendix B, it is also shown that as N tends to infinity

$$a_N \to m\omega \csc(\omega\tau), \tag{49}$$

$$f_N \to \tfrac{1}{2}m\omega \cot(\omega\tau), \tag{50}$$

$$g_N \to \tfrac{1}{2}m\omega \cot(\omega\tau). \tag{51}$$

Therefore, the radial propagator becomes

$$K_l(r'', r'; \tau) = -i(r'r'')^{-\frac{1}{2}}m\omega \csc(\omega\tau) \times \exp[\tfrac{1}{2}im\omega(r'^2 + r''^2)\cot(\omega\tau)] \times I_{l+\frac{1}{2}}[-im\omega r'r'' \csc(\omega\tau)]. \tag{52}$$

As a particular case, the propagator of the two-dimensional oscillator can be obtained:

$$K(r'', \phi''; r', \phi'; \tau) = \frac{m\omega}{2\pi i \sin(\omega\tau)} \exp\left\{\frac{im\omega}{2\sin(\omega\tau)} \times [(r'^2 + r''^2)\cos(\omega\tau) - 2r'r''\cos(\phi'' - \phi')]\right\}. \tag{53}$$

For the one-dimensional oscillator,

$$K(r'', r'; \tau) = \left(\frac{m\omega}{2\pi i \sin(\omega\tau)}\right)^{\frac{1}{2}} \exp[\tfrac{1}{2}im\omega(r'^2 + r''^2)\cot(\omega\tau)]. \tag{54}$$

In the limit where ω vanishes, the propagator (52) reduces to that of a free particle in three dimensions. In the same limit, the propagator (54) leads to the one-dimensional free-particle case

$$K_0(r'', r'; \tau) = \left(\frac{m}{2\pi i\tau}\right)^{\frac{1}{2}} \exp\left[\frac{im}{2\tau}(r'' - r')^2\right]. \tag{55}$$

From Eqs. (33) and (52) follows the useful relation

$$\int \exp\left\{i\int\left[\tfrac{1}{2}m\dot{r}^2 - \frac{\nu^2 - \frac{1}{4}}{r^2} - \frac{m\omega^2}{r^2} - \tfrac{1}{2}m\omega^2r^2\right]dt\right\}\mathfrak{D}r = -i(r'r'')^{\frac{1}{2}}m\omega \csc(\omega\tau) \exp[\tfrac{1}{2}im\omega(r'^2 + r''^2)\cot(\omega\tau)] \times I_\nu[-im\omega r'r'' \csc(\omega\tau)], \tag{56}$$

for $\mathrm{Re}(\nu) > -1$.

The Charged Particle in a Uniform Magnetic Field

The Lagrangian for a particle of charge e moving in a constant uniform magnetic field B, which is applied along the z axis, is

$$L = \tfrac{1}{2}m[\dot{\mathbf{r}}^2 + 2\omega(x\dot{y} - y\dot{x})], \tag{57}$$

or, in cylindrical coordinates (r, θ, z),

$$L = \tfrac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2 + 2\omega r^2\dot{\theta}) + \tfrac{1}{2}m\dot{z}^2, \tag{58}$$

where $\frac{1}{2}eB/m = \omega$ is the Larmor frequency of the charged particle. Introduction of a new angular variable ϕ such that

$$\phi = \theta + \omega t \tag{59}$$

and

$$\dot{\theta}^2 + 2\omega\dot{\theta} = \dot{\phi}^2 - \omega^2$$

casts the Lagrangian (29) in the form

$$L = \tfrac{1}{2}m(\dot{r}^2 + r^2\dot{\phi}^2 - \omega^2r^2) + \tfrac{1}{2}m\dot{z}^2. \tag{60}$$

The corresponding partial action in the time interval ϵ is

$$S(\mathbf{r}_j, \mathbf{r}_{j-1}) = \tfrac{1}{2}m(r_j^2 + r_{j-1}^2)/\epsilon - (m/\epsilon)r_jr_{j-1}\cos(\phi_j - \phi_{j-1}) - \tfrac{1}{2}\epsilon m\omega^2r_j^2 + \tfrac{1}{2}m(\Delta z_j)^2/\epsilon. \tag{61}$$

It is clear that the motion of a charged particle in a uniform magnetic field is equivalent to a combination of two-dimensional harmonic oscillation and free motion perpendicular to the plane of oscillation. Correspondingly, the action in a given time interval can be separated into contributions from the harmonic oscillation and the free motion. Thus, the propagator for this system is a product of the propagators for a harmonic oscillator in the (r, ϕ) plane and a free particle in the z direction. That is,

$$K(r'', \phi'', z''; r', \phi', z'; \tau) = K(r'', \phi''; r', \phi'; \tau)K_0(z'', z'; \tau). \tag{62}$$

The propagators on the right-hand side of Eq. (62) have been expressed in Eqs. (53) and (55). Transforming the variable ϕ back into the real angular variable θ by Eq. (59) leads to the desired propagator

$$K(r'', \theta'', z''; r', \theta', z'; \tau) = \left(\frac{m}{2\pi i}\right)^{\frac{3}{2}}\frac{\omega\tau}{\sin(\omega\tau)} \times \exp\left\{\frac{im\omega}{2\sin(\omega\tau)}[(r'^2 + r''^2)\cos(\omega\tau) - 2r'r''\cos(\theta'' - \theta' + \omega\tau)] + \frac{im}{2\tau}(z'' - z')^2\right\}. \tag{63}$$

It is well known that the simple replacement of τ in the propagator by $-i(kT)^{-1}$, where k is the Boltzman constant and T the temperature, enables one to write down the density matrix in statistical mechanics. Following this procedure, we obtain the density matrix for an ensemble of charged particles

in a uniform magnetic field as

$$\rho(r'', r'; T) = \left(\frac{mkT}{2\pi}\right)^{\frac{3}{2}} \frac{\omega k^{-1}T^{-1}}{\sin(\omega k^{-1}T^{-1})} \exp\left\{-\tfrac{1}{2}mkT \times \left[\frac{2i\omega}{kT} r'r'' \sin(\theta' - \theta'') + \frac{\omega}{kT}(r'^2 + r''^2)\coth\left(\frac{\omega}{kT}\right) - \frac{2\omega}{kT}\coth\left(\frac{\omega}{kT}\right)\cos(\theta'' - \theta') + (z'' - z')^2\right]\right\}, \tag{64}$$

which is of the same form as that derived by Sondheimer and Wilson.[5]

The Rigid Rotator

The expression (33) for the l wave is more convenient than (16) for evaluating the propagator of a rigid rotator. Let r_0 be the radius of the sphere on which the rotator is constrained. Then let $\delta(r_j - r_0)$ take the place of $\exp[-i\epsilon V(r_j)]$ in the radial propagator (33); that is,

$$K_l(r_0; \tau) = r_0^{-2} \lim_{N\to\infty} \left(\frac{m}{2\pi i\epsilon}\right)^{\frac{1}{2}} \times \prod_{j=1}^{N} \left\{\exp\left[\frac{im}{2\epsilon}(r_j - r_{j-1})^2 - \frac{i\epsilon l(l+1)}{r_j r_{j-1}}\right]\delta(r_j - r_0)\right\} \prod^{N-1}(dr). \tag{65}$$

After integration, the following simple form results:

$$K_l(r_0; \tau) = r_0^{-2} \exp\left[\frac{\tau}{2im}\frac{l(l+1)}{r_0}\right]. \tag{66}$$

Thus Eq. (15) gives, for this rotator,

$$K(\theta'', \phi''; \theta', \phi'; \tau) = r_0^{-1} \sum_{l=0}^{\infty} \sum_{n=-l}^{l} \exp\left[\frac{\tau l(l+1)}{2imr_0}\right] Y_l^{n*}(\theta'', \phi'') Y_l^n(\theta', \phi'). \tag{67}$$

The Particle in an Inverse-Square Potential

For a particle in an attractive potential

$$V(r) = k^2/r^2, \tag{68}$$

the derivation of the propagator is a trivial matter when one utilizes relation (56), setting $\omega = 0$ and replacing $l + \frac{1}{2}$ by $[(l + \frac{1}{2})^2 + k^2]^{\frac{1}{2}}$. To see the situation in more detail, we start with the radial function (31), which now takes the form

$$R_l(r_j, r_{j-1}) = \frac{i\epsilon}{2mr_j r_{j-1}} \times \exp\left\{\frac{im}{2\epsilon}(\Delta r_j)^2 - \frac{i\epsilon[l(l+1) + k^2]}{2mr_j r_{j-1}}\right\}. \tag{69}$$

Within the approximation adopted, the asymptotic expansion formula (22) enables us to rewrite the radial function as

$$R_l(r_j, r_{j-1}) = \left(\frac{i\pi\epsilon}{2mr_j r_{j-1}}\right)^{\frac{1}{2}} \times \exp\left[\frac{im(r_j^2 + r_{j-1}^2)}{2\epsilon}\right] I_\lambda\left(\frac{mr_j r_{j-1}}{i\epsilon}\right), \tag{70}$$

where

$$\lambda(l) = [(l + \tfrac{1}{2})^2 + k^2]^{\frac{1}{2}}. \tag{71}$$

Since the radial integrations are independent of $\lambda(l)$, there results from Eq. (70) in much the same manner that Eq. (33) comes from Eq. (31) the radial propagator for the l wave

$$K_l(r'', r'; \tau) = (r'r'')^{\frac{1}{2}}(-im/\tau) \times \exp[\tfrac{1}{2}im(r'^2 + r''^2)/\tau] I_\lambda(-imr'r''/\tau), \tag{72}$$

with λ defined by (71). By Eq. (15), the propagator for a particle in the potential (68) is

$$K(r'', \theta'', \phi''; r', \theta', \phi'; \tau) = (r'r'')^{-\frac{1}{2}}\left(\frac{m}{i\tau}\right) \exp\left[\frac{im}{2\tau}(r'^2 + r''^2)\right] \times \sum_{l=0}^{\infty} \sum_{n=-l}^{l} I_{\lambda(l)}\left(\frac{mr'r''}{i\tau}\right) Y_l^{n*}(\theta'', \phi'') Y_l^n(\theta', \phi'). \tag{73}$$

ACKNOWLEDGMENT

The authors wish to acknowledge helpful discussions with Professor Jack H. Smith.

APPENDIX A: DERIVATION OF FORMULA (41)

Consider the contour integral

$$\oint_\Gamma e^{-\alpha z^2} I_\nu(az) I_\nu(bz) z\, dz \tag{A1}$$

for $\mathrm{Re}(\nu) > -1$ and $\mathrm{Re}(\alpha) > 0$. As is shown in Fig. 1, Γ is a closed contour consisting of a path from A to B along the positive real axis, a circular arclike path from B to C, a path from C to D along the line with $\arg(z) = 3\pi/4$, and a small circular arclike path from D to A about the origin. The integrand is regular in the z plane cut along the negative real axis. As a consequence, the integral (A1) vanishes. Since the contributions from the two arclike paths disappear when the appropriate limits are invoked, we have

$$i\int_0^\infty e^{i\alpha z^2} I_\nu(az/i^{\frac{1}{2}}) I_\nu(bz/i^{\frac{1}{2}}) z\, dz + \int_0^\infty e^{-\alpha z^2} I_\nu(za) I_\nu(bz) z\, dz = 0. \tag{A2}$$

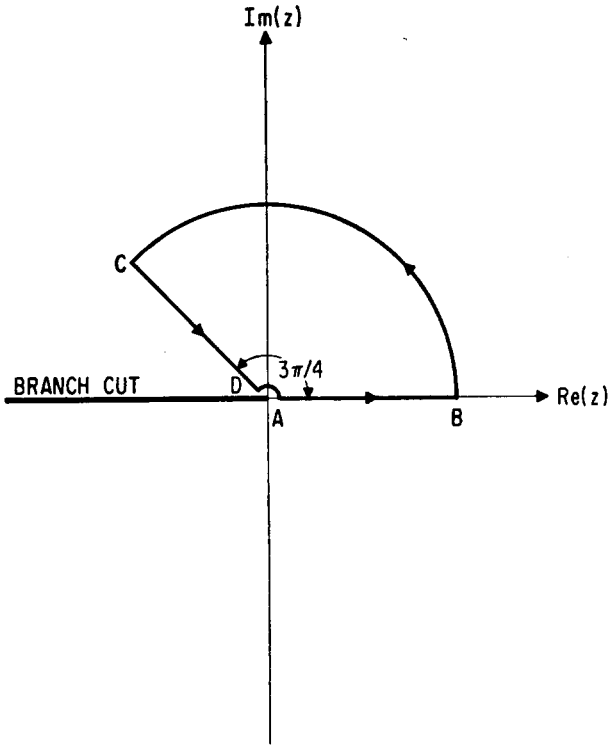

FIG. 1. Contour Γ taken in Eq. (A1).

Thus, by Weber's formula[7]

$$\int_0^\infty e^{-\alpha z^2} I_\nu(az) I_\nu(bz) z\, dz = \frac{1}{2\alpha} \exp\left[\frac{a^2+b^2}{4\alpha}\right] I_\nu\left(\frac{ab}{2\alpha}\right), \tag{A3}$$

we obtain our formula (41), after replacing a by $a/i^{\frac{1}{2}}$ and b by $b/i^{\frac{1}{2}}$.

APPENDIX B: DETERMINATION OF THE COEFFICIENTS a, f, AND g

Let λ_j be $2\alpha_j/\beta$ and define the finite product of λ_j^{-1}:

$$\Lambda_k = \prod_{j=1}^{k} \lambda_j^{-1}. \tag{B1}$$

Then the coefficients defined in Eqs. (46), (47), and (48) are all expressible in terms of β and Λ_k:

$$a_N = \beta \Lambda_{N-1}, \tag{B2}$$

$$f_N = \tfrac{1}{2}\beta\left(1 - \sum_{j=1}^{N-1} \Lambda_j \Lambda_{j-1}\right), \tag{B3}$$

$$g_N = \tfrac{1}{2}\beta(1 - \Lambda_N'/\Lambda_{N-1}). \tag{B4}$$

[7] See G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge University Press. Cambridge, England, 1962), 2nd ed., p.395.

Now consider a series

$$X_k = \sum_{j=0}^{k} (-1)^j \binom{k+j+1}{2j+1} \eta^{2j+1}. \tag{B5}$$

By induction, it is straightforward to show that

$$X_{k+1} + X_{k-1} = X_1 X_k. \tag{B6}$$

It is apparent that

$$\lambda_k = X_k / X_{k-1} \tag{B7}$$

satisfies the relation

$$\lambda_{k+1} + \lambda_k^{-1} = \lambda_1, \tag{B8}$$

which coincides with Eq. (43) for $\lambda_j = 2\alpha_j/\beta$. From (B7) it immediately follows that

$$\Lambda_k = X_1 / X_k, \tag{B9}$$

where η is $\omega\epsilon$.

Let N and η be such that $N\eta$ remains finite for all N. Then

$$X_{k-1} \to \sin(k\eta) \tag{B10}$$

as N goes to infinity. To see this, compare the sum of the first n terms of X_{k-1},

$$T_n = \sum_{j=0}^{n<k} (-1)^j \binom{k+j}{2j+1} \eta^{2j+1}, \tag{B11}$$

with that of the series for $\sin(k\eta)$,

$$S_n = \sum_{j=0}^{n<k} (-1)^j \frac{(k\eta)^{2j+1}}{(2j+1)!}; \tag{B12}$$

that is,

$$|T_n - S_n| < [n(n+1)\eta/k] \sinh(k\eta), \tag{B13}$$

from which the convergence (B10) is obvious.

Accordingly, we have

$$a_N \to \beta\eta \csc(N\eta), \tag{B14}$$

$$f_N \to \tfrac{1}{2}\beta\eta \cot(N\eta). \tag{B15}$$

It is also clear that $(\Lambda_k \Lambda_{k-1})$ converges uniformly to $\eta^2 \csc[(k+1)\eta] \csc(k\eta)$ in the same limit. Therefore, we may write

$$\lim_{N\to\infty} \sum_{j=1}^{N-1} (\Lambda_k \Lambda_{k-1})^{-1} = \eta \int_\eta^{N\eta} \csc^2 x\, dx \tag{B16}$$

and determine the limiting value of g as

$$g_N \to \tfrac{1}{2}\beta\eta \cot(N\eta). \tag{B17}$$

In Eqs. (B14), (B15), and (B17), let $\beta\eta = m\omega$ and $N\eta = \omega\tau$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Inverse Functions of the Products of Two Bessel Functions

WILLIAM SOLLFREY*
The RAND Corporation, Santa Monica, California

(Received 27 December 1968)

Special cases of the inverse function of the product of two spherical Bessel functions have been found recently by other writers as ${}_1F_2$ hypergeometric functions. We give the general expression as the derivative of a product of spherical Bessel functions. These results have also been found in the classical literature.

A recent article[1] considers the problem of finding the inverse function of the product of two spherical Bessel functions. The results are given in terms of ${}_1F_2$ hypergeometric functions. Define the inverse function by the equation

$$\int_0^\infty j_l(kr)j_{l+m}(kr)g_{l,m}(kr')\,dk = \delta(r - r'). \quad (1)$$

Then Ref. 1 gives formulas in terms of ${}_1F_2$ functions for $m = 0, 1, 2$, and explicit results for $g_{0,0}$, $g_{1,0}$, $g_{2,0}$, $g_{0,1}$, and $g_{0,2}$. They conjecture that the higher-order inverse functions can be found by their methods.

This problem has been treated in the classical literature. The inverse function $g_{l,0}$ was found, in a form differing by an integration by parts, by Bateman,[2] and the higher-order forms by Fox.[3] All these results are available in a standard text.[4] The general and quite simple formula is

$$g_{l,m}(x) = \frac{8x^2}{\pi}\frac{d}{dx}x^2 n_l(x)j_{l+m}(x), \quad (2)$$

where $n_l(x)$ denotes the spherical Neumann function. Equation (2) reduces to all the special cases treated in Ref. 1 except for $g_{0,2}$, for which it differs by a constant. Since it may be easily verified that j_l and j_{l+2} are themselves orthogonal over the range, the inverse function is not unique to an arbitrary additive constant. Furthermore, if the spherical Bessel functions are replaced by the corresponding expression in cylindrical Bessel functions, Eq. (2) is valid when l is not an integer.

To demonstrate the result of Eq. (2), the method of Ref. 4 will be followed. If two functions $f(x)$ and $g(x)$ satisfy the relation of Eq. (1), then $g(x)$ is given in terms of the Mellin transform of $f(x)$ by[5]

$$g(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} x^{-s}\,ds/F(1 - s), \quad (3)$$

$$F(s) = \int_0^\infty x^{s-1}f(x)\,dx. \quad (4)$$

The path of integration in Eq. (3) must lie in the strip where both $F(s)$ and $F(1 - s)$ are analytic. In terms of cylindrical Bessel functions, using $f(x)$ from Eq. (1), the result is obtained:

$$F(1 - s) = \frac{\pi}{2}\int_0^\infty dx J_{l+\frac{1}{2}}(x)J_{l+m+\frac{1}{2}}(x)x^{-s-1}. \quad (5)$$

This integral is a standard form[6] and yields

$$F(1 - s) = \frac{\pi}{2^{s+2}} \times \frac{\Gamma(s+1)\Gamma(l+\frac{1}{2}m+\frac{1}{2}-\frac{1}{2}s)}{\Gamma(\frac{1}{2}s+\frac{1}{2}m+1)\Gamma(\frac{1}{2}s+l+\frac{1}{2}m+\frac{3}{2})\Gamma(\frac{1}{2}s-\frac{1}{2}m+1)}. \quad (6)$$

Thus, $g_{l,m}$ is given by

$$g_{l,m}(x) = \frac{8}{\pi}\frac{1}{2\pi i}\int ds \times \frac{2^{2s}\Gamma(s+\frac{1}{2}m+1)\Gamma(s+l+\frac{1}{2}m+\frac{3}{2})\Gamma(s-\frac{1}{2}m+1)}{x^{2s}\Gamma(2s+1)\Gamma(l+\frac{1}{2}m+\frac{1}{2}-s)}, \quad (7)$$

where s of Eq. (6) has been replaced by $2s$, and the abscissa of integration lies between $-\frac{1}{2}$ and 0. Regardless of the parity of m, the poles of the first two gamma functions in the numerator are canceled by the poles of $\Gamma(2s + 1)$ in the denominator. The path of integration may be closed by a large semicircle in the left half-plane and the integral over the semicircle tends to zero. The third factor in the numerator has poles at $s = -n - 1 + \frac{1}{2}m$, where n takes on all positive

* This research is supported by the United States Air Force under Project RAND. Views or conclusions contained in this abridgment of RAND Memorandum RM-5886-PR should not be interpreted as representing the official opinion or policy of the United States Air Force.

[1] H. A. Mavromatis and K. Schilcher, J. Math. Phys. **9**, 1627 (1968).

[2] H. Bateman, Proc. London Math. Soc. **4**, 461 (1906).

[3] C. Fox, Proc. London Math. Soc. **29**, 401 (1929).

[4] E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals* (Oxford University Press, London, England, 1948), pp. 212–216.

[5] Reference 4, p. 214.

[6] G. N. Watson, *Treatise on the Theory of Bessel Functions* (The Macmillan Company, New York, 1948), p. 403.

integer values and zero. Those poles, for which $0 \leq n \leq \frac{1}{2}m - 1$, lie to the right of the integration path and do not contribute to the integral. Those poles, for which $\frac{1}{2}m \leq n \leq m - 1$, have vanishing residues. The residues of the remaining poles may be evaluated and the reflection formula $\Gamma(z)\Gamma(1 - z) = \pi/\sin \pi z$ may be used to simplify the expression. Thus,

$$g_{l,m}(x) = 4x^2(-1)^{l+m+1} \sum_{m}^{\infty} \frac{(-1)^n}{n!} \times \frac{\Gamma(2n + 2 - m)(\frac{1}{2}x)^{2n-m}}{\Gamma(n + 1 - m)\Gamma(n + \frac{1}{2} - l - m)\Gamma(n + l + \frac{3}{2})} \quad (8)$$

$$= 4x^2(-1)^{l+1} \sum_{0}^{\infty} \frac{(-1)^n}{n!} \times \frac{\Gamma(2n + m + 2)(\frac{1}{2}x)^{2n+m}}{\Gamma(n + m + 1)\Gamma(n + \frac{1}{2} - l)\Gamma(n + l + m + \frac{3}{2})}. \quad (9)$$

Now we have the general formula[7]

$$J_\mu(x)J_\nu(x) = \sum_{0}^{\infty} \frac{(-1)^n\Gamma(2n + \mu + \nu + 1)(\frac{1}{2}x)^{2n+\mu+\nu}}{n!\,\Gamma(n + \mu + \nu + 1)\Gamma(n + \mu + 1)\Gamma(n + \nu + 1)}. \quad (10)$$

Setting $\mu = -l - \frac{1}{2}$, $\nu = l + m + \frac{1}{2}$, this series becomes

$$J_{-l-\frac{1}{2}}(x)J_{l+m+\frac{1}{2}}(x) = \sum_{0}^{\infty} \frac{(-1)^n\Gamma(2n + m + 1)(\frac{1}{2}x)^{2n+m}}{n!\,\Gamma(n + m + 1)\Gamma(n + \frac{1}{2} - l)\Gamma(n + l + m + \frac{3}{2})}. \quad (11)$$

[7] Reference 6, p. 147.

All the gamma functions in Eqs. (9) and (11) match except the one in the numerator. Multiply Eq. (11) by x and differentiate, and the two series become identical. Thus,

$$g_{l,m}(x) = 4x^2(-1)^{l+1} \frac{d}{dx} xJ_{-l-\frac{1}{2}}(x)J_{l+m+\frac{1}{2}}(x). \quad (12)$$

The identity $N_{l+\frac{1}{2}}(x) = (-1)^{l+1}J_{-l-\frac{1}{2}}(x)$ and the return from cylindrical to spherical Bessel functions now yield Eq. (2). The analysis may be carried through in the same manner if l is not an integer. Only one of the three sets of numerator poles will cancel, and two series of the type of Eq. (9) result. Both may be identified as Bessel function products and combined to yield the form of Eq. (2) in cylindrical Bessel functions.

For large x, Eq. (2) has the asymptotic form

$$g_{l,m}(x) \to \frac{8x^2}{\pi}(-1)^{l+1} \cos\left(2x - \frac{m\pi}{2}\right), \quad (13)$$

which includes all the cases of Ref. 1. For small x, the result is

$$g_{l,m}(x) \to -\frac{(m + 1)\Gamma(l + \frac{1}{2})x^{m+2}}{\pi 2^{m-2}\Gamma(l + m + \frac{3}{2})}. \quad (14)$$

Since this vanishes for $x = 0$, the calculation of the potential from the phase shift, as discussed in Ref. 1, becomes much more practical. The structure of Eq. (2) makes it clear that the inverse function $g_{l,m}(x)$ can always be written in terms of algebraic and trigonometric functions when l and m are both integers. If m is an even positive integer, an arbitrary constant may be added to $g_{l,m}$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Group Theory and Mixed Atomic Configurations*

J. C. MORRISON†
Physics Department, The Johns Hopkins University, Baltimore, Maryland

(Received 20 December 1968)

A group-theoretical scheme is introduced to classify the states of an atomic system having two open shells. States labeled according to this scheme may be written

$$|(l_A + l_B)\alpha Q_A(S_AL_A)J_A \times \beta Q_B(S_BL_B)J_B,\ QM_QJM_J\rangle,$$

where the quasispins Q_A and Q_B are coupled together to form a total quasispin Q. Although these states are, in general, mixtures of different configurations $l_A^n l_B^q$, it is found that they serve as a convenient basis for the calculation of matrix elements in $(l_A + l_B)^N$. The matrix elements of operators between the states of two configurations are obtained from these matrix elements by means of a unitary transformation. As an example matrix elements of the Coulomb interaction within $(f + p)^N$ are calculated.

I. INTRODUCTION

In 1958, Elliott[1] showed that it is possible to classify the states of mixed configurations according to the representations of certain continuous groups. More recently, Feneuille[2,3] has chosen a particular group-theoretical scheme which is useful in the general case. He has also expressed the Coulomb interaction within the configurations $(d + s)^N$ as a sum of seven operators, which have well-defined transformation properties under the operations of these groups, and he has tabulated the matrix elements of these basic operators.

The particular groups which Feneuille chose to classify the states of $(l_A + l_B)^N$ are not compatible with Jj coupling between the shells, which is often the "best" coupling scheme. However, this is not a disadvantage in those cases for which it is necessary to form and to diagonalize the entire energy matrix.

The purpose of this paper is to consider two other possible choices of continuous groups—each of which is compatible with Jj coupling. Operators are introduced which have well-defined transformation properties with respect to these groups and in terms of which any scalar interaction within $(l_A + l_B)^N$ may be expressed. It is found that all of these basic operators may be written as a sum of operators of the form $\{\mathbf{T}_A^{(PK)}\mathbf{U}_B^{(P'K)}\}_{00}^{(R0)}$, where, for instance, $\mathbf{T}_A$ is an operator acting only within the A shell having quasispin rank P and rank in the total spin-orbital space K. $\mathbf{T}_A$ and $\mathbf{U}_B$ are coupled together to form a scalar operator with a total quasispin rank of R.

II. CLASSIFICATION OF STATES

The first group-theoretical scheme which we shall use is, in part,

$$R(8l_A + 8l_B + 8) \supset SU(2) \times SP(4l_A + 4l_B + 4)$$
$$\supset SU(2) \times SP(4l_A + 2) \times SP(4l_B + 2). \quad (1)$$

Because the relevant irreducible representations of the rotation group $R(8l_A + 8l_B + 8)$ are all of the type $(\frac{1}{2}, \frac{1}{2}, \cdots, \pm\frac{1}{2})$, the irreducible representations of $SU(2) \times SP(4l_A + 4l_B + 4)$ may be simply denoted by (Q, v), where Q is the total quasispin and v the total seniority. Similarly, the irreducible representations of $SP(4l_A + 2)$ and $SP(4l_B + 2)$ may be denoted by v_A and v_B, where, for instance, v_A is the seniority in the A shell. The scheme is completed by the reductions

$$SP(4l_A + 2) \times SP(4l_B + 2)$$
$$\supset (SU(2) \times R(2l_A + 1)) \times (SU(2) \times R(2l_B + 1))$$
$$\supset (SU(2) \times R(3)) \times (SU(2) \times R(3))$$
$$\supset R(3) \times R(3) \supset R(3). \quad (2)$$

A state of $(l_A + l_B)^N$ can be described in this scheme by writing

$$|(l_A + l_B)^N Qv(v_A w_A \tau_A (S_AL_A)J_A \times v_B w_B \tau_B (S_BL_B)J_B), JM\rangle, \quad (3)$$

where the usual spectroscopic notation is used. According to the reduction (2), the spin and orbital quantum numbers in the $A(B)$ shell are coupled together to form $J_A(J_B)$. J_A and J_B are then coupled together to form a total J.

Although states labeled according to this scheme have a definite total quasispin, in general they are mixtures of different configurations $l_A^{N_A} l_B^{N_B}$.[4] For instance, the states of $(f + p)^5$ which have a total

* This work was partially supported by the United States Atomic Energy Commission.

† Present address: Argonne National Laboratory, Argonne, Illinois.

[1] J. P. Elliott, Proc. Roy. Soc.(London)**A245**, 128 (1958).

[2] S. Feneuille, J. Phys. **28**, 61 (1967).

[3] S. Feneuille, J. Phys. **28**, 315 (1967).

[4] R. D. Lawson and M. H. Macfarlane, Nucl. Phys. **66**, 80 (1965).

quasispin $\frac{7}{2}$ and seniorities 2 and 1 in the f and p shells are linear combinations of states of f^4p and f^2p^3. We shall find, however, that the states of this scheme serve as a very convenient basis for the calculation of matrix elements of operators acting within $(l_A + l_B)^N$. The matrix elements of operators between the states of a single configuration or joining the states of two different configurations may be obtained from these matrix elements by means of a unitary transformation.

In order to define this unitary transformation, we introduce a second group-theoretical scheme for which the representation labels are compatible with the configuration labels N_A and N_B. We write

$$\begin{aligned} R(8l_A + 8l_B + 8) &\supset R(8l_A + 4) \times R(8l_B + 4) \\ &\supset (SU(2) \times SP(4l_A + 2)) \\ &\quad \times (SU(2) \times SP(4l_B + 2)). \end{aligned} \quad (4)$$

Irreducible representations of

$$(SU(2) \times SP(4l_A + 2)) \times (SU(2) \times SP(4l_B + 2))$$

are denoted by $(Q_A, v_A) \times (Q_B, v_B)$, where, for instance, (Q_A, v_A) is the quasispin and seniority in the A shell. We again complete the scheme with the reductions (2) and label our states

$$\begin{aligned} &|l_A^{N_A} l_B^{N_B} Q_A v_A w_A \tau_A (S_A L_A) J_A \\ &\qquad \times Q_B v_B w_B \tau_B (S_B L_B) J_B, JM\rangle. \end{aligned} \quad (5)$$

The generators of the continuous groups which appear in these two schemes may be found elsewhere.[5]

For eigenstates of the first scheme,

$$M_Q = -\tfrac{1}{2}(2l_A + 2l_B + 2 - N).$$

N is thus related to the azimuthal quantum number of total quasispin. Similarly, for eigenstates of the second scheme,

$$\begin{aligned} M_Q^A &= -\tfrac{1}{2}(2l_A + 1 - N_A), \\ M_Q^B &= -\tfrac{1}{2}(2l_B + 1 - N_B). \end{aligned}$$

The phases of the eigenstates of the first scheme can be chosen so that the two sets of states are related by the equation

$$\begin{aligned} &|(l_A + l_B)^N Q v (v_A w_A \tau_A (S_A L_A) J_A \\ &\qquad \times v_B w_B \tau_B (S_B L_B) J_B), JM\rangle \\ &= \sum_{M_{Q^A} M_{Q^B}} (Q_A M_Q^A Q_B M_Q^B \mid Q_A Q_B Q M_Q) \\ &\qquad \times |l_A^{N_A} l_B^{N_B} Q_A v_A w_A \tau_A (S_A L_A) J_A \\ &\qquad \times Q_B v_B w_B \tau_B (S_B L_B) J_B, JM\rangle. \end{aligned} \quad (6)$$

We can thus pass from one scheme to the other by either coupling or uncoupling in the quasispin space of A and B particles. Accordingly, we shall hence-forth label a state within the first scheme as

$$\begin{aligned} &|(l_A + l_B) Q_A v_A w_A \tau_A (S_A L_A) J_A \\ &\qquad \times Q_B v_B w_B \tau_B (S_B L_B) J_B, QM_Q JM_Q\rangle, \end{aligned} \quad (7)$$

thereby making its quasispin structure explicit.

The idea of coupling quasispins was first introduced by Kerman *et al.*[6] in dealing with a nuclear pairing interaction. The simplified pairing Hamiltonian which they used could be written in terms of the quasispin generators and so the technique was very convenient. However, atomic interactions or more realistic nuclear interactions, which involve position, momentum, and spin operators, do not seem to have been dealt with in this way.

III. CLASSIFICATION OF TENSOR OPERATORS

We denote by $a_\alpha^\dagger$ and $b_\beta^\dagger$ creation operators for l_A and l_B electrons, respectively. Following Judd,[7] we introduce special absorption operators $\tilde{a}_\alpha$ and $\tilde{b}_\beta$, where, for instance, $\tilde{a}_\alpha$ is related to the adjoint of $a_\alpha^\dagger$ by the equation

$$\tilde{a}_{m_s m_l} = (-)^{s + l_A - m_s - m_l} (a^\dagger_{-m_s - m_l})^\dagger.$$

The operators $a_\alpha^\dagger$ and $\tilde{a}_\alpha$ have a spin rank of s and an orbital rank of l_A. They transform according to the $(10\cdots0)$ representation of $SP(4l_A + 2)$ and

$$SP(4l_A + 4l_B + 4)$$

and according to the $(0\cdots0)$ representation of $SP(4l_B + 2)$. The operators $b_\beta^\dagger$ and $\tilde{b}_\beta$ have analogous transformation properties.

Having assigned ranks to the a's and b's, we can form other tensor operators by coupling them together in various ways. In particular, we can write

$$\begin{aligned} \mathbf{W}^{(\kappa k)}(l_a, l_a) &= -(a^\dagger \tilde{a})^{(\kappa k)}, \\ \mathbf{W}^{(\kappa k)}(l_a, l_b) &= -(a^\dagger \tilde{b})^{(\kappa k)}, \\ \mathbf{W}^{(\kappa k)}(l_b, l_a) &= -(b^\dagger \tilde{a})^{(\kappa k)}, \\ \mathbf{W}^{(\kappa k)}(l_b, l_b) &= -(b^\dagger \tilde{b})^{(\kappa k)}. \end{aligned}$$

The transformation properties of the operators $\mathbf{W}^{(\kappa k)}(l_a, l_a)$ and $\mathbf{W}^{(\kappa k)}(l_b, l_b)$ under the operations of the groups $SP(4l_A + 2) \times SP(4l_B + 2)$ and $SP(4l_A + 4l_B + 4)$ are easily obtained. They have been given, for instance, by Judd.[8] Since the operators $\mathbf{W}^{(\kappa k)}(l_a, l_b)$ and $\mathbf{W}^{(\kappa k)}(l_b, l_a)$ are both bilinear in the a's and b's, they transform according to the $(10\cdots0) \times (10\cdots0)$ representation of $SP(4l_A + 2) \times SP(4l_B + 2)$ and we can introduce linear combinations of them which also correspond to definite representations of $SP(4l_A + 4l_B + 4)$. We write

$$\mathbf{U}^{(\kappa k)}(l_a, l_b) = -\{(a^\dagger \tilde{b})^{(\kappa k)} + (\tilde{a} b^\dagger)^{(\kappa k)}\}, \quad (8)$$

$$\mathbf{G}^{(\kappa k)}(l_a, l_b) = -\{(a^\dagger \tilde{b})^{(\kappa k)} - (\tilde{a} b^\dagger)^{(\kappa k)}\}. \quad (9)$$

[5] B. R. Judd, "Group Theory in Atomic Spectroscopy," in *Group Theory and Its Applications*, E. M. Loebl, Ed. (Academic Press Inc., New York, 1968); H. T. Wadzinski, thesis, The Johns Hopkins University, 1968.

[6] A. K. Kerman, R. D. Lawson, and M. H. Macfarlane, Phys. Rev. **124**, 162 (1961).

[7] B. R. Judd, *Second Quantization and Atomic Spectroscopy* (The Johns Hopkins Press, Baltimore, Md., 1967).

[8] B. R. Judd, Physica **33**, 174 (1967).

TABLE I. Classification of tensor operators.

Operator	Representation $SP(4l_A+2)\times SP(4l_B+2)$	Representation $SP(4l_A+4l_B+4)$
$\mathbf{W}^{(\kappa k)}(l_a,l_a)$ ($\kappa+k$ even)	$(110\cdots0)\times(0\cdots0)$	$(110\cdots0)$
$\mathbf{W}^{(\kappa k)}(l_a,l_a)$ ($\kappa+k$ odd)	$(20\cdots0)\times(0\cdots0)$	$(20\cdots0)$
$\mathbf{W}^{(\kappa k)}(l_a,l_a)$ ($\kappa+k=0$)	$(0\cdots0)\times(0\cdots0)$	$(0\cdots0)$
$\mathbf{W}^{(\kappa k)}(l_b,l_b)$ ($\kappa+k$ even)	$(0\cdots0)\times(110\cdots0)$	$(110\cdots0)$
$\mathbf{W}^{(\kappa k)}(l_b,l_b)$ ($\kappa+k$ odd)	$(0\cdots0)\times(20\cdots0)$	$(20\cdots0)$
$\mathbf{W}^{(\kappa k)}(l_b,l_b)$ ($\kappa+k=0$)	$(0\cdots0)\times(0\cdots0)$	$(0\cdots0)$
$\mathbf{U}^{(\kappa k)}(l_a,l_b)$	$(10\cdots0)\times(10\cdots0)$	$(110\cdots0)$
$\mathbf{G}^{(\kappa k)}(l_a,l_b)$	$(10\cdots0)\times(10\cdots0)$	$(20\cdots0)$

The operators $\mathbf{U}^{(\kappa k)}$ are linear combinations of the operators $a^\dagger_\alpha\tilde{b}_\beta - b^\dagger_\beta\tilde{a}_\alpha$ and, hence, belong to the antisymmetric part of the direct product representation $(10\cdots0)\boxtimes(10\cdots0)$ of $SP(4l_A+4l_B+4)$, which is $(110\cdots0)+(0\cdots0)$. Since it is not possible to construct a scalar with respect to $SP(4l_A+4l_B+4)$ from the $\mathbf{U}^{(\kappa k)}$, these operators transform according to the $(110\cdots0)$ representation of that group. Similarly, the $\mathbf{G}^{(\kappa k)}$ transform according to the symmetric part of the direct product representation, and hence belong to the $(20\cdots0)$ representation of $SP(4l_A+4l_B+4)$.

The anticommutation relations between $\tilde{a}$ and $b^\dagger$ lead to expressions for $\mathbf{U}^{(\kappa k)}$ and $\mathbf{G}^{(\kappa k)}$ in terms of the $\mathbf{W}^{(\kappa k)}(l_a,l_b)$ and $\mathbf{W}^{(\kappa k)}(l_b,l_a)$:

$$\mathbf{U}^{(\kappa k)}(l_a,l_b) = \mathbf{W}^{(\kappa k)}(l_a,l_b) + (-)^{l_A+l_B+\kappa+k}\mathbf{W}^{(\kappa k)}(l_b,l_a), \quad (10)$$

$$\mathbf{G}^{(\kappa k)}(l_a,l_b) = \mathbf{W}^{(\kappa k)}(l_a,l_b) - (-)^{l_A+l_B+\kappa+k}\mathbf{W}^{(\kappa k)}(l_b,l_a). \quad (11)$$

If l_A and l_B have the same parity, $\mathbf{U}^{(\kappa k)}(l_a,l_b)$ and $\mathbf{G}^{(\kappa k)}(l_a,l_b)$ are identical to the operators $\mathbf{W}^{+(\kappa k)}(l_a,l_b)$ and $\mathbf{W}^{-(\kappa k)}(l_a,l_b)$ introduced by Feneuille.[3] The complete classification of the operators $\mathbf{W}^{(\kappa k)}(l_a,l_a)$, $\mathbf{W}^{(\kappa k)}(l_b,l_b)$, $\mathbf{U}^{(\kappa k)}(l_a,l_b)$, and $\mathbf{G}^{(\kappa k)}(l_a,l_b)$ with respect to the groups $SP(4l_A+2)\times SP(4l_B+2)$ and $SP(4l_A+4l_B+4)$ is given in Table I.

IV. DECOMPOSITION OF TWO-PARTICLE SCALAR OPERATORS

Of the possible coupled products which we can form with pairs of tensor operators, the products

$$\{\mathbf{W}^{(k\kappa)}(l_a,l_a)\mathbf{W}^{(\kappa'k')}(l_a,l_a)\}^{(tt)0}$$

and

$$\{\mathbf{W}^{(\kappa k)}(l_b,l_b)\mathbf{W}^{(\kappa'k')}(l_b,l_b)\}^{(tt)0}$$

act entirely within a single atomic shell. Since methods for resolving operators of this kind into group-theoretical parts are already available, we shall not consider them here. Of the other possible products,

$$\{\mathbf{U}^{(\kappa k)}(l_a,l_b)\mathbf{U}^{(\kappa'k')}(l_a,l_b)\}^{(tt)0}$$

and

$$\{\mathbf{W}^{(\kappa k)}(l_a,l_a)\mathbf{U}^{(\kappa'k')}(l_a,l_b)\}^{(tt)0},$$

with $\kappa+k$ even, odd, or zero, are typical examples. Each of these sets transforms according to a direct product representation of $SP(4l_A+2)\times SP(4l_B+2)$. It is these representations that we wish to resolve into their irreducible parts.

We make use of a theorem of Racah[9] to factor the appropriate Clebsch–Gordan (CG) coefficients into two parts:

$$\langle((v_aw_as_al_a\times v_bw_bs_bl_b)\kappa k,(\bar{v}_a\bar{w}_a\bar{s}_a\bar{l}_a\times\bar{v}_b\bar{w}_b\bar{s}_b\bar{l}_b)\kappa'k',tt,0|\ V_aW_a(\kappa_ak_a)K\times V_bW_b(\kappa_bk_b)K,0\rangle$$
$$=\langle((s_al_a)(s_bl_b))\kappa k,((\bar{s}_a\bar{l}_a)(\bar{s}_b\bar{l}_b))\kappa'k',tt,0\,|\,(s_a\bar{s}_a)\kappa_a(l_a\bar{l}_a)k_a,K;(s_b\bar{s}_b)\kappa_b(l_b\bar{l}_b)k_b,K;0\rangle$$
$$\times\langle(v_aw_as_al_a+\bar{v}_a\bar{w}_a\bar{s}_a\bar{l}_a\,|\,V_aW_a\kappa_ak_a)(v_bw_bs_bl_b+\bar{v}_b\bar{w}_b\bar{s}_b\bar{l}_b\,|\,V_bW_b\kappa_bk_b)\rangle. \quad (12)$$

The first part, which is itself a CG coefficient for $R_3\times R_3$, may be expanded as a sum of products of 3-j symbols. The resulting expression can then be simplified using the graphical techniques of Jucys *et al.*[10] to obtain

$$\langle((s_al_a)(s_bl_b))\kappa k,((\bar{s}_a\bar{l}_a)(\bar{s}_b\bar{l}_b))\kappa'k',tt,0\,|\,(s_a\bar{s}_a)\kappa_a(l_a\bar{l}_a)k_a,K;(s_b\bar{s}_b)\kappa_b(l_b\bar{l}_b)k_b,K;0\rangle$$
$$=(-1)^{\kappa_b+k_a+K+t}\{[\kappa][k][\kappa'][k'][t][\kappa_a][k_a][\kappa_b][k_b][K]\}^{\frac{1}{2}}\begin{Bmatrix}\kappa_a&\kappa_b&t\\k_b&k_a&K\end{Bmatrix}\begin{pmatrix}s_a&s_b&\kappa\\\bar{s}_a&\bar{s}_b&\kappa'\\\kappa_a&\kappa_b&t\end{pmatrix}\begin{pmatrix}l_a&l_b&k\\\bar{l}_a&\bar{l}_b&k'\\k_a&k_b&t\end{pmatrix}. \quad (13)$$

We can now form operators which are linear combinations of the $\{\mathbf{U}^{(\kappa k)}(l_a,l_b)\mathbf{U}^{(\kappa'k')}(l_a,l_b)\}^{(tt)0}$ and which

[9] G. Racah, Phys. Rev. 76, 1352 (1949).
[10] A. P. Jucys, I. B. Levinson, and V. V. Vanagas, *Mathematical Apparatus of the Theory of Angular Momentum* (Israel Program for Scientific Translations, Jerusalem, 1962).

correspond to single representations of

$$SP(4l_A + 2) \times SP(4l_B + 2).$$

We write

$$\eta((\kappa_a k_a)K, (\kappa_b k_b)K, 0) = \sum_{\substack{\kappa k \\ \kappa' k' \\ t}} \langle((sl_A)(sl_B))\kappa k, ((sl_A)(sl_B))\kappa' k', tt, 0 \mid (ss)\kappa_a(l_A l_A)k_a, K; (ss)\kappa_b(l_B l_B)k_b, K; 0\rangle \times \{\mathbf{U}^{(\kappa k)}(l_a, l_b)\mathbf{U}^{(\kappa' k')}(l_a, l_b)\}^{(tt)0}. \quad (14)$$

Using Eq. (8) for the $\mathbf{U}^{(\kappa k)}$ and explicitly carrying out the recoupling, we obtain expressions for η in the following cases:

(1) $\kappa_a + k_a$ and $\kappa_b + k_b$ both even:

$$\eta((\kappa_a k_a)K, (\kappa_b k_b)K, 0) = -(\tfrac{8}{3})^{\frac{1}{2}}\{\mathbf{A}^{(1(\kappa_a k_a)K)}\mathbf{B}^{(1(\kappa_b k_b)K)}\}_{00}^{(20)} - (\tfrac{1}{3})^{\frac{1}{2}} \times \{\mathbf{A}^{(1(\kappa_a k_a)K)}\mathbf{B}^{(1(\kappa_b k_b)K)}\}_{00}^{(00)}; \quad (15)$$

(2) $\kappa_a + k_a$ and $\kappa_b + k_b$ both odd:

$$\eta((\kappa_a k_a)K, (\kappa_b k_b)K, 0) = \{\mathbf{A}^{(0(\kappa_a k_a)K)}\mathbf{B}^{(0(\kappa_b k_b)K)}\}_{00}^{(00)}; \quad (16)$$

(3) $\kappa_a + k_a$ and $\kappa_b + k_b$ of opposite parity:

$$\eta((\kappa_a k_a)K, (\kappa_b k_b)K, 0) = 0; \quad (17)$$

(4) $\kappa_a = k_a = \kappa_b = k_b = 0$:

$$\eta^{(2)}((\kappa_a k_a)K, (\kappa_b k_b)K, 0) = -(\tfrac{8}{3})^{\frac{1}{2}}\{\mathbf{A}^{(1(00)0)}\mathbf{B}^{(1(00)0)}\}_{00}^{(20)} - (\tfrac{1}{3})^{\frac{1}{2}}\{\mathbf{A}^{(1(00)0)}\mathbf{B}^{(1(00)0)}\}_{00}^{(00)} - ([l_B]/[l_A])^{\frac{1}{2}}N_A - ([l_A]/[l_B])^{\frac{1}{2}}N_B + ([l_A][l_B])^{\frac{1}{2}},$$

$$\eta^{(1)}((\kappa_a k_a)K, (\kappa_b k_b)K, 0) = ([l_B]/[l_A])^{\frac{1}{2}}N_A + ([l_A]/[l_B])^{\frac{1}{2}}N_B. \quad (18)$$

Here $\mathbf{A}^{(P\kappa k)}$ is identical to the $\mathbf{X}^{(P\kappa k)}(l_a, l_a)$ used by Feneuille,[2] and $\mathbf{B}^{(P\kappa k)}$ is identical to $\mathbf{X}^{(P\kappa k)}(l_b, l_b)$. In cases (1) and (2), η is a pure two-body operator. However, for $\kappa_a = k_a = \kappa_b = k_b = 0$, η has both one-body and two-body parts. We have denoted these as $\eta^{(1)}$ and $\eta^{(2)}$.

Similarly, we form those linear combinations of the operators, $\{\mathbf{W}^{(\kappa k)}(l_a, l_a)\mathbf{U}^{(\kappa' k')}(l_a, l_b)\}^{(tt)0}$ with $\kappa + k$ even, which correspond to definite representations of $SP(4l_A + 2) \times SP(4l_B + 2)$:

$$\zeta(VW(\kappa_a k_a)K, (10\cdots0)^2(10\cdots0)l_B K, 0) = \sum_{\kappa+k \text{ even}} ((110\cdots0)w\kappa k + (10\cdots0)^2(10\cdots0)l_A \mid VW\kappa_a k_a) \times \sum_{\substack{\kappa' k' \\ t}} \langle((\kappa k)(00))\kappa k, ((sl_A)(sl_B))\kappa' k', tt, 0 \mid (\kappa s)\kappa_a(kl_A)k_a, K; (0s)s(0l_B)l_B, K; 0\rangle \times \{\mathbf{W}^{(\kappa k)}(l_a, l_a)\mathbf{U}^{(\kappa' k')}(l_a, l_b)\}^{(tt)0}. \quad (19)$$

Again, using Eq. (8) and carrying out the recoupling, we obtain

$$\zeta(VW(\kappa_a k_a)K, (10\cdots0)^2(10\cdots0)l_B K, 0) = \sum_{\kappa+k \text{ even}} ((110\cdots0)w\kappa k + l_A \{\mid VW\kappa_a k_a) \times [(\tfrac{2}{3})^{\frac{1}{2}}\{(\mathbf{A}^{(1\kappa k)}\mathbf{a})^{(\frac{3}{2}(\kappa_a k_a)K)}\mathbf{b}^{(\frac{1}{2}K)}\}_{00}^{(20)} - (\tfrac{1}{3})^{\frac{1}{2}}\{(\mathbf{A}^{(1\kappa k)}\mathbf{a})^{(\frac{1}{2}(\kappa_a k_a)K)}\mathbf{b}^{(\frac{1}{2}K)}\}_{00}^{(00)}]. \quad (20)$$

$\mathbf{a}_{m_q\alpha}$ denotes the triple tensor whose $m_q = +\frac{1}{2}$ component is $a_\alpha^\dagger$ and whose $m_q = -\frac{1}{2}$ component is $\tilde{a}_\alpha$. Similarly, $\mathbf{b}^{(\frac{1}{2}K)}$ is obtained from the triple tensor $\mathbf{b}_{m_q\beta}$ by coupling the spin and orbital components to form a tensor of rank K in the total spin-orbital space.

For the special values $\kappa_a = \frac{1}{2}$, $k_a = l_A$, ζ has a one-particle part given by

$$\zeta^{(1)}((10\cdots0)^2(10\cdots0)l_A K, (10\cdots0)^2(10\cdots0)l_B K, 0) = \sum_{\kappa+k \text{ even}} ([\kappa][k]/[s][l_A])^{\frac{1}{2}}((110\cdots0)w\kappa k + l_A \{\mid V^2 W l_A)\{a^{\dagger(K)}\tilde{b}^{(K)}\}^{(0)}.$$

For all other values of κ_a and k_a, ζ is a pure two-body operator.

A similar analysis may be applied to the cases $\kappa + k$ odd and $\kappa + k = 0$, and to all other possible coupled pairs of tensors. Hence, any scalar two-body operator within $(l_A + l_B)^N$ can be expressed as a sum of operators of the form $\{\mathbf{T}_A^{(PK)}\mathbf{U}_B^{(P')K}\}_{00}^{(R0)}$—in addition to certain one-body operators which may be easily dealt with.

V. MATRIX ELEMENTS

We have seen that it is possible to label the states of $(l_A + l_B)^N$,

$$|(l_A + l_B)\alpha Q_A(S_A L_A)J_A \times \beta Q_B(S_B L_B)J_B, QM_Q JM_J\rangle,$$

and to express operators as linear combinations of tensors of the type

$$\{\mathbf{T}_A^{(PB)}\mathbf{U}_B^{(P'K')}\}^{(RT)}.$$

A generalization of Eq. (7.1.5) of Edmonds[11] yields

$$\langle((l_A + l_B)\alpha Q_A(S_AL_A)J_A \times \beta Q_B(S_BL_B)J_B, QJ\| \{\mathbf{T}_A^{(PK)}\mathbf{U}_B^{(P'K')}\}^{(RT)} \times \|(l_A + l_B)\alpha' Q'_A(S'_AL'_A)J'_A \times \beta' Q'_B(S'_BL'_B)J'_B, Q'J'\rangle$$
$$= (-)^{2Qx}\{[Q][R][Q'][J][T][J']\}^{\frac{1}{2}}\begin{Bmatrix} Q_A & Q'_A & P \\ Q_B & Q'_B & P' \\ Q & Q' & R \end{Bmatrix}\begin{Bmatrix} J_A & J'_A & K \\ J_B & J'_B & K' \\ J & J' & T \end{Bmatrix}$$
$$\times \langle\alpha Q_A(S_AL_A)J_A\| \mathbf{T}_A^{(PK)} \|\alpha' Q'_A(S'_AL'_A)J'_A\rangle\langle\beta Q_B(S_BL_B)J_B\| \mathbf{U}_B^{(P'K')} \|\beta' Q'_B(S'_BL'_B)J'_B\rangle, \quad (21)$$

where x is equal to the number of creation and absorption operators in $\mathbf{T}_A^{(PK)}$.

The reduced matrix elements of $\mathbf{a}^{(\frac{1}{2}K)}$ and $\mathbf{A}^{P((\kappa k)K)}$ for $v_A \geq v'_A$ are given by the equations

$$(l_A\alpha Q_A(S_AL_A)J_A \|\mathbf{a}^{(\frac{1}{2}K)}\| l_A\alpha' Q'_A(S'_AL'_A)J'_A)$$
$$= -r^{-1}\{v_A[S_A][L_A][J_A][K][J'_A]\}^{\frac{1}{2}}\begin{Bmatrix} S_A & S'_A & \frac{1}{2} \\ L_A & L'_A & l_A \\ J_A & J'_A & K \end{Bmatrix}(l_A^{v_A}\alpha Q_AS_AL_A\{|\, l_A^{v_A-1}\alpha' Q'_AS'_AL'_A), \quad (22)$$

$$(l_A\alpha Q_A(S_AL_A)J_A \|\mathbf{A}^{(P(\kappa k)K)}\| l_A\alpha' Q'_A(S'_AL'_A)J'_A)$$
$$= (-1)^{2Q_A+1}s^{-1}\{2[J_A][K][J'_A]\}^{\frac{1}{2}}\begin{Bmatrix} S_A & S'_A & \kappa \\ L_A & L'_A & k \\ J_A & J'_A & K \end{Bmatrix}(l_A^{v_A}\alpha Q_AS_AL_A \|\mathbf{W}^{(\kappa k)}\| l_A^{v_A}\alpha' Q'_AS'_AL'_A), \quad \text{for } \kappa + k \neq 0, \quad (23)$$

$$(l_A\alpha Q_A(S_AL_A)J_A \|\mathbf{A}^{(1(00)0)}\| l_A\alpha' Q_A(S_AL_A)J_A) = -2\{Q_A(Q_A + 1)(2Q_A + 1)(2J_A + 1)/(2l_A + 1)\}^{\frac{1}{2}}, \quad (24)$$

where

$$r = \begin{pmatrix} Q_A & \frac{1}{2} & Q'_A \\ Q_A & \frac{1}{2} & -Q'_A \end{pmatrix} \quad \text{and} \quad s = \begin{pmatrix} Q_A & P & Q'_A \\ Q_A & 0 & -Q_A \end{pmatrix}.$$

All of the other operators $\mathbf{T}_A^{P(K)}$, which we encountered in the previous section, were obtained by coupling together the operators $\mathbf{a}^{(\frac{1}{2}K)}$ and $\mathbf{A}^{(P(\kappa k)K)}$. Their matrix elements may be evaluated by the usual tensor methods.

The reduced matrix elements of the $\mathbf{U}_B^{(PK)}$ may, of course, be obtained from these equations also.

VI. COULOMB INTERACTION IN $(f + p)^N$

As an example of the theory, we consider the Coulomb interaction in the $(f + p)$ shell. The Coulomb Hamiltonian is given by

$$H = \sum_{i>j}\{14F_0(f)\mathbf{W}_i^{(00)}(f,f)\cdot\mathbf{W}_j^{(00)}(f,f)$$
$$+ 6F_0(p)\mathbf{W}_i^{(00)}(p,p)\cdot\mathbf{W}_j^{(00)}(p,p) + 2(21)^{\frac{1}{2}}F_0(f,p)$$
$$\times(\mathbf{W}_i^{(00)}(f,f)\cdot\mathbf{W}_j^{(00)}(p,p) + \mathbf{W}_i^{(00)}(p,p)\cdot\mathbf{W}_j^{(00)}(f,f))$$
$$+ 168F_2(f)\mathbf{W}_i^{(02)}(f,f)\cdot\mathbf{W}_j^{(02)}(f,f) + 12F_2(p)\mathbf{W}_i^{(02)}(p,p)\cdot\mathbf{W}_j^{(02)}(p,p)$$
$$+ 28(14)^{\frac{1}{2}}F_2(f,p)(\mathbf{W}_i^{(02)}(f,f)\cdot\mathbf{W}_j^{(02)}(p,p) + \mathbf{W}_i^{(02)}(p,p)\cdot\mathbf{W}_j^{(02)}(f,f))$$
$$+ 308F_4(f)\mathbf{W}_i^{(04)}(f,f)\cdot\mathbf{W}_j^{(04)}(f,f) + 1848F_6(f)$$
$$\times\mathbf{W}_i^{(06)}(f,f)\cdot\mathbf{W}_j^{(06)}(f,f) + 126G_2(f,p)\mathbf{U}_i^{(02)}(f,p)\cdot\mathbf{U}_j^{(02)}(f,p)$$
$$- 56G_4(\ ,p)\mathbf{U}_i^{(04)}(f,p)\cdot\mathbf{U}_j^{(04)}(f,p) - 4(21)^{\frac{1}{2}}H_2(f,p)$$
$$\times(\mathbf{W}_i^{(02)}(f,f)\mathbf{U}_j^{(02)}(f,p) + \mathbf{U}_i^{(02)}(f,p)\cdot\mathbf{W}_j^{(02)}(f,f))$$
$$- 6(6)^{\frac{1}{2}}K_2(f,p)(\mathbf{W}_i^{(02)}(p,p)\mathbf{U}_j^{(02)}(f,p) + \mathbf{U}_i^{(02)}(f,p)\cdot\mathbf{W}_j^{(02)}(p,p))$$
$$- 4(462)^{\frac{1}{2}}H_4(f,p)(\mathbf{W}_i^{(04)}(f,f)\cdot\mathbf{U}_j^{(04)}(f,p) + \mathbf{U}_i^{(04)}(f,p)\cdot\mathbf{W}_j^{(04)}(f,f))\},$$

where we have used the integrals F_k and G_k defined by Condon and Shortley[12] and in addition set

$$F_0(f,p) = F^{(0)}(f,p),$$
$$H_2(f,p) = R^2(ff,fp)/25,$$
$$H_4(f,p) = R^4(ff,fp)/297, \quad (25)$$
$$K_2(f,p) = R^2(fp,pp)/25.$$

H contains terms which act entirely within one atomic shell and also terms of the kind

$$\sum_{i>j}\mathbf{W}_i^{(0r)}(f,f)\cdot\mathbf{W}_j^{(0r)}(p,p),$$
$$\sum_{i>j}\mathbf{U}_i^{(0r)}(f,p)\cdot\mathbf{U}_j^{(0r)}(f,p),$$

[11] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, N.J., 1960).

[12] E. U. Condon and G. H. Shortley, *The Theory of Atomic Spectra* (Cambridge University Press, New York, 1935).

and

$$\sum_{i>j} \mathbf{W}_i^{(0r)}(f,f)\cdot\mathbf{U}_j^{(0r)}(f,p).$$

Using Eqs. (13) and (14), we find that the second of these "interaction" terms is related to the η's by the equation

$$\sum_{i>j}\mathbf{U}_i^{(0r)}(f,p)\cdot\mathbf{U}_j^{(0r)}(f,p) = (\tfrac{1}{4})\sum_{\kappa k K}(-)^{k+r}[r]([K])^{\frac{1}{2}}\begin{Bmatrix}3 & 3 & k\\ 1 & 1 & r\end{Bmatrix} \times \eta^{(2)}((\kappa k)K,(\kappa k)K,0). \quad (26)$$

Similarly, Eqs. (13) and (19) yield

$$\sum_{i>j}\mathbf{W}_i^{(0r)}(f,f)\cdot\mathbf{U}_j^{(0r)}(f,p) = -(\tfrac{1}{4})[r]\begin{Bmatrix}3 & 1 & r\\ 1 & 3 & 1\end{Bmatrix} \times((110\cdots0)^1 r + f\{|VW^2P)\sum_K([K])^{\frac{1}{2}} \times \zeta^{(2)}(VW^2PK,(10\cdots0)^2PK,0), \quad \text{for } r \text{ even.} \quad (27)$$

The operator $\sum_{i>j}\mathbf{W}_i^{(0r)}(f,f)\cdot\mathbf{W}_j^{(0r)}(p,p)$ is itself the two-body part of an operator having definite transformation properties with respect to $SP(4l_A+2)\times SP(4l_B+2)$, and it may easily be expressed as a sum of operators of the form $\{\mathbf{A}^{(PK)}\mathbf{B}^{(PK)}\}^{(R0)}$.

To illustrate our method in detail, we calculate the matrix elements of $\sum_{i>j}\mathbf{U}_i^{(02)}(f,p)\cdot\mathbf{U}_j^{(02)}(f,p)$ between the states $|f^4p_{\frac{5}{2}}{}^3H_6\times 1^2P_{\frac{3}{2}},{}^{1\,5}_{-\frac{1}{2}}\rangle$ and $|f^2p^3_{\frac{5}{2}}{}^3H_6\times 1^2P_{\frac{3}{2}},{}^{1\,5}_{-\frac{1}{2}}\rangle$ for which $Q_A=\frac{5}{2}$ and $Q_B=1$. According to our method, we will relate these states to states of definite total quasispin, carry out the calculation of matrix elements within the total quasispin scheme, and, finally, obtain the desired matrix elements by means of a unitary transformation.

The transformation equations relating the two sets of states may be written

$$|(f+p)^3_{\frac{5}{2}}H_6\times 1^2P_{\frac{3}{2}},{}^{5}_{\frac{5}{2}}{}^{1\,5}_{-\frac{1}{2}}\rangle = (\tfrac{5}{2}-\tfrac{3}{2}\,1-1\,|\,\tfrac{5}{2}\,1\,\tfrac{5}{2}-\tfrac{5}{2})\,|f^4p^3_{\frac{5}{2}}H_6\times 1^2P_{\frac{3}{2}},{}^{1\,5}_{-\frac{1}{2}}\rangle + (\tfrac{5}{2}-\tfrac{5}{2}\,1\,0\,|\,\tfrac{5}{2}\,1\,\tfrac{5}{2}-\tfrac{5}{2})\,|f^2p^3_{\frac{5}{2}}{}^3H_6\times 1^2P_{\frac{3}{2}},{}^{1\,5}_{-\frac{1}{2}}\rangle, \quad (28)$$

$$|(f+p)^3_{\frac{5}{2}}H_6\times 1^2P_{\frac{3}{2}},{}^{7}_{\frac{5}{2}}{}^{1\,5}_{-\frac{1}{2}}\rangle = (\tfrac{5}{2}-\tfrac{3}{2}\,1-1\,|\,\tfrac{5}{2}\,1\,\tfrac{7}{2}-\tfrac{5}{2})\,|f^4p^3_{\frac{5}{2}}H_6\times 1^2P_{\frac{3}{2}},{}^{1\,5}_{-\frac{1}{2}}\rangle + (\tfrac{5}{2}-\tfrac{5}{2}\,1\,0\,|\,\tfrac{5}{2}\,1\,\tfrac{7}{2}-\tfrac{5}{2})\,|f^2p^3_{\frac{5}{2}}{}^3H_6\times 1^2P_{\frac{3}{2}},{}^{1\,5}_{-\frac{1}{2}}\rangle. \quad (29)$$

We shall, henceforward, denote the original states simply as $|1\rangle$ and $|2\rangle$, respectively, and make the quasispin structure of the new basis states explicit by denoting them as $|(\frac{5}{2}\,1)^{\frac{5}{2}}\rangle$ and $|(\frac{5}{2}\,1)^{\frac{7}{2}}\rangle$. Then, Eqs. (28) and (29) become simply

$$|(\tfrac{5}{2}\,1)Q\rangle = \sum_{P=1}^{2}U_{QP}\,|P\rangle, \quad (30)$$

where

$$U=\begin{pmatrix}(\frac{2}{7})^{\frac{1}{2}} & -(\frac{5}{7})^{\frac{1}{2}}\\ (\frac{5}{7})^{\frac{1}{2}} & (\frac{2}{7})^{\frac{1}{2}}\end{pmatrix}.$$

The matrix elements of $\sum_{i>j}\mathbf{U}_i^{(02)}(f,p)\cdot\mathbf{U}_j^{(02)}(f,p)$ in the total quasispin scheme are obtained from the corresponding matrix elements of

$$\{\mathbf{A}^{(P(\kappa k)K)}\mathbf{B}^{(P(\kappa k)K)}\}^{(R0)}_{00}$$

by using Eqs. (15), (16), (18), and (26). This calculation is greatly simplified if the matrix elements of

$$\{\mathbf{A}^{(P(\kappa k)K)}\mathbf{B}^{(P(\kappa k)K)}\}^{(R0)}$$

for $\kappa+k\neq 0$ are immediately factored in the following way:

$$\langle(\tfrac{5}{2}\,1)Q|\,\{\mathbf{A}^{(P(\kappa k)K)}\mathbf{B}^{(P(\kappa k)K)}\}^{(R0)}_{00}\,|(\tfrac{5}{2}\,1)Q'\rangle = C(\kappa,k,K)D_{QQ'}(P,R), \quad \text{for} \quad \kappa+k\neq 0.$$

The Wigner–Eckart theorem together with Eqs. (21) and (23) yield the following expressions for these quantities:

$$C(\kappa,k,K)=(-1)^{\kappa+k+K}[\kappa][k]([K])^{\frac{1}{2}}\begin{Bmatrix}1 & 1 & \kappa\\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2}\end{Bmatrix}\begin{Bmatrix}5 & 5 & k\\ 3 & 3 & 3\end{Bmatrix}\begin{Bmatrix}\frac{1\,5}{2} & 6 & \frac{3}{2}\\ K & \frac{3}{2} & 6\end{Bmatrix}\begin{Bmatrix}1 & 1 & \kappa\\ 5 & 5 & k\\ 6 & 6 & K\end{Bmatrix}\begin{Bmatrix}\frac{1}{2} & \frac{1}{2} & \kappa\\ 1 & 1 & k\\ \frac{3}{2} & \frac{3}{2} & K\end{Bmatrix},$$

$$D_{QQ'}(P,R)=(-1)^{Q+\frac{7}{2}}2^4\cdot3\cdot11\cdot13\{[Q][R][Q']\}^{\frac{1}{2}}\begin{Bmatrix}Q & R & Q'\\ \frac{5}{2} & 0 & -\frac{5}{2}\end{Bmatrix}\begin{Bmatrix}\frac{5}{2} & \frac{5}{2} & P\\ 1 & 1 & P\\ Q & Q' & R\end{Bmatrix}\Big/\begin{Bmatrix}\frac{5}{2} & P & \frac{5}{2}\\ \frac{5}{2} & 0 & -\frac{5}{2}\end{Bmatrix}\begin{Bmatrix}1 & P & 1\\ 1 & 0 & -1\end{Bmatrix}.$$

The matrix elements of the η's may then be written

$$\langle(\tfrac{5}{2}\,1)Q|\,\eta((\kappa k)K,(\kappa k)K,0)\,|(\tfrac{5}{2}\,1)Q'\rangle = C(\kappa,k,K)E_{QQ'}(P), \quad \text{for} \quad \kappa+k\neq 0, \quad (31)$$

where P is the quasispin label of η in the f and p shells.

For $\kappa+k$ even ($P=1$), the two-by-two E matrix is obtained from $[D(1,2)]$ and $[D(1,0)]$ by carrying out the addition (15):

$$[E(1)] = -(\tfrac{8}{3})^{\frac{1}{2}}[D(1,2)] - (\tfrac{1}{3})^{\frac{1}{2}}[D(1,0)].$$

For $\kappa+k$ odd ($P=0$),

$$[E(0)] = [D(0,0)].$$

Similarly, for the case $\kappa + k = 0$, the matrix elements of η can be obtained by using Eqs. (21), (24), and (18). The contribution of the single-particle operator

$$-([l_B]/[l_A])^{\frac{1}{2}}N_A - ([l_A]/[l_B])^{\frac{1}{2}}N_B + ([l_A][l_B])^{\frac{1}{2}}$$

may be easily evaluated using the transformation Eq. (30). We write

$$\langle(\tfrac{5}{2}\,1)Q|\,\eta^{(2)}((\kappa k)K,(\kappa k)K,0)\,|(\tfrac{5}{2}\,1)Q'\rangle = A_{QQ'}, \quad \text{for } \kappa = k = 0. \quad (32)$$

According to Eqs. (26), (31), and (32), the matrix elements of $\sum_{i>j}\mathbf{U}_i^{(02)}(f,p)\cdot\mathbf{U}_j^{(02)}(f,p)$ in the total quasispin scheme, which we shall denote $M_{QQ'}$, are given by the equation

$$\langle(\tfrac{5}{2}\,1)Q|\sum_{i>j}\mathbf{U}_i^{(02)}(f,p)\cdot\mathbf{U}_j^{(02)}(f,p)\,|(\tfrac{5}{2}\,1)Q'\rangle = \tfrac{5}{84}(21)^{\frac{1}{2}}A_{QQ'} + (\tfrac{5}{4})\sum_{\kappa kK}(-)^k([K])^{\frac{1}{2}} \times \begin{Bmatrix}3 & 3 & k\\ 1 & 1 & r\end{Bmatrix} C(\kappa,k,K)E_{QQ'}(P).$$

In this way $[M]$ is found to be

$$[M] = \begin{pmatrix} -(\frac{8\,4\,3}{5\,8\,8}) & -(\frac{1\,9}{2\,9\,4})(10)^{\frac{1}{2}} \\ -(\frac{1\,9}{2\,9\,4})(10)^{\frac{1}{2}} & -(\frac{1\,9}{4\,9}) \end{pmatrix}.$$

The matrix elements of $\sum_{i>j}\mathbf{U}_i^{(02)}(f,p)\cdot\mathbf{U}_j^{(02)}(f,p)$ between the states $|f^4p^{\frac{5}{2}3}H_6\times{}^1P_{\frac{3}{2}},\frac{1}{2}\rangle$ $(|1\rangle)$ and $|f^2p^{\frac{3}{2}\frac{5}{2}3}H_6\times{}^1P_{\frac{3}{2}},\frac{1}{2}\rangle$ $(|2\rangle)$ are obtained from $[M]$ by means of the unitary transformation

$$U^+[M]U = \begin{pmatrix} -\frac{9}{14} & (\frac{1\,3}{8\,4})(10)^{\frac{1}{2}} \\ (\frac{1\,3}{8\,4})(10)^{\frac{1}{2}} & -\frac{5\,5}{8\,4} \end{pmatrix}.$$

Each of these matrix elements was also calculated by the conventional method that seemed most convenient. The (1, 1) element was calculated by fractional parentage, the (2, 2) element was calculated by Slater determinants, and the off-diagonal elements were calculated by a method recently developed by Shore.[13] Complete agreement was obtained.

The calculation of these four matrix elements was easier and more straightforward by our method than by the conventional methods. This was true even though the states which we chose correspond to a "stretch" case for which the conventional methods could be simplified considerably.

[13] B. W. Shore, Phys. Rev. **139**, A1042 (1965).

Our method has the useful feature that it provides general equations which may be applied directly. The equations do not contain sums over dummy indices or sums over magnetic quantum numbers which require further simplification.

VI. CONCLUSION

We have found that by coupling together the quasispins of A and B particles, we may produce states having a total quasispin label which serve as a convenient basis for the calculation of matrix elements within $(l_A + l_B)^N$. Since these states have a total quasispin value, the Wigner–Eckart theorem may be used to factor the N dependence from the matrix elements, and the resulting reduced matrix elements, which are independent of N, can be factored into A and B parts. Thus the matrix elements of an operator acting within the $(l_A + l_B)$ shell can be calculated from the matrix elements of other operators acting solely within A or within B. This factorization of the matrix elements of $(l_A + l_B)$ should be very convenient for the study of complex configurations such as $(f + p)^N$. It is also hoped that it will provide some insight into these more complicated problems.

This method of evaluating matrix elements can obviously be extended to deal with matrix elements between configurations having several open shells. The basic operators and states could then be formed by coupling three or more quasispins.

Throughout the above analysis there has been an equivalence between quasispin labels and representation labels of the symplectic group. For instance, in the case $(f + p)^N$, the first term on the rhs of Eq. (15) transforms according to the $(11110\cdots0)$ representation of SP_{20} and the second term transforms according to the $(220\cdots0)$ representation of SP_{20}. Thus the quasispin labels 2 and 0 are equivalent to the symplectic labels $(11110\cdots0)$ and $(220\cdots0)$. The idea of coupling quasispins, however, is much more lucid than the idea of combining representations of $SP(4l_A + 2) \times SP(4l_B + 2)$ to form a representation of $SP(4l_A + 4l_B + 4)$.

ACKNOWLEDGMENTS

I am indebted to Dr. H. T. Wadzinski for suggesting that the transformation relating the two sets of states amounted to coupling the quasispins. I would like to thank Dr. B. R. Judd for his encouragement.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

On a Condition for Completeness*

JACK WONG
Lawrence Radiation Laboratory, University of California, Livermore, California

(Received 15 April 1967)

It is shown that a completeness relation for the eigensolutions of a non-Hermitian operator H can be derived even if the resolvent operator $R(H)$ of H is allowed to have poles of higher order than just simple poles, as required by Fonda, Ghirardi, Weber, and Rimini. A class of operators satisfying the requirements of this note is cited.

In describing a physical system by a Hamiltonian H, it is important to know whether or not the eigensolutions of the Hamiltonian H are complete in the carrier Hilbert space. Recently Fonda *et al.*[1] have shown that the eigensolutions of a non-Hermitian operator H are complete in a Hilbert space if the resolvent operator $R(z)$ $[\equiv (zI - H)^{-1}]$ of H has only simple poles at points of the point spectrum $\sigma_p(H)$ (among other assumptions). The purpose of this note is to point out that the above assumption can be relaxed as follows: The resolvent operator $R(z)$ of the non-Hermitian operator H has a Laurent expansion of *finite* order about any points $E_a, E_b, E_c, \cdots$ of the point spectrum $\sigma_p(H)$. The idea that such a relaxation is possible is principally motivated by the mathematical properties of compact (completely continuous) operators[2] and a recent study on the spectral properties of a class of non-Hermitian operators by the author.[3] Compact operators possess spectral properties that mimic those of matrices, and the class of non-Hermitian operators of Ref. 3 possess point spectra with properties like those of compact operators and a continuous spectrum that is real. These properties coincide almost completely with the assumptions of Fonda *et al.*[1] in their proof of the completeness of the eigensolutions of H.

Since our intention is to relax only one of the conditions of Ref. 1, we state a modified version of their assumptions as follows:

(1) The resolvent operator $R(z) = (zI - H)^{-1}$ is defined on an everywhere-dense set D in a Hilbert space, such that for all ϕ, ψ in D the function $(\phi, R(z)\psi)$ is analytic in z except for isolated poles of finite order and for cuts. The function $(\phi, R(z)\psi)$ has a *Laurent expansion* about any of its isolated poles:

$$(\phi, R(z)\psi) = \sum_{n=-\nu(b)}^{\infty} (z - E_b)^n (\phi, B_n\psi),$$
$$\nu(b) = (\text{order of the pole } E_b) < \infty, \qquad (1)$$
$$(\phi, B_n\psi) = \frac{1}{2\pi i}\int_{\Gamma_b} (z - E_b)^{-n-1}(\phi, R(z)\psi)\, dz,$$

where the contour Γ_b encloses only the singularity E_b. Since ϕ and ψ are arbitrary functions belonging to a dense set D, it is meaningful to speak of the B_n's as operators with domain D. We assume that the range of any of the operators B_n is in D, so that the operator product B_mB_n makes sense. The function $(\phi, R(z)\psi)$ has a branch cut lying along the real axis $E_0 \leq \mathrm{Re}\,(z) < \infty$. The "projection" operators p's are defined by the following formulas:

$$\frac{1}{2\pi i}\int_{\Gamma_b} (\phi, R(z)\psi)\, dz = (\phi, B_{-1}\psi) = (\phi, P_b\psi)$$

and

$$(-)\frac{1}{2\pi i}\{(\phi, R(E + i0)\psi) - (\phi, R(E - i0)\psi)\}$$
$$= (\phi, P(E)\psi), \quad E_0 \leq E < \infty. \qquad (2)$$

The operator $P(E)$ has the usual meaning of a projection operator associated with the continuous spectrum.[4]

Remarks: In case that (a) the operator H is closed, (b) $(zI - H)$ has a range dense in the Hilbert space, and (c) the inverse operator $R(z) = (zI - H)^{-1}$ is continuous, then $R(z)$ is guaranteed to be analytic in the resolvent set $\rho(H)$.[5]

(2) The function $(\phi, R(z)\psi)$ has only a finite number of poles and none of the poles fall on the branch cut $E_0 \leq \mathrm{Re}\,(z) < \infty$.

* Work performed under the auspices of the U.S. Atomic Energy Commission.

[1] L. Fonda, G. C. Ghirardi, T. Weber, and A. Rimini, J. Math. Phys. 7, 1643 (1966).

[2] A. C. Zaamen, *Linear Analysis* (Interscience Publishers, Inc., New York, 1953).

[3] J. Wong, J. Math. Phys. 8, 2039 (1967).

[4] See, e.g., N. I. Akhiezer and I. M. Glazman, *Theory of Linear Operators in Hilbert Space* (Frederick Ungar Publ. Co., New York, 1963), Vol. II.

[5] See, e.g., A. E. Taylor, *Introduction to Functional Analysis* (John Wiley & Sons, Inc., New York, 1964), Chap. 5.

(3) Asymptotically, the function $(\phi, R(z)\psi) \to z^{-1}(\phi, I\psi)$ as $|z| \to \infty$ for all z differing from the cut $E_0 \leq \mathrm{Re}\,(z) < \infty$.

The modification occurs in assumption (1) in allowing a Laurent expansion of finite order about any pole E_b, rather than just a simple pole.

The product B_mB_n of two operator coefficients B_m and B_n can be computed with the aid of the operator identity

$$(z' - z)R(z)R(z') = R(z) - R(z'), \quad z \text{ and } z' \in \rho(H), \tag{3}$$

$$\begin{aligned} B_mB_n &= \frac{1}{(2\pi i)^2}\int_{\Gamma_b} dz \int_{\Gamma_{b'}} dz'(z - E_b)^{-m-1} \\ &\quad \times (z' - E_b)^{-n-1}R(z)R(z') \\ &= \frac{1}{(2\pi i)^2}\int_{\Gamma_b} dz \int_{\Gamma_{b'}} dz'(z - E_b)^{-m-1} \\ &\quad \times (z' - E_b)^{-n-1}(z' - z)^{-1}[R(z) - R(z')] \\ &= [1 - \theta(m) - \theta(n)]B_{m+n+1}, \quad m, n \geq \nu(b), \end{aligned}$$

where

$$\begin{aligned} \theta(m) &= 1, \quad \text{if} \quad m \geq 0, \\ &= 0, \quad \text{if} \quad m < 0, \end{aligned} \tag{4}$$

and for ease of computation, one contour (say Γ_b') should be placed completely inside the other contour (Γ_b). By letting m and n assume various admissible integral values, relations between various B_m can be deduced:

$$\begin{aligned} &B_{-1} = (B_{-1})^2, \quad \text{or } P_b = P_b^2, && \text{a projection operator,} \\ &B_{-m} = (B_{-2})^{m-1} \equiv N_b^{m-1}, && m = 2, 3, \cdots, \nu(b), \\ &(N_b)^{\nu(b)} = 0, && \text{a nilpotent operator,} \\ &B_n = (-)^n B_0^{n+1}, && n = 0, 1, \cdots, \infty, \end{aligned} \tag{5}$$

and

$$\begin{aligned} B_{-2}B_{-1} = B_{-1}B_{-2} &= P_bN_b = N_b, \\ B_0B_{-1} = B_{-1}B_0 &= P_bB_0 = 0, \\ B_0B_{-2} = B_{-2}B_0 &= N_bB_0 = 0. \end{aligned} \tag{6}$$

By a method of computation similar to Eq. (4), the product of two operator coefficients A_m and B_n associated with two distinct isolated poles E_a and E_b vanishes for m and n less than zero:

$$\begin{aligned} A_mB_n = B_nA_m &= \frac{1}{(2\pi i)^2}\int_{\Gamma_a} dz \int_{\Gamma_{b'}} dz'(z - E_a)^{-m-1} \\ &\quad \times (z' - E_b)^{-n-1}R(z)R(z') \\ &= \frac{1}{(2\pi i)^2}\int_{\Gamma_a} dz \int_{\Gamma_{b'}} dz'(z - E_a)^{-m-1} \\ &\quad \times (z' - E_b)^{-n-1}(z' - z)^{-1}[R(z) - R(z')] \\ &= 0, \quad m, n < 0. \end{aligned}$$

In particular, we have the following important identities:

$$\begin{aligned} A_{-1}B_{-1} &= P_aP_b = P_bP_a = 0, \\ A_{-1}B_{-2} &= P_aN_b = N_bP_a = 0, \\ A_{-2}B_{-1} &= N_aP_b = P_bN_a = 0. \end{aligned} \tag{7}$$

It should be noticed that the properties of Eq. (7) are those of projection operators. With the definition of Eq. (2) and with the operator identity Eq. (3), the operator product $P_bP(E)$ and $N_bP(E)$ can be shown to satisfy the following projection operator identities:

$$\begin{aligned} P_bP(E) &= P(E)P_b = 0, \\ N_bP(E) &= P(E)N_b = 0, \quad E_0 \leq E < \infty. \end{aligned} \tag{8}$$

The completeness relation of the projection operators $P_a, P_b, P_c, \cdots$, and $P(E)$ can be derived with the assumptions (1), (2), and (3) by equating two analytically identical expressions (as was done in Ref. 1):

$$\begin{aligned} Q &= \frac{1}{2\pi i}\int_{\Gamma_1} dz(\phi, R(z)\psi) = \sum_b (\phi, P_b\psi), \\ Q &= \frac{1}{2\pi i}\int_{\Gamma_2} dz(\phi, R(z)\psi) = (\phi, \psi) - \int_{E_0}^{\infty} dE(\phi, P(E)\psi), \end{aligned} \tag{9}$$

where Γ_1 is a counterclockwise-oriented contour encircling only the poles of $(\phi, R(z)\psi)$ and Γ_2 is the analytic deformation of Γ_1 = (1) a counterclockwise-oriented circle at $|z| \simeq \infty$ and (2) a contour wrapping around the cut $E_0 \leq \mathrm{Re}\,(z) < \infty$. Equating the two expressions of Eq. (9) and dropping the arbitrary functions of ϕ and ψ of an *everywhere-dense* set D, we arrive at the completeness relations

$$\sum_b P_b + \int_{E_0}^{\infty} dEP(E) = I. \tag{10}$$

The remainder of the discussion on the completeness relation should be the same as that of Ref. 1 except for their Eq. (13) because of our Eq. (6). We have instead

$$H = \sum_b (E_bP_b + N_b) + \int_{E_0}^{\infty} EP(E)\,dE. \tag{11}$$

Hence,

$$P_bH = HP_b = E_bP_b + N_b. \tag{12}$$

But this should present no difficulty as far as the completeness relation is concerned.

Now we cite a class of physically reasonable operators that are special examples for this note. The class of operators discussed in Ref. 3 possess compact operator properties for their point spectrum and selfadjoint operator properties for their continuous spectrum. They can be easily shown to satisfy all the assumptions (1), (2), and (3). A completeness

relation for them can be written down readily with the biorthonormal sets of eigenfunctions $\{\phi_b\}$ and $\{\psi_b\}$, where

$$H\psi_b = E_b\psi,$$
$$H^\dagger\phi_b = E_b^*\phi_b, \quad (\phi_a, \psi_b) = \delta_{ab}. \tag{13}$$

Furthermore, the eigenmultiplicities $\nu(b)$ of the eigenvalues E_b and E_b^* must be the same because, in the proof of Theorem A of Ref. 3, the eigenspace of E_b is the null space of $T(z) - I$, where $T(z)$ is a *compact* operator. With the help of the outgoing-wave boundary condition for the unrenormalizable eigenvectors ϕ^+, ψ^+ (see Ref. 1), we write a completeness relation

$$\sum_{\substack{b \\ m(b)=1,2,\cdots,\nu(b)}} |\psi_{b,m(b)}\rangle\langle\phi_{b,m(b)}| + \int_{E_0}^{\infty} dE\, |\psi_E^+\rangle\langle\phi_E^+| = I, \tag{14}$$

where $\{\psi_{b,m(b)}\}$ and $\{\phi_{b',m'(b)}\}$ span the associated eigenspaces of E_b and E_b^*, respectively, and can be chosen such that

$$(\phi_{b',m'(b)}, \psi_{b,m(b)}) = \delta_{b'b}\delta_{m'(b)m(b)}.$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Partial Sums of Vacuum Polarization Graphs*

JAMES P. FINK
Department of Mathematics and Department of Physics
University of Pittsburgh, Pittsburgh, Pennsylvania

(Received 17 January 1969)

This paper contains results on the logarithmic asymptotic behavior of certain partial sums of photon self-energy graphs.

I. INTRODUCTION

The question of obtaining logarithmic asymptotic estimates for renormalized Feynman integrals, in particular, those arising from the vacuum polarization graphs of quantum electrodynamics, has been considered in Refs. 1 and 2. In Ref. 1, techniques were developed enabling one to obtain rigorous logarithmic estimates for a single Feynman graph. The problem of summing different graphs was not considered there. In Ref. 2, renormalization-group arguments were used to obtain logarithmic estimates on the sum of all the photon self-energy graphs of a given order.

In this paper, we use the result of Jost and Luttinger[3,4] and the techniques of Ref. 1 to obtain logarithmic estimates on certain partial sums of photon self-energy graphs. All of the notation and terminology used in the following is described in detail in Ref. 1.

II. SUMMING GRAPHS

The analysis of sums of graphs using the results of Ref. 1 seems, in general, to be a difficult problem.

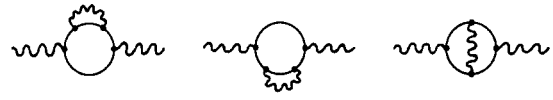

FIG. 1. Fourth-order photon self-energy graphs.

For a special collection of graphs, however, we can determine upper bounds on the logarithmic asymptotic coefficients of their sums.

We begin by considering the three fourth-order photon self-energy graphs shown in Fig. 1. Each graph has logarithmic asymptotic coefficient 2,[1] but in summing these graphs the $\log^2(q^2/\lambda^2)$ terms cancel so that the sum has logarithmic asymptotic coefficient 1.[3,4]

Consider now any three nth-order photon self-energy graphs which can be obtained from the fourth-order graphs in Fig. 1 by inserting the same m irreducible insertions into each one. For example, the eighth-order graphs in Fig. 2 are obtained from those in Fig. 1 by inserting first the irreducible insertion shown in Fig. 3(a) followed by the irreducible insertion shown in Fig. 3(b).

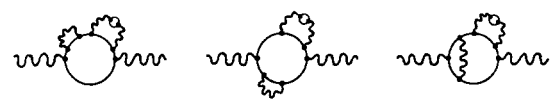

FIG. 2. Eighth-order photon self-energy graphs.

* Supported in part by the U.S. Atomic Energy Commission under Contract No. AT(30-1)-3829.

[1] J. Fink, J. Math. Phys. **9**, 1389 (1968).
[2] J. Fink, Phys. Rev. **170**, 1580 (1968).
[3] R. Jost and J. M. Luttinger, Helv. Phys. Acta **23**, 201 (1950).
[4] J. Bjorken and S. Drell, *Relativistic Quantum Fields* (McGraw-Hill Book Company, New York, 1965).

relation for them can be written down readily with the biorthonormal sets of eigenfunctions $\{\phi_b\}$ and $\{\psi_b\}$, where

$$H\psi_b = E_b\psi,$$
$$H^\dagger\phi_b = E_b^*\phi_b, \quad (\phi_a, \psi_b) = \delta_{ab}. \tag{13}$$

Furthermore, the eigenmultiplicities $\nu(b)$ of the eigenvalues E_b and E_b^* must be the same because, in the proof of Theorem A of Ref. 3, the eigenspace of E_b is the null space of $T(z) - I$, where $T(z)$ is a *compact* operator. With the help of the outgoing-wave boundary condition for the unrenormalizable eigenvectors ϕ^+, ψ^+ (see Ref. 1), we write a completeness relation

$$\sum_{\substack{b \\ m(b)=1,2,\cdots,\nu(b)}} |\psi_{b,m(b)}\rangle\langle\phi_{b,m(b)}| + \int_{E_0}^{\infty} dE\, |\psi_E^+\rangle\langle\phi_E^+| = I, \tag{14}$$

where $\{\psi_{b,m(b)}\}$ and $\{\phi_{b',m'(b)}\}$ span the associated eigenspaces of E_b and E_b^*, respectively, and can be chosen such that

$$(\phi_{b',m'(b)}, \psi_{b,m(b)}) = \delta_{b'b}\delta_{m'(b)m(b)}.$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Partial Sums of Vacuum Polarization Graphs*

JAMES P. FINK
Department of Mathematics and Department of Physics
University of Pittsburgh, Pittsburgh, Pennsylvania

(Received 17 January 1969)

This paper contains results on the logarithmic asymptotic behavior of certain partial sums of photon self-energy graphs.

I. INTRODUCTION

The question of obtaining logarithmic asymptotic estimates for renormalized Feynman integrals, in particular, those arising from the vacuum polarization graphs of quantum electrodynamics, has been considered in Refs. 1 and 2. In Ref. 1, techniques were developed enabling one to obtain rigorous logarithmic estimates for a single Feynman graph. The problem of summing different graphs was not considered there. In Ref. 2, renormalization-group arguments were used to obtain logarithmic estimates on the sum of all the photon self-energy graphs of a given order.

In this paper, we use the result of Jost and Luttinger[3,4] and the techniques of Ref. 1 to obtain logarithmic estimates on certain partial sums of photon self-energy graphs. All of the notation and terminology used in the following is described in detail in Ref. 1.

II. SUMMING GRAPHS

The analysis of sums of graphs using the results of Ref. 1 seems, in general, to be a difficult problem.

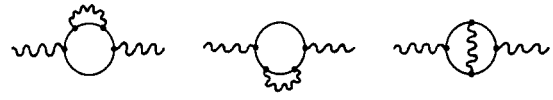

FIG. 1. Fourth-order photon self-energy graphs.

For a special collection of graphs, however, we can determine upper bounds on the logarithmic asymptotic coefficients of their sums.

We begin by considering the three fourth-order photon self-energy graphs shown in Fig. 1. Each graph has logarithmic asymptotic coefficient 2,[1] but in summing these graphs the $\log^2(q^2/\lambda^2)$ terms cancel so that the sum has logarithmic asymptotic coefficient 1.[3,4]

Consider now any three nth-order photon self-energy graphs which can be obtained from the fourth-order graphs in Fig. 1 by inserting the same m irreducible insertions into each one. For example, the eighth-order graphs in Fig. 2 are obtained from those in Fig. 1 by inserting first the irreducible insertion shown in Fig. 3(a) followed by the irreducible insertion shown in Fig. 3(b).

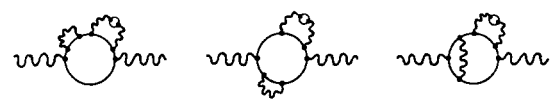

FIG. 2. Eighth-order photon self-energy graphs.

* Supported in part by the U.S. Atomic Energy Commission under Contract No. AT(30-1)-3829.

[1] J. Fink, J. Math. Phys. **9**, 1389 (1968).
[2] J. Fink, Phys. Rev. **170**, 1580 (1968).
[3] R. Jost and J. M. Luttinger, Helv. Phys. Acta **23**, 201 (1950).
[4] J. Bjorken and S. Drell, *Relativistic Quantum Fields* (McGraw-Hill Book Company, New York, 1965).

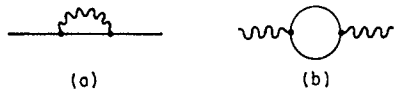

FIG. 3. Irreducible insertions added to the graphs of Fig. 1 to obtain the graphs of Fig 2.

Each of the three nth-order graphs so obtained has logarithmic asymptotic coefficient $m + 2$.[1] Hence, the logarithmic asymptotic coefficient of their sum is less than or equal to $m + 2$. Using the known cancellation in the sum of the fourth-order graphs in Fig. 1, we can improve this upper bound.

Let $\beta^{(n)}$, $n \geq 4$, be the logarithmic asymptotic coefficient of the sum of three nth-order photon self-energy graphs obtained as described above. We know that

$$\beta^{(4)} = 1,$$
$$\beta^{(n)} \leq m + 2.$$

Furthermore, since the dimension of the Feynman integral corresponding to an nth-order photon self-energy graph is $2n$, we obtain[1] from that

$$\beta^{(n)} = \sum_{i=1}^{2n} p_i - 2n,$$

where the dimension numbers p_i are defined explicitly in Ref. 1. Thus,

$$\begin{aligned}\beta^{(n)} &= \sum_{i=1}^{2n} p_i - 2n \\ &= \sum_{i=1}^{2\cdot 4} p_i - 2\cdot 4 + \sum_{i=2\cdot 4+1}^{2n} p_i - (2n - 2\cdot 4) \\ &= \beta^{(4)} + \sum_{i=2\cdot 4+1}^{2n} p_i - 2(n-4) \\ &\leq 1 + m.\end{aligned}$$

Here we have used the fact that $\beta^{(4)} = 1$. The estimate on the remaining terms is derived in Ref. 1.

We have, therefore, proved the following theorem:

Theorem 1: Let G be the sum of any three nth-order ($n \geq 4$) photon self-energy graphs obtainable from the three fourth-order photon self-energy graphs by inserting the same m irreducible insertions into each. Then G has logarithmic asymptotic coefficient $\beta(G) \leq m + 1$.

Two corollaries follow immediately.

Corollary 1: Let G be a sum of nth-order ($n \geq 4$) photon self-energy graphs consisting of sums G_i, $i = 1, \cdots, k$, of the type described in Theorem 1. Let m_i, $i = 1, \cdots, k$, be the number of irreducible insertions added to the fourth-order graphs in order to obtain G_i. Then G has logarithmic asymptotic coefficient

$$\beta(G) \leq \max_{i=1,\cdots,k} \{m_i\} + 1.$$

The second corollary follows by observing that the maximum number of irreducible insertions that can be added to a fourth-order photon self-energy graph in order to obtain an nth-order graph is $(n-4)/2$.

Corollary 2: Let G be the sum of three nth-order ($n \geq 4$) photon self-energy graphs as described in Theorem 1. Then G always has logarithmic asymptotic coefficient $\beta(G) \leq \frac{1}{2}n - 1$.

III. CONCLUDING REMARKS

Corollary 2 gives one hope that perhaps these techniques can be used to show that the highest power of the logarithm cancels when all the graphs of a given order are summed together. This result was derived in Ref. 2 using renormalization-group arguments, but it does not follow from the above results as they stand.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

On Fluids of Particles with Short-Range Repulsion and Weak Long-Range Attractive Interaction*

JOHN B. JALICKEE AND ARNOLD J. F. SIEGERT
Physics Department, Northwestern University, Evanston, Illinois
AND
DAVID J. VEZZETTI
Physics Department, University of Illinois at Chicago Circle, Chicago, Illinois

(Received 2 August 1968)

We have obtained an appreciably simplified derivation of the expansion in inverse range of the thermodynamic functions of a system of particles with short-range repulsion and weak long-range attractive interaction. We assume all properties of the system with only the short-range repulsive interaction to be known and consider this system as a reference system for the purpose of a perturbation expansion. We derive corrections to the van der Waals equation and to the Maxwell construction. The terms in our expansion are integrals over products of the modified Ursell functions of the reference system only and factors which are the transformed long-range potential. A simple rule is given to determine the order, in the inverse range of the attractive potential, of each term in the expansion. The simplification of the derivation is achieved by avoiding the expansion and subsequent resummation of any functions pertaining to the reference system.

SECTION 1

Because of the tremendous mathematical difficulties of calculating the thermodynamic functions of systems of interacting particles, even in the classical limit and with pair interactions only, it is natural to look for perturbation methods. In such a method, one has to start from a reference system whose properties are assumed to be known. Expansions in number density or fugacity use as reference system the ideal gas, and their limitations are well known.

For the case of the Ising model of the ferromagnet, it was suggested first by Brout[1] that one should consider as perturbation a long-range, weak interaction with the reciprocal effective number of neighbors as expansion parameter, because it was surmised that the Weiss theory of ferromagnetism would become exact in the limit of infinitely weak, infinitely long-range integrable interaction potential.

For the fluid, the corresponding procedure of using as reference system a system of particles with strong short-range repulsive interaction and as a perturbation a weak long-range attractive interaction has been used by many authors.[2] The attractive part of the potential energy of interaction is usually taken to be of the form $-\gamma^{\nu}\Phi(\gamma|\mathbf{r}_1 - \mathbf{r}_2|)$, where ν is the dimensionality of the space and the parameter γ^{-1} is a measure of the range of the attractive force. It is now established that for a wide class of repulsive potentials and for a wide class of functions Φ, in the limit $\gamma \to 0$, the equation of state is determined by applying the Maxwell construction to the van der Waals type equation

$$p(\rho) = p_h(\rho) - \tfrac{1}{2}v_0\rho^2,$$

where ρ is the number density, $p(\rho)$ and $p_h(\rho)$ the pressures of the system and the reference system, respectively, and $v_0 = \int \Phi(\mathbf{x})\, d^{\nu}x$.[3] Corrections to the van der Waals equation, in the form of expansions in powers of γ, explicit to order $\gamma^{2\nu}$, have been obtained by several authors: γ^{ν} by Hemmer,[4] $\gamma^{2\nu-1}$ by Zittartz,[5] and $\gamma^{2\nu}$ by Lebowitz, Stell, and Baer.[2] The last authors also develop a method of obtaining the expansion to all orders of γ.

We are aware of the objections to the expansion in powers of γ. It fails in an obvious manner at the critical point of the van der Waals theory, and it predicts a phase transition of the one-dimensional model for $\gamma \neq 0$ which, from the exact calculation of Kac, Uhlenbeck, and Hemmer,[3] is known not to occur. It is not known whether partial resummation of the expansion can remove these shortcomings. The expansion may, however, be valid, at least as an asymptotic series, for the two- and three-dimensional fluid in single-phase regions including the first-order phase transition at any fixed temperature above or below the critical temperature for sufficiently small γ. For this reason, and because of the desirability of

* Program at Northwestern University supported by Office of Naval Research and National Science Foundation.

[1] R. Brout, Phys. Rev. **118**, 1009 (1960).

[2] See J. L. Lebowitz, G. Stell, and S. Baer, J. Math. Phys. **6**, 1282 (1965), and references given there.

[3] For the one-dimensional system of particles with hard-rod interaction and $\Phi(x) = e^{-|x|}$, this was established by M. Kac, G. E. Uhlenbeck, and P. C. Hemmer, J. Math. Phys. **4**, 216 (1963). The general result was obtained heuristically by N. G. van Kampen, Phys. Rev. **135**, 362 (1964), and made rigorous by J. L. Lebowitz and O. Penrose, J. Math. Phys. **7**, 98 (1966).

[4] P. C. Hemmer, J. Math. Phys. **5**, 75 (1964).

[5] J. Zittartz, Z. Physik **180**, 219 (1964).

having the expansion in a transparent form, facilitating attempts at resummation, we have reconsidered the expansion and a possible modification.

In the present paper, we derive corrections to the van der Waals equation and to the Maxwell construction. Our derivation is made comparatively simple and straightforward by avoiding any expansion and subsequent resummation of functions pertaining to the reference system. The terms in our expansion are given as integrals over products of the modified Ursell functions[6] of the reference system only and factors which are the transformed potential of the Debye type, but of long range. This potential is obtainable by Fourier inversion from the Fourier transforms of the two-particle correlation function of the reference system and the original attractive potential. We give a simple rule for determining the order in γ of any term in the expansion.

We use the method of random functions, originated by Kac.[7] In Sec. 2, we introduce our notation and state restrictions on the potential in order to use the method of random functions in its most convenient form. In Sec. 3, we write the grand partition function of the system as an average over random functions (function-space integral) and motivate the expansion of the integrand. In Sec. 4, we carry out the expansion and the average over the random functions. The terms of this expansion are conveniently written as diagrams. While our diagrams consist of circles (hypervertices) and lines, like those of Lebowitz, Stell, and Baer,[2] the interpretation of these diagrams is different. Our circles represent the modified Ursell functions of the reference system and are independent of the long-range part of the potential. In Sec. 5, we show how to determine the order in γ of any term in the expansion. In Sec. 6, we obtain the correction terms to pressure and number density and to the Maxwell construction in the grand canonical form and give the formal perturbation expansion of the Helmholtz free energy as function of the density. Section 7 contains a summary and discussion of our results.

SECTION 2

We consider a system of particles in a ν-dimensional cube of volume V. The interaction potential is taken to consist of a hard-core[8] interaction and an attractive pair potential $-v(\mathbf{r}) \leq 0$. We restrict our considerations to functions $v(\mathbf{r})$ which can be represented as the covariance of real-valued Gaussian random functions $\phi(r)$ with

$$\langle \phi(\mathbf{r}) \rangle_{\mathrm{av}\ \phi} = 0 \tag{2.1}$$

and

$$\langle \phi(\mathbf{r})\phi(\mathbf{r}') \rangle_{\mathrm{av}\ \phi} = \beta v(\mathbf{r} - \mathbf{r}'), \tag{2.2}$$

where $\beta = (kT)^{-1}$ with k Boltzmann's constant and T the absolute temperature. The restriction to real-valued random functions requires that the eigenvalues v_σ of $v(\mathbf{r} - \mathbf{r}')$, considered as a kernel, are positive and that $v(0)$ is finite.[9] An explicit representation of the functions $\phi(\mathbf{r})$ can be obtained by expanding them in the eigenfunctions of $v(\mathbf{r} - \mathbf{r}')$ in the volume V, i.e.,

$$\phi(\mathbf{r}) = \sum_\sigma c_\sigma (\beta v_\sigma)^{\frac{1}{2}} u_\sigma(\mathbf{r}), \tag{2.3}$$

where

$$\int_V d^\nu r' v(\mathbf{r} - \mathbf{r}') u_\sigma(\mathbf{r}') = v_\sigma u_\sigma(\mathbf{r}). \tag{2.4}$$

In (2.3) the random variables c_σ are defined by the probability density

$$\prod_\sigma \left(\frac{dc_\sigma}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}c_\sigma^2} \right).$$

For convenience, we make the interaction potentials (both hard-core and attractive parts) periodic in space with period $V^{1/\nu}$, and impose Born–von Kármán conditions. In this case, the eigenfunctions $u_\sigma(\mathbf{r})$ are simply the trigonometric functions and we have, for the largest eigenvalue v_0 and corresponding eigenfunction $u_0(\mathbf{r})$,

$$v_0 = \int_V v(\mathbf{r})\, d^\nu r, \tag{2.5}$$

$$u_0(\mathbf{r}) = V^{-\frac{1}{2}}. \tag{2.6}$$

While the assumption of periodicity makes $v(\mathbf{r})$ volume dependent, for finite volumes, it is generally accepted that the correct thermodynamic functions are obtained in this way in the thermodynamic limit.

We are interested in the case where the attractive potential is weak and long range, and we introduce a reciprocal range γ by assuming that for the infinite system the potential is of the form $-\gamma^\nu \Phi(\gamma\, |\mathbf{r}_1 - \mathbf{r}_2|)$. The potential $v(\mathbf{r})$ is then defined by making this potential periodic.

We take $\Phi(\mathbf{r})$ to satisfy the conditions of Lebowitz and Penrose.[3] These guarantee that the equation of state of our system, in the limit $\gamma \to 0$, approaches the van der Waals type equation[10] with Maxwell construction.

[6] J. L. Lebowitz and J. K. Percus, J. Math. Phys. **4**, 1495 (1963).
[7] M. Kac, Phys. Fluids **2**, 8 (1959).
[8] By "hard core" we mean any repulsive potential which satisfies the conditions of Lebowitz and Penrose, Ref. 3.
[9] The restriction to real-valued random functions could easily be avoided, but we need the positivity of the v_σ and the finiteness of $v(0)$ for carrying out the estimates of Appendix B.
[10] By "van der Waals type" we mean the van der Waals equation with the van der Waals approximation to the hard-core pressure replaced by the exact hard-core pressure.

SECTION 3

We denote by $S_N(\mathbf{r}_1, \cdots, \mathbf{r}_N)$ the Boltzmann factor for the hard-core[8] potential alone. The Gibbs integral for our system is

$$Z_N(\beta, V) = \int_V d^{\nu N} r S_N \exp\left[\tfrac{1}{2}\beta \sum_{i \neq j} v(\mathbf{r}_i - \mathbf{r}_j)\right], \quad (3.1)$$

and introducing the random functions defined in the previous section, we obtain the well-known representation

$$Z_N(\beta, V) = \exp\left[-\tfrac{1}{2}N\beta v(0)\right] \times \left\langle \int_V d^{\nu N} r S_N \exp\left[\sum_{i=1}^{N} \phi(\mathbf{r}_i)\right]\right\rangle_{\text{av } \phi}, \quad (3.2)$$

which is easily verified by using the representation (2.3) and carrying out the integral over the variables c_σ.

The grand partition function at fugacity z, defined by

$$Q(z) \equiv \sum_{N=0}^{\infty} \frac{z^N}{N!} Z_N, \quad (3.3)$$

becomes

$$Q(z) = \left\langle \sum_{N=0}^{\infty} \frac{(ze^{-\frac{1}{2}\beta v(0)})^N}{N!} \times \int_V d^{\nu N} r S_N \exp\left[\sum_{i=1}^{N} \phi(\mathbf{r}_i)\right]\right\rangle_{\text{av } \phi}$$
$$= \int_{-\infty}^{\infty} \cdots \int \prod_\sigma \left(\frac{dc_\sigma}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}c_\sigma^2}\right) \sum_{N=0}^{\infty} \frac{(ze^{-\frac{1}{2}\beta v(0)})^N}{N!} \times \int_V d^{\nu N} r S_N \exp\left[\sum_{\sigma,i} c_\sigma (\beta v_\sigma)^{\frac{1}{2}} u_\sigma(\mathbf{r}_i)\right]. \quad (3.4)$$

For the extrema of the integrand, one has the set of equations

$$c_\tau = \frac{\partial}{\partial c_\tau} \ln \sum_{N=0}^{\infty} \frac{(ze^{-\frac{1}{2}\beta v(0)})^N}{N!} \times \int_V d^{\nu N} r S_N \exp\left[\sum_{\sigma,i} c_\sigma (\beta v_\sigma)^{\frac{1}{2}} u_\sigma(\mathbf{r}_i)\right]. \quad (3.5)$$

Carrying out the differentiation explicitly, one finds that these equations have some simple solutions, namely,

$$c_\tau^{(0)} = 0, \quad \text{for} \quad \tau \neq 0, \quad (3.6)$$

and $c_0^{(0)}$ given by the roots of the equation

$$c_0/(\beta v_0 V)^{\frac{1}{2}} = \rho_h\{z \exp\left[-\tfrac{1}{2}\beta v(0) + c_0(\beta v_0/V)^{\frac{1}{2}}\right]\}, \quad (3.7)$$

where $\rho_h(y)$ is the density of the hard-core system at fugacity y. This equation becomes more transparent

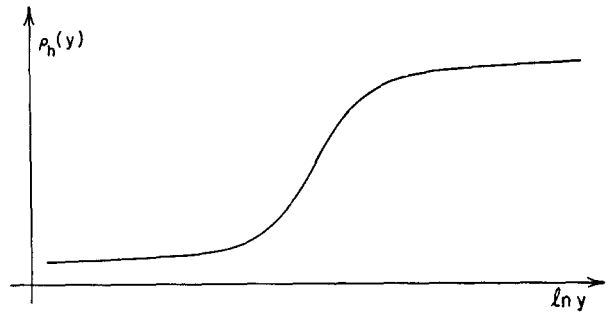

FIG. 1. Hard-core density vs ln (fugacity), schematic.

with the notation

$$\ln y = \ln z - \tfrac{1}{2}\beta v(0) + c_0(\beta v_0/V)^{\frac{1}{2}} \equiv \mu, \quad (3.8)$$

$$\rho_h(y) = \tilde{\rho}_h(\mu), \quad (3.9)$$

$$\ln z - \tfrac{1}{2}\beta v(0) \equiv \xi, \quad (3.10)$$

so that

$$c_0(\beta v_0/V)^{\frac{1}{2}} = \mu - \xi. \quad (3.11)$$

We then have

$$(\mu - \xi)/\beta v_0 = \tilde{\rho}_h(\mu). \quad (3.12)$$

$\tilde{\rho}_h(\mu)$ is a nondecreasing nonnegative bounded function and, from any of the known approximations, one expects the shape indicated schematically in Fig. 1. The possible values of μ and, therefore, of $c_0^{(0)}$ are thus obtained by the construction familiar from the Weiss theory of magnetism. Choosing that root of Eq. (3.12) which maximizes the integrand in (3.4), this construction is the grand canonical form of the Maxwell construction applied to the van der Waals type equation. The construction is discussed in Appendix A.

With this as motivation, we separate ϕ into its spatial average and its spatially inhomogeneous part $\psi(\mathbf{r})$, defined by

$$\psi(\mathbf{r}) \equiv \phi(\mathbf{r}) - c_0(\beta v_0/V)^{\frac{1}{2}}. \quad (3.13)$$

Equation (3.4) then becomes

$$Q(z) = \left\langle \left\langle \sum_{N=0}^{\infty} \frac{y^N}{N!} \int_V d^{\nu N} r S_N \times \exp\left[\sum_{i=1}^{N} \psi(\mathbf{r}_i)\right]\right\rangle_{\text{av } \psi} \right\rangle_{\text{av } c_0}, \quad (3.14)$$

where y is given by (3.8). For $\psi(\mathbf{r}) = 0$, we obtain

$$Q(z) \cong Q_0(z) \equiv \langle \exp\left[\beta V P_h(y)\right]\rangle_{\text{av } c_0}, \quad (3.15)$$

where $P_h(y)$ is the pressure of the hard-core system at fugacity y. According to the preceding arguments, this gives the van der Waals equation with Maxwell construction.

Quite generally, the representation (3.14) can be written in the form

$$Q(z) = \langle \exp [\beta V P_h(y) + V h(y)] \rangle_{\text{av } c_0}, \quad (3.16)$$

where $h(y)$ is defined by

$$h(y) \equiv V^{-1} \ln \left\langle \sum_{N=0}^{\infty} \frac{y^N}{N!} \int_V d^{\nu N} r S_N \times \exp \left[\sum_{i=1}^{N} \psi(\mathbf{r}_i) \right] \right\rangle_{\text{av } \psi} - \beta P_h(y). \quad (3.17)$$

It is now convenient to introduce a functional $A\{\psi\}$ defined by

$$A\{\psi\} \equiv \ln \sum_{N=0}^{\infty} \frac{y^N}{N!} \int_V d^{\nu N} r S_N \exp \left[\sum_{i=1}^{N} \psi(\mathbf{r}_i) \right] - \ln \sum_{N=0}^{\infty} \frac{y^N}{N!} \int_V d^{\nu N} r S_N, \quad (3.18)$$

so that (3.17) can be written in the form

$$h(y) = V^{-1} \ln \langle \exp A\{\psi\} \rangle_{\text{av } \psi}. \quad (3.19)$$

We have seen that setting $\psi(\mathbf{r}) = 0$ gives the van der Waals type equation of state. Thus, it is natural to expand the functional $A\{\psi\}$ in a functional Taylor series about $\psi(\mathbf{r}) = 0$. This expansion is well known to be

$$A\{\psi\} = \sum_{n=1}^{\infty} \frac{1}{n!} \int_V d^{\nu} r_1 \cdots \int_V d^{\nu} r_n \mu_n(\mathbf{r}_1, \cdots, \mathbf{r}_n \mid y) \psi(\mathbf{r}_1) \cdots \psi(\mathbf{r}_n), \quad (3.20)$$

where the functions $\mu_n(\mathbf{r}_1, \cdots, \mathbf{r}_n \mid y)$ are the modified Ursell functions for a hard-core system at fugacity y. They are defined in Ref. 6 and denoted there by $\hat{\mathcal{F}}_n$. In terms of the Ursell (or correlation) functions for a hard-core system $\mathcal{F}_n(\mathbf{r}_1, \cdots, \mathbf{r}_n \mid y)$, the first few are given by

$$\begin{aligned}
\mu_1(\mathbf{r}_1 \mid y) &= \mathcal{F}_1(\mathbf{r}_1 \mid y), \\
\mu_2(\mathbf{r}_1, \mathbf{r}_2 \mid y) &= \mathcal{F}_2(\mathbf{r}_1, \mathbf{r}_2 \mid y) + \mathcal{F}_1(\mathbf{r}_1 \mid y)\delta(\mathbf{r}_1 - \mathbf{r}_2), \\
\mu_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3 \mid y) &= \mathcal{F}_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3 \mid y) \\
&\quad + \mathcal{F}_2(\mathbf{r}_1, \mathbf{r}_2 \mid y)\delta(\mathbf{r}_2 - \mathbf{r}_3) + \mathcal{F}_2(\mathbf{r}_1, \mathbf{r}_3 \mid y)\delta(\mathbf{r}_1 - \mathbf{r}_2) \\
&\quad + \mathcal{F}_2(\mathbf{r}_2, \mathbf{r}_3 \mid y)\delta(\mathbf{r}_1 - \mathbf{r}_3) \\
&\quad + \mathcal{F}_1(\mathbf{r}_1 \mid y)\delta(\mathbf{r}_1 - \mathbf{r}_2)\delta(\mathbf{r}_2 - \mathbf{r}_3). \quad (3.21)
\end{aligned}$$

Because of the Born–von Kármán conditions, all the functions μ_n are translationally invariant and, in particular, $\mu_1(\mathbf{r}_1 \mid y)$ is independent of $\mathbf{r}_1$. In the fluid, the range of μ_n will be of order n times the range of the repulsive potential, which will be used later. Since $\int_V \psi(\mathbf{r})\, d^{\nu} r = 0$, by definition, the term linear in ψ in (3.20) vanishes. Keeping only the term quadratic in ψ in (3.20) and writing out the average over ψ as a multiple integral over the random variables c_ϵ introduced in Sec. 2, we obtain

$$\ln \langle \exp A\{\psi\} \rangle_{\text{av } \psi} \cong -\tfrac{1}{2} \sum_{\sigma \neq 0} \ln (1 - \beta v_\sigma \mu_{2,\sigma}), \quad (3.22)$$

where the coefficients $\mu_{2,\sigma}$ are defined by[11]

$$\int_V \mu_2(\mathbf{r}, \mathbf{r}' \mid y) u_\sigma(\mathbf{r}')\, d^{\nu} r' = \mu_{2,\sigma} u_\sigma(\mathbf{r}). \quad (3.23)$$

The result (3.22) was obtained by Zittartz.[5] The result of Hemmer[4] may be obtained from (3.22) by replacing $\mu_{2,\sigma}$ by $\mu_{2,0}$ where

$$\mu_{2,0} = \rho_h \partial \rho_h / \partial \beta P_h. \quad (3.24)$$

Equation (3.24) is a special case of the general fluctuation theorem

$$\int_V d^{\nu} r_2 \cdots \int_V d^{\nu} r_n \mu_n(\mathbf{r}_1, \cdots, \mathbf{r}_n \mid y) = \partial^n \beta P_h(y) / \partial (\ln y)^n. \quad (3.25)$$

Zittartz's result contains all contributions to $h(y)$ of orders γ^{ν} through $\gamma^{2\nu-1}$ inclusive. (It also contains, in general, higher-order terms.) Hemmer's result gives the terms proportional to γ^{ν} exactly, because the coefficients v_σ decrease rapidly while the coefficients $\mu_{2,\sigma}$ decrease slowly in Fourier space.

In order to obtain terms of higher order in γ, we make a transformation in function space similar to the one which turned out to be useful in the case of the Ising model.[12] We introduce a function $\chi(\mathbf{r} - \mathbf{r}')$, periodic with period $V^{1/\nu}$, and its expansion

$$\chi(\mathbf{r} - \mathbf{r}') = \sum_\sigma \chi_\sigma u_\sigma(\mathbf{r}) u_\sigma(\mathbf{r}'). \quad (3.26)$$

This function is arbitrary except for the restriction

$$1 - \beta v_\sigma \chi_\sigma > \theta, \quad (3.27)$$

where θ is a fixed positive number which we impose. Starting from the form

$$\langle e^{A\{\psi\}} \rangle_{\text{av } \psi} = \int_{-\infty}^{\infty} \cdots \int \prod_\sigma{}' \left(\frac{dc_\sigma}{(2\pi)^{\frac{1}{2}}} \right) \exp \left[-\tfrac{1}{2} \sum_\sigma{}' c_\sigma^2 + A \left\{ \sum_\sigma{}' c_\sigma (\beta v_\sigma)^{\frac{1}{2}} u_\sigma(r) \right\} \right], \quad (3.28)$$

where the prime indicates exclusion of $\sigma = 0$, we add and subtract in the exponent

$$\tfrac{1}{2} \sum_\sigma{}' c_\sigma^2 \beta v_\sigma \chi_\sigma \quad (3.29)$$

[11] Under the assumption of Born–von Kármán conditions, the function $\mu_2(\mathbf{r}, \mathbf{r}' \mid y)$ depends only on the relative vectorial distance $\mathbf{r} - \mathbf{r}'$, so that its eigenfunctions are also the trigonometric functions $u_\sigma(\mathbf{r})$.

[12] A. J. F. Siegert and D. J. Vezzetti, J. Math. Phys. 9, 2173 (1968).

and obtain

$$\langle e^{A\{\psi\}}\rangle_{\mathrm{av}\,\psi} = \int_{-\infty}^{\infty}\cdots\int \prod_\sigma{}' \left(\frac{dc_\sigma}{(2\pi)^{\frac{1}{2}}}\right) \times \exp\left[-\tfrac{1}{2}\sum_\sigma{}' c_\sigma^2(1-\beta v_\sigma\chi_\sigma) + A\left\{\sum_\sigma{}' c_\sigma(\beta v_\sigma)^{\frac{1}{2}}u_\sigma(\mathbf{r})\right\} - \tfrac{1}{2}\sum_\sigma{}' c_\sigma^2\beta v_\sigma\chi_\sigma\right]. \tag{3.30}$$

We now introduce new variables b_σ defined by

$$b_\sigma \equiv c_\sigma(1-\beta v_\sigma\chi_\sigma)^{\frac{1}{2}} \tag{3.31}$$

and obtain

$$\langle e^{A\{\psi\}}\rangle_{\mathrm{av}\,\psi} = \prod_\sigma{}'(1-\beta v_\sigma\chi_\sigma)^{-\frac{1}{2}}\int_{-\infty}^{\infty}\cdots\int \prod_\sigma{}' \left(\frac{db_\sigma}{(2\pi)^{\frac{1}{2}}}\right) \times \exp\left[-\tfrac{1}{2}\sum_\sigma{}' b_\sigma^2 + A\left\{\sum_\sigma{}' b_\sigma\left(\frac{\beta v_\sigma}{1-\beta v_\sigma\chi_\sigma}\right)^{\frac{1}{2}}u_\sigma(\mathbf{r})\right\} - \tfrac{1}{2}\sum_\sigma{}' b_\sigma^2\frac{\beta v_\sigma\chi_\sigma}{1-\beta v_\sigma\chi_\sigma}\right]. \tag{3.32}$$

Introducing new random functions $\tilde{\psi}(\mathbf{r})$ defined by

$$\tilde{\psi}(\mathbf{r}) \equiv \sum_\sigma{}' b_\sigma\left(\frac{\beta v_\sigma}{1-\beta v_\sigma\chi_\sigma}\right)^{\frac{1}{2}}u_\sigma(\mathbf{r}), \tag{3.33}$$

where the random variables b_σ are defined by the probability density

$$\prod_\sigma{}'\left(\frac{db_\sigma}{(2\pi)^{\frac{1}{2}}}e^{-\frac{1}{2}b_\sigma^2}\right), \tag{3.34}$$

we can eliminate the explicit representation and write (3.32) in the form

$$\langle e^{A\{\psi\}}\rangle_{\mathrm{av}\,\psi} = \prod_\sigma{}'(1-\beta v_\sigma\chi_\sigma)^{-\frac{1}{2}}\left\langle \exp\left[A\{\tilde{\psi}\} - \tfrac{1}{2}\int_V d^\nu r_1\, d^\nu r_2\tilde{\psi}(\mathbf{r}_1)\chi(\mathbf{r}_1-\mathbf{r}_2)\tilde{\psi}(\mathbf{r}_2)\right]\right\rangle_{\mathrm{av}\,\tilde{\psi}}. \tag{3.35}$$

The functions $\tilde{\psi}(\mathbf{r})$ are Gaussian random functions with mean value $\langle\tilde{\psi}(\mathbf{r})\rangle_{\mathrm{av}\,\tilde{\psi}} = 0$ and covariance

$$\langle\tilde{\psi}(\mathbf{r})\tilde{\psi}(\mathbf{r}')\rangle_{\mathrm{av}\,\tilde{\psi}} = \beta\tilde{v}(\mathbf{r},\mathbf{r}'\mid\beta), \tag{3.36}$$

where

$$\tilde{v}(\mathbf{r},\mathbf{r}'\mid\beta) = \tilde{v}(\mathbf{r}-\mathbf{r}'\mid\beta) = \sum_\sigma{}'\tilde{v}_\sigma u_\sigma(\mathbf{r})u_\sigma(\mathbf{r}') \tag{3.37}$$

with

$$\tilde{v}_\sigma \equiv v_\sigma/(1-\beta v_\sigma\chi_\sigma). \tag{3.38}$$

We note that the function

$$\tilde{v}_1(\mathbf{r}-\mathbf{r}'\mid\beta) \equiv \tilde{v}(\mathbf{r}-\mathbf{r}'\mid\beta) + \tilde{v}_0/V \tag{3.39}$$

satisfies the integral equation

$$\tilde{v}_1(\mathbf{r}-\mathbf{r}'\mid\tau) = v(\mathbf{r}-\mathbf{r}') + \tau\int_V d^\nu r''K(\mathbf{r}-\mathbf{r}'')\tilde{v}_1(\mathbf{r}''-\mathbf{r}'\mid\tau), \tag{3.40}$$

where

$$K(\mathbf{r}-\mathbf{r}'') = \int_V d^\nu r' v(\mathbf{r}-\mathbf{r}')\chi(\mathbf{r}'-\mathbf{r}''). \tag{3.41}$$

This transformation thus extracts from the average a factor of the form of Zittartz's approximation (3.22). For the special choice of the arbitrary function χ, namely $\chi = \mu_2$, this factor becomes identical with the result of Zittartz.

This factor can be expressed in terms of the function $\tilde{v}(\mathbf{r}\mid\tau)$. Since

$$\frac{\partial}{\partial\tau}\ln\prod_\sigma{}'(1-\tau v_\sigma\chi_\sigma)^{-\frac{1}{2}} = \tfrac{1}{2}\sum_\sigma{}'\tilde{v}_\sigma\chi_\sigma = \tfrac{1}{2}V\int_V d^\nu r\tilde{v}(\mathbf{r}\mid\tau)\chi(\mathbf{r}), \tag{3.42}$$

we have

$$-\tfrac{1}{2}\sum_\sigma{}'\ln(1-\beta v_\sigma\chi_\sigma) = \tfrac{1}{2}V\int_0^\beta d\tau\int_V d^\nu r\tilde{v}(\mathbf{r}\mid\tau)\chi(\mathbf{r}). \tag{3.43}$$

With these results, we may now write $h(y)$ in the form

$$h(y) = \tfrac{1}{2}\int_0^\beta d\tau\int_V d^\nu r\tilde{v}(\mathbf{r}\mid\tau)\chi(\mathbf{r}) + V^{-1}\ln\left\langle\exp\left[A_3\{\tilde{\psi}\} + \tfrac{1}{2}\int_V d^\nu r_1\, d^\nu r_2 \times\tilde{\psi}(r_1)(\mu_2(\mathbf{r}_1,\mathbf{r}_2\mid y) - \chi(\mathbf{r}_1,\mathbf{r}_2))\tilde{\psi}(r_2)\right]\right\rangle_{\mathrm{av}\,\tilde{\psi}}, \tag{3.44}$$

where

$$A_3\{\tilde{\psi}\} \equiv \sum_{n=3}^{\infty}\frac{1}{n!}\int_V d^\nu r_1\cdots d^\nu r_n\mu_n(\mathbf{r}_1,\cdots,\mathbf{r}_n\mid y)\tilde{\psi}(\mathbf{r}_1)\cdots\tilde{\psi}(\mathbf{r}_n). \tag{3.45}$$

Before we proceed to a systematic expansion of $h(y)$, we note that we can quickly derive the next-order result, which includes the result of Lebowitz, Stell, and Baer.[2] With the above-mentioned special choice of $\chi(\mathbf{r}_1,\mathbf{r}_2)$, we expand the logarithm of the average in (3.44) in a cumulant expansion, i.e.,

$$\ln\langle e^{A_3\{\tilde{\psi}\}}\rangle_{\mathrm{av}\,\tilde{\psi}} = \langle A_3\{\tilde{\psi}\}\rangle_{\mathrm{av}\,\tilde{\psi}} + \tfrac{1}{2}[\langle A_3^2\{\tilde{\psi}\}\rangle_{\mathrm{av}\,\tilde{\psi}} - \langle A_3\{\tilde{\psi}\}\rangle^2_{\mathrm{av}\,\tilde{\psi}}] + \cdots \tag{3.46}$$

and retain only the first nonvanishing term in each cumulant (the term with μ_4 in $\langle A_3\rangle$ and the terms with

two μ_3's in $[\langle A_3^2\rangle - \langle A_3\rangle^2]$). We then have, as we will show later, all terms which contribute to $h(y)$ through order $\gamma^{3\nu-1}$. This may be shown to contain all contributions to the free energy through order $\gamma^{3\nu-1}$ and thus includes the terms through order $\gamma^{2\nu}$ which have been calculated explicitly by Lebowitz, Stell, and Baer.

SECTION 4

We now proceed to a systematic evaluation of $h(y)$ with the special choice $\chi = \mu_2$. Defining $h^{(1)}(y)$ by

$$e^{Vh^{(1)}(y)} \equiv \langle e^{A_3\{\tilde{\varphi}\}}\rangle_{\text{av }\tilde{\varphi}}, \tag{4.1a}$$

we have

$$e^{Vh^{(1)}(y)} = \left\langle \prod_{n=3}^{\infty} \exp \int_V \frac{\mu_n(\mathbf{r}_1, \cdots, \mathbf{r}_n \mid y)}{n!} \prod_{j=1}^{n} \tilde{\varphi}(\mathbf{r}_j)\, d^{\nu n} r \right\rangle_{\text{av }\tilde{\varphi}}. \tag{4.1b}$$

Expanding each factor of the product, we obtain

$$\begin{aligned} e^{Vh^{(1)}(y)} &= \left\langle \sum_{m_3=0}^{\infty} \sum_{m_4=0}^{\infty} \cdots \prod_{n=3}^{\infty} \frac{1}{m_n!} \left(\int_V \frac{\mu_n}{n!} \prod_{j=1}^{n} \tilde{\varphi}(\mathbf{r}_j)\, d^{\nu n} r \right)^{m_n} \right\rangle_{\text{av }\tilde{\varphi}} \\ &= 1 + \sum_{\alpha=3}^{\infty} {\sum_{m_3, m_4, \cdots}}' \prod_{n=3}^{\infty} \frac{1}{m_n!\,(n!)^{m_n}} \\ &\quad \times \int_V \prod_{n=3}^{\infty} (\mu_n)^{(m_n)} \left\langle \prod_{j=1}^{\alpha} \tilde{\varphi}(\mathbf{r}_j) \right\rangle_{\text{av }\tilde{\varphi}} d^{\nu\alpha} r, \end{aligned} \tag{4.2}$$

where the prime indicates the restriction

$$\sum_{n=3}^{\infty} n m_n = \alpha \tag{4.3}$$

and where

$$\prod_{n=3}^{\infty} (\mu_n)^{(m_n)}$$

is the product of m_3 factors μ_3, m_4 factors μ_4, etc., with the arguments r_j $(j = 1, \cdots, \alpha)$ in natural order.

The average in (4.2) is zero if α is odd. For α even, we have

$$\langle \tilde{\varphi}(\mathbf{r}_1) \cdots \tilde{\varphi}(\mathbf{r}_\alpha)\rangle_{\text{av }\tilde{\varphi}} = \sum_{\text{(pairings)}} \beta\tilde{v}(\mathbf{r}_{i_1} - \mathbf{r}_{i_2}) \cdots \beta\tilde{v}(\mathbf{r}_{i_{\alpha-1}} - \mathbf{r}_{i_\alpha}). \tag{4.4}$$

The sum is to be taken over all different ways of pairing the variables, that is, over all different ways of choosing $\alpha/2$ pairs (i_1, i_2), (i_3, i_4), $\cdots$, $(i_{\alpha-1}, i_\alpha)$ from the numbers $1, 2, \cdots, \alpha$. Two ways of pairing count as different only if they cannot be made the same by a permutation of the pairs and a pair is defined regardless of order.

The terms resulting from substituting (4.4) into (4.2) can now be associated with labeled diagrams. The diagrams are defined as follows: A circle containing n points $\mathbf{r}_{i_1}, \cdots, \mathbf{r}_{i_n}$ represents $\mu_n(\mathbf{r}_{i_1}, \cdots, \mathbf{r}_{i_n} \mid y)$. A line connecting points $\mathbf{r}_{i_k}$ and $\mathbf{r}_{i_{k'}}$ represents $\beta\tilde{v}(\mathbf{r}_{i_k} - \mathbf{r}_{i_{k'}})$. Each integral in (4.2) is therefore associated with a diagram consisting of S circles ($S \geq 1$) containing the appropriate number of points, these points labeled $1, 2, \cdots, \alpha$, and $\alpha/2$ lines connecting these points such that each point is connected to one and only one other point by a line. By observing the order in V or by the usual diagram arguments, it now follows that

$$Vh^{(1)}(y) = \sum_{\alpha=4}^{\infty} {\sum_{m_3, m_4, \cdots}}' \prod_{n=3}^{\infty} \frac{1}{m_n!\,(n!)^{m_n}} \int_V \prod_{n=3}^{\infty} (\mu_n)^{(m_n)} \times \sum_{\text{(pairings)}}' \beta\tilde{v}(\mathbf{r}_{i_1} - \mathbf{r}_{i_2}) \cdots \beta\tilde{v}(\mathbf{r}_{i_{\alpha-1}} - \mathbf{r}_{i_\alpha})\, d^{\nu\alpha} r, \tag{4.5}$$

where the prime on the last sum indicates that only such pairings are permitted which result in a connected diagram.

Many of the integrals have of necessity the same value, because they arise from different ways of labeling the variables of integration. It is therefore sufficient to take into account in Eq. (4.5) only a subset $\mathcal{G}$ of the original diagrams obtained by selecting only one diagram G from all the connected diagrams which give the same value, I_G, for the integral. Thus, we have

$$h^{(1)}(y) = V^{-1} \sum_{G \in \mathcal{G}} \prod_{n=3}^{\infty} \frac{1}{m_n!\,(n!)^{m_n}} \mathcal{N}_G I_G, \tag{4.6}$$

where $\mathcal{N}_G$ is the number of integrals in (4.5) which have the value I_G. The number $\mathcal{N}_G$ can be obtained as the product of the total number of permutations of the points in each circle times the total number of permutations of the circles with the same number of points divided by the number S_G of such permutations which give the same pairing. The first number is just

$$\prod_{n=3}^{\infty} m_n!\,(n!)^{m_n}$$

and cancels the denominator in (4.6). We can thus write (4.6) in the form

$$h^{(1)}(y) = V^{-1} \sum_{G \in \mathcal{G}} I_G / S_G. \tag{4.7}$$

The symmetry number S_G can be written as follows: Consider the graph G with the circles labeled from 1 to $\sum_n m_n$, and the points unlabeled. Let t_{ij} be the number of lines connecting circle i and circle j (t_{ii} is not necessarily zero). The matrix t_{ij} characterizes the graph with labeled circles and unlabeled points

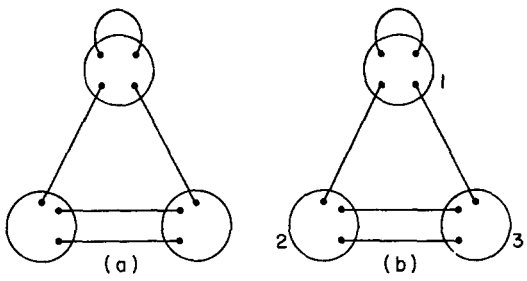

FIG. 2. Example of diagram.

uniquely. Then

$$S_G = \prod_{i<j} t_{ij}! \prod_i (2^{t_{ii}} t_{ii}!) S'_G, \tag{4.8}$$

where S'_G is the number of permutations, including the identity, of the labels $1, 2, \cdots, \sum_{n=3}^{\infty} m_n$ which leave the matrix t_{ij} invariant. The factor preceding S'_G is just the number of ways of labeling the points in each circle which all give rise to the same pairing.

As an example, we calculate the coefficient of

$$I_G \equiv V^{-1}\int_V \mu_4(1,2,3,4)\mu_3(5,6,7)\mu_3(8,9,10) \times \beta^5 \tilde{v}(1,2)\tilde{v}(3,5)\tilde{v}(4,10)\tilde{v}(6,9)\tilde{v}(7,8)\, d^{10\nu}r, \tag{4.9}$$

which corresponds to the diagram in Fig. 2(a). To evaluate the coefficient by (4.8) we label the circles as indicated in Fig. 2(b) and obtain $t_{11} = t_{12} = t_{13} = 1$, $t_{23} = 2$, all others being equal to zero, and $S'_G = 2$, since only the identity and the permutation $\binom{123}{132}$ leave the t_{ij} invariant. Thus $S_G = (2!)(2^1)(2) = 8$.

To evaluate the coefficient from (4.6) we note that $m_3 = 2$, $m_4 = 1$. To obtain $\mathcal{N}_G$, the number of pairings, we note that the re-entrant line of circle 1 can be chosen in $\binom{4}{2}$ ways and the two lines connecting circles 2 and 3 can be chosen in $\binom{3}{2}^2 \cdot 2$ ways. The remaining point in circle 2 can be paired with either of the two remaining points in circle 1, giving another factor 2, and we obtain $\mathcal{N}_G = 216$ and

$$\prod \frac{1}{m_n!\,(n!)^{m_n}} \mathcal{N}_G = \frac{1}{2!\,1!\,(3!)^2 4!} \cdot \frac{4!}{(2!)^2} \frac{(3!)^2}{(2!)^2} \cdot 2 = \frac{1}{8}, \tag{4.10}$$

in agreement with the value obtained from (4.8). It is clear that (4.8) is more convenient than (4.6), especially for more elaborate diagrams.

SECTION 5

Having defined the diagrams and determined the symmetry numbers S_G, we will now estimate the order in γ of any given diagram. The range of the factors $\tilde{v}(\mathbf{r}_i, \mathbf{r}_j)$ is at least of order γ^{-1}, while the range of μ_n is independent of γ by definition and is of order n times the hard-core diameter. For purposes of estimation, we can then replace each μ_n by

$$\mu_n(\mathbf{r}_1, \cdots, \mathbf{r}_n \mid y) \sim \prod_{k=2}^{n} \delta(\mathbf{r}_1 - \mathbf{r}_k) \int_V \mu_n(\mathbf{r}_1, \cdots, \mathbf{r}_n \mid y)\, d^\nu r_2 \cdots d^\nu r_n = \prod_{k=2}^{n} \delta(\mathbf{r}_1 - \mathbf{r}_k) \frac{\partial^n \beta P_h(y)}{\partial(\ln y)^n}. \tag{5.1}$$

The order in γ of any integral I_G is then the order of

$$V^{-1}(\tilde{v}(0))^S \int_V d^\nu r_1 \cdots d^\nu r_S \prod_{1\le i<j\le S} [\tilde{v}(\mathbf{r}_i - \mathbf{r}_j)]^{t_{ij}}, \tag{5.2}$$

where $S \equiv \sum_i t_{ii}$ and where the matrix t_{ij} was defined in Sec. 4, and S is the number of circles in the diagram. For $v(\mathbf{r}) = \gamma^\nu \Phi(\gamma \mathbf{r})$, we find, using (5.1) in (3.39)–(3.41) with $\chi = \mu_2$,

$$\tilde{v}(\mathbf{r}) = \gamma^\nu f(\gamma \mathbf{r}). \tag{5.3}$$

The exact form of μ_2 would contribute to $\tilde{v}$ only terms of higher order in γ. Introducing new coordinates $\mathbf{x}_i = \gamma \mathbf{r}_i$, we obtain from (5.2)

$$(\gamma^\nu f(0))^S \int_V d^\nu x_2 \cdots d^\nu x_S \gamma^{\nu(1-S)} \prod_{1\le i<j\le S} [\gamma^\nu f(\mathbf{x}_i - \mathbf{x}_j)]^{t_{ij}}, \tag{5.4}$$

where S is defined as after Eq. (5.2). If $f(0)$ and

$$\int_V d^\nu x_2 \cdots d^\nu x_S \prod_{1\le i<j\le S} [f(\mathbf{x}_i - \mathbf{x}_j)]^{t_{ij}}$$

are finite even for $V \to \infty$ (the proof and discussion are in Appendix B), then (5.4) is of order

$$\gamma^{\nu[S'+1-S]} \tag{5.5}$$

in γ. Here,

$$S' \equiv \sum_{1\le i\le j\le S} \tau_{ij}.$$

The sum in the bracket is just the number of lines in the diagram which we will denote by B. We can summarize by writing, for a diagram with B bonds

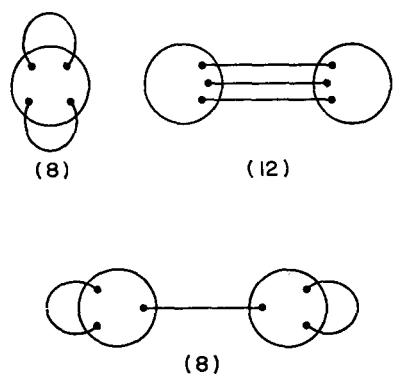

FIG. 3. Diagrams needed through order $\gamma^{3\nu-1}$. The last of these diagrams gives a contribution of zero. Symmetry numbers are shown in parenthesis.

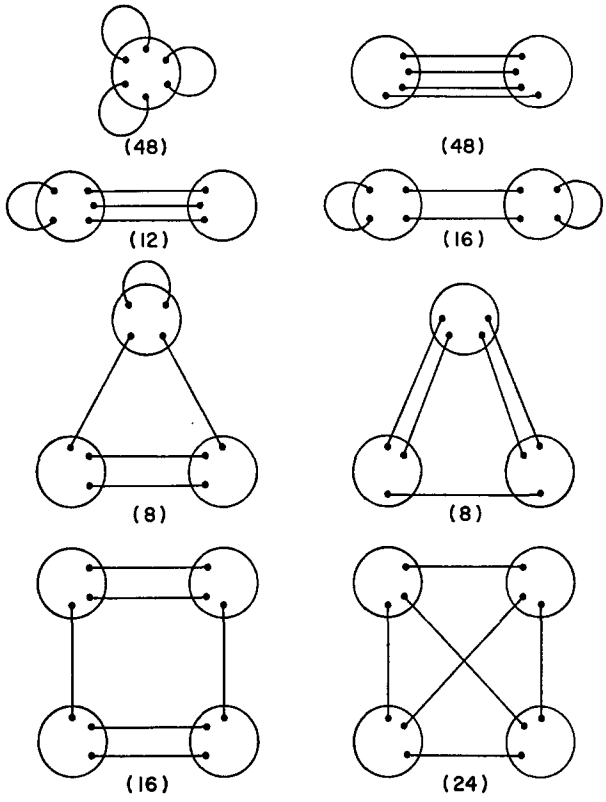

FIG. 4. Diagrams needed through order $\gamma^{4\nu-1}$. Symmetry numbers are shown in parenthesis.

and S circles, the dominant contribution is of order

$$\gamma^{\nu[B+1-S]}. \tag{5.6}$$

In this manner, each diagram in the sum we are considering may be classified according to its dominant order in γ, and the sum of diagrams may be ordered accordingly.

From (5.6) we can prove that each term in $h^{(1)}(y)$ is of order at least $\gamma^{2\nu}$. This is seen as follows. Since each circle contains at least three points, we have $B \geq \frac{3}{2}$. The expression (5.6) then shows that all diagrams with S circles are of order at least $\gamma^{\nu[S/2+1]}$. For $S \geq 2$, we obtain terms of order $\gamma^{2\nu}$ or higher. If $S = 1$, then $B \geq 2$, and (5.6) then gives $\gamma^{2\nu}$ or higher. Q.E.D.

In order to obtain all terms whose dominant contributions are of order $\gamma^{\nu M}$, we simply write

$$B + 1 - S = M \tag{5.7}$$

and consider the various $B\,(\geq 2)$ and $S\,(\geq 1)$ $(B \geq 3S/2)$ which can give this M. To obtain the expansion through order $\gamma^{\nu M}$, only diagrams with not more than $2(M-1)$ circles need to be considered. (Since the next dominant term is of order $\gamma^{\nu(M+1)}$, this actually gives all terms through order $\gamma^{\nu(M+1)-1}$.)

For instance, for $M = 2$, we have to take diagrams with $S = 1, 2$, i.e., the diagrams in Fig. 3. The last of these diagrams has the value zero. The first two diagrams of Fig. 3 are all that we need in order to compute the thermodynamic functions through order $\gamma^{3\nu-1}$, since the remaining diagrams contribute $\gamma^{3\nu}$ and higher. These two diagrams include the result of Lebowitz, Stell, and Baer.[2] To compute the thermodynamic functions to order $\gamma^{4\nu-1}$ we need those diagrams whose dominant contributions are of order $\gamma^{3\nu}$. For $M = 3$, we need $S = 1, 2, 3, 4$. The only ones of these diagrams which are not necessarily zero are shown in Fig. 4.

SECTION 6

We have shown how to evaluate the function $h(y)$ as an expansion in γ. Combining our results to this point, we have

$$h(y) = \tfrac{1}{2}\int_0^\beta d\tau \int_V d^\nu r \tilde{v}(\mathbf{r} \mid \tau)\mu_2(\mathbf{r} \mid y) + h^{(1)}(y), \tag{6.1}$$

where $h^{(1)}(y)$ is given by the diagram sum (4.7).

To determine the thermodynamic functions of our system we must evaluate the pressure given by

$$\beta P(z) = \lim_{V\to\infty} \frac{1}{V} \ln \int_{-\infty}^{\infty} \frac{dc_0}{(2\pi)^{\frac{1}{2}}} \times \exp\left[-\tfrac{1}{2}c_0^2 + \beta V P_h(y) + V h(y)\right], \tag{6.2}$$

where y is given by (3.8). The saddle-point approximation gives

$$\beta P(z) = \max_{c_0}\left[-\frac{c_0^2}{2V} + \beta P_h(y) + h(y)\right]. \tag{6.3}$$

The extrema are the roots of the equation

$$c_0 = V\frac{\partial}{\partial c_0}\left[\beta P_h(y) + h(y)\right]. \tag{6.4}$$

Since

$$\frac{\partial}{\partial c_0} = \left(\frac{\beta v_0}{V}\right)^{\frac{1}{2}} \frac{\partial}{\partial \ln y}, \tag{6.5}$$

the extrema are the roots of the equation

$$\ln y - \ln\left(z e^{-\beta v(0)/2}\right) = \beta v_0\left[\rho_h(y) + \frac{\partial h(y)}{\partial \ln y}\right], \tag{6.6}$$

where $\rho_h(y)$ is the density of the hard-core system at fugacity y,

$$\rho_h(y) = \frac{\partial \beta P_h(y)}{\partial \ln y}. \tag{6.7}$$

We then have

$$\beta P(z) = -\frac{\left[\ln y^* - \ln\left(z e^{-\beta v(0)/2}\right)\right]^2}{2\beta v_0} + \beta P_h(y^*) + h(y^*), \tag{6.8}$$

where y^* is that root of (6.6) which gives the largest value of $P(z)$.

To obtain the density $\rho(z)$, we differentiate (6.8) and get

$$\rho(z) = \frac{d\beta P(z)}{d \ln z} = \left[\frac{\partial}{\partial \ln z} + \frac{d \ln y^*}{d \ln z}\frac{\partial}{\partial \ln y^*}\right]\beta P(z) = \frac{\ln y^* - \ln (ze^{-\beta v(0)/2})}{\beta v_0}, \quad (6.9)$$

since the second term in the brackets does not contribute because of (6.8). Using (6.6) once more, we have

$$\rho(z) = \rho_h(y^*) + \frac{\partial h(y^*)}{\partial \ln y^*} \quad (6.10)$$

and

$$\beta P(z) = \beta P_h(y^*) - \tfrac{1}{2}\beta v_0 \rho^2(z) + h(y^*). \quad (6.11)$$

Equations (6.6), (6.10), and (6.11) constitute our formal result. For $h(y) \to 0$, this result approaches the van der Waals equation with Maxwell construction in the grand canonical form as discussed in Sec. 3, starting with Eq. (3.8).

In order to compare our results with those of previous authors who have given the Helmholtz free energy as a function of density (ρ) through order $\gamma^{2\nu}$, we have to eliminate the fugacity from our equations. We first change variables from y^* and z to ρ_h and ρ by the definitions

$$\rho(z) \equiv \rho, \quad (6.12a)$$

$$\rho_h(y^*) \equiv \beta_h, \quad (6.12b)$$

$$P(z) \equiv P(z(\rho)) \equiv p(\rho), \quad (6.12c)$$

$$P_h(y^*) \equiv P_h(y^*(\rho_h)) \equiv p_h(\rho_h), \quad (6.12d)$$

$$h(y^*) \equiv \eta(\rho_h), \quad (6.12e)$$

$$\partial h(y^*)/\partial \ln y^* \equiv \eta_1(\rho_h). \quad (6.12f)$$

Equations (6.10) and (6.11) then become

$$\beta p(\rho) = \beta p_h(\rho_h) + \eta(\rho_h) - \tfrac{1}{2}\beta v_0 \rho^2 \quad (6.13)$$

and

$$\rho = \rho_h + \eta_1(\rho_h). \quad (6.14)$$

Eliminating ρ_h from these equations, we obtain $p(\rho)$. This is conveniently done by using the relation between Helmholtz free energy per particle $A(\rho)$ and the fugacity z:

$$\beta A(\rho) = \ln z - \beta P(z)/\rho(z). \quad (6.15)$$

Using (6.6) for $\ln z$ and (6.13) and (6.14), we obtain

$$\beta A(\rho) = \beta A_h(\rho_h) + \tfrac{1}{2}\beta v(0) - \tfrac{1}{2}\beta v_0(\rho_h + \eta_1(\rho_h))^2 + \frac{\beta p_h(\rho_h)\eta_1(\rho_h) - \rho_h \eta(\rho_h)}{\rho_h(\rho_h + \eta_1(\rho_h))} \equiv F(\rho_h), \quad (6.16)$$

with

$$\beta A_h(\rho_h) \equiv \ln y^* - \beta P_h(y^*)/\rho_h(y^*) \quad (6.17)$$

and $A_h(\rho_h)$ is the Helmholtz free energy per particle of the hard-core system at density ρ_h. Using (6.14), we can express the left side of Eq. (6.16) in powers of η_1:

$$\beta A(\rho) = \beta A(\rho_h + \eta_1(\rho_h)) = \beta A(\rho_h) + \sum_{n=1}^{\infty} \frac{(\eta_1(\rho_h))^n}{n!}\frac{d^n}{d\rho_h^n}\beta A(\rho_h). \quad (6.18)$$

Inserting (6.18) into (6.16), we see that we have completely eliminated the variable ρ. Thus ρ_h takes on the status of independent variable, and we can rename it ρ. Hence, we obtain

$$\beta A(\rho) = F(\rho) - \sum_{n=1}^{\infty} \frac{(\eta_1(\rho))^n}{n!}\frac{d^n}{d\rho^n}\beta A(\rho) \equiv F(\rho) - \hat{q}\beta A(\rho), \quad (6.19)$$

where the last line defines the (differential) operator $\hat{q}$. Solving this equation by iteration, we obtain the final result

$$\beta A(\rho) = \sum_{n=0}^{\infty}(-\hat{q})^n F(\rho). \quad (6.20)$$

By this equation, we have succeeded in expressing the free energy in terms of the hard-core functions A_h and p_h, which we assume are known, and the function η_1. The equation of state follows from (6.20) in the usual manner:

$$\beta p(\rho) = \rho^2 \frac{\partial}{\partial \rho}\beta A(\rho) = \rho^2 \frac{\partial}{\partial \rho}\sum_{n=0}^{\infty}(-\hat{q})^n F(\rho). \quad (6.21)$$

For reference purposes and for our future use, we give $A(\rho)$ explicitly to third order in η:

$$\beta A(\rho) = \beta a_h(\rho) - \tfrac{1}{2}\beta v_0 \rho + \tfrac{1}{2}\beta v(0) - [\eta(\rho)/\rho] + \frac{1}{2}\frac{(\eta'(\rho))^2}{\beta p_h'(\rho)} - \frac{1}{6\rho}\frac{\partial}{\partial \rho}\left(\frac{\rho^2(\eta'(\rho))^3}{(\beta p_h'(\rho))^2}\right) + O(\eta^4), \quad (6.22)$$

where the prime indicates differentiation with respect to ρ. The terms explicitly exhibited in (6.22) are sufficient to obtain the Helmholtz free energy to order $\gamma^{4\nu-1}$. Evaluated to order $\gamma^{2\nu}$, this equation gives the results of Lebowitz, Stell, and Baer,[2] which in turn contains the results of Zittartz[5] and Hemmer.[4]

Since our result in its grand canonical form includes the Maxwell construction, it is understood that this construction in its appropriate form is to be applied to these equations.

SECTION 7

In this section we summarize and discuss our results, first in a special form, then in a more general form.

The pressure $P(z)$ and number density $\rho(z)$ of the system described in Sec. 2 are obtained in the form

$$\beta P(z) = \beta P_h(y^*) - \tfrac{1}{2}\beta v_0 \rho^2(z) + h(y^*), \quad (7.1)$$

$$\rho(z) = \rho_h(y^*) + \frac{\partial h(y^*)}{\partial \ln y^*}, \quad (7.2)$$

where y^* is that root of the equation

$$\ln y - \ln (z e^{-\frac{1}{2}\beta v(0)}) = \beta v_0 \left[\rho_h(y) + \frac{\partial h(y)}{\partial \ln y} \right], \quad (7.3)$$

which maximizes the pressure $P(z)$. Here $P_h(y)$ and $\rho_h(y)$ are the pressure and number density of the hard-core system at fugacity y and v_0 is defined by Eq. (2.5). The function $h(y)$ is given by

$$h(y) = \tfrac{1}{2}\int_0^\beta d\tau \int_V d^\nu r \tilde{v}(\mathbf{r} \mid \tau) \mu_2(\mathbf{r} \mid y) + h^{(1)}(y) \quad (7.4)$$

and $h^{(1)}(y)$ is a diagram series defined as follows. Draw S circles ($S \geq 1$) labeled $C_1, \cdots, C_j, \cdots, C_S$, each containing $n_j \geq 3$ points such that $\sum_{j=1}^S n_j$ is even. Connect these points by lines such that each point is connected to one and only one other point. (Lines connecting two points in the same circle are permitted.) Disconnected diagrams and diagrams which would become disconnected upon removal of one line are excluded. With any diagram G, we associate an integral I_G in the following way. Label the points in circle C_j by $\mathbf{r}_{j,1}, \mathbf{r}_{j,2}, \cdots, \mathbf{r}_{j,n_j}$. For each circle C_j write a factor $\mu_{n_j}(\mathbf{r}_{j,1}, \cdots, \mathbf{r}_{j,n_j} \mid y)$, where μ_n is the nth modified Ursell function of the hard-core system at fugacity y, defined in Ref. 6 and denoted there by $\hat{\mathcal{F}}_n$. For a line connecting points $\mathbf{r}_{j,k}$ and $\mathbf{r}_{j',k'}$ write a factor $\beta\tilde{v}(\mathbf{r}_{j,k} - \mathbf{r}_{j',k'})$, where $\tilde{v}$ is defined by Eqs. (3.38)–(3.41) with the special choice

$$\chi(\mathbf{r} - \mathbf{r}') = \mu_2(\mathbf{r}, \mathbf{r}' \mid y).$$

I_G is then the integral over all coordinates of the product of these factors. With any diagram G, we also associate a symmetry number S_G. Let t_{ij} be the number of lines connecting circles C_i and C_j and let S'_G be the number of permutations of the labels of the circles which leave all t_{ij} invariant. Then

$$S_G = \prod_{i \leq j} (t_{ij}!) \prod_i 2^{t_{ii}} S'_G. \quad (7.5)$$

We call two diagrams G and G' equivalent if the integrands of I_G and $I_{G'}$ are the same, except for the labeling of the variables of integration. Divide the class of diagrams described above into subclasses of equivalent diagrams. Select one member from each subclass to form the class $\mathfrak{G}$. Then

$$h^{(1)}(y) = V^{-1} \sum_{G \in \mathfrak{G}} I_G / S_G. \quad (7.6)$$

We have shown that a diagram with B bonds and S circles cannot contribute any power of γ less than $\nu[B - S + 1]$ and that the coefficient of $\gamma^{\nu[B-S+1]}$ is finite provided that

$$1 - \beta v_0 \frac{\partial \rho_h(y)}{\partial \ln y} > \theta, \quad (7.7)$$

where θ is any fixed, positive number. The meaning of this condition is explained more fully in Appendix B, Eqs. (B5)–(B6). The requirement is a direct consequence of the condition (3.27) imposed on the arbitrary function $\chi(\mathbf{r}, \mathbf{r}')$ and the special choice $\chi(\mathbf{r}, \mathbf{r}') = \mu_2(\mathbf{r}, \mathbf{r}' \mid y)$. This condition can be weakened by not making this special choice of χ. We hope to exploit the freedom of choosing χ in a future paper and for that reason have kept the present development general except for the diagram series and the estimates of the individual terms. The diagrams for the case $\chi \neq \mu_2$ are of the same type, except that now circles with two points are permitted and are interpreted as $[\mu_2(\mathbf{r}, \mathbf{r}' \mid y) - \chi(\mathbf{r}, \mathbf{r}')]$, and the bonds are interpreted as $\tilde{v}(\mathbf{r} - \mathbf{r}' \mid \beta)$, defined by the original equation (3.37). The estimate of these diagrams is not difficult if condition (3.27) is satisfied. However, the introduction of $\chi \neq \mu_2$ may be expected to lead to interesting results only when this condition is relaxed to the extent of having $1 - \beta v_\sigma \chi_\sigma$ (for some σ) go to zero for $\gamma \to 0$ though not as strongly as $1 - \beta v_0 \mu_{2,0}$.

With $\chi = \mu_2$, for temperatures below the van der Waals critical temperature, our method can yield $h(y)$ only outside an excluded interval of y values as discussed in Appendix B following Eq. (B5). For small enough γ, the value of y^* given by Eq. (7.3), with $\partial h(y)/\partial \ln y$ determined by truncating our expansion at a finite order in γ, will not fall in the excluded interval. This means that our result predicts a phase transition to any finite order in γ, and therefore cannot be used to decide whether a phase transition occurs. In fact, Kac, Uhlenbeck, and Hemmer[3] show that their one-dimensional model has no phase transition for any $\gamma > 0$, but that the expansion of their exact result in powers of γ exhibits a phase transition to any order in γ. In the context of our work, this can be understood physically because the Kac, Uhlenbeck, and Hemmer model has, for small $\gamma > 0$, even well below the van der Waals critical temperature, instead of a phase transition, a very high compressibility in the region near the van der Waals first-order transition, and therefore has very large density fluctuations in that

region. It is then inappropriate to consider the spatially inhomogeneous part of the random functions as a small perturbation. If, however, a first-order phase transition occurs for $\gamma > 0$, as is expected for the three-dimensional case, and if the compressibility is not abnormally large, then it is physically reasonable to treat the inhomogeneous part of the random field as a perturbation, because under those conditions large density fluctuations occur only in the coexistence region which does not enter into the calculations in the grand canonical formulation.

Our γ expansion is not subject to the objection—raised by some of the previous authors against their results—that the two-particle distribution function does not vanish when the distance between the two particles is less than the hard-core diameter. The two-particle distribution function obtained from our result by functional differentiation with respect to the *repulsive* potential vanishes in the hard-core term by term, when the repulsive potential becomes a potential with hard core. This will be shown in detail in a forthcoming publication.

APPENDIX A

The van der Waals theory with Maxwell construction takes the following form in the grand canonical formalism. In Fig. 5, we have $\rho_h(y) \equiv \tilde{\rho}_h(\mu)$ plotted versus $\mu \equiv \ln y$. To solve (3.12) graphically, we draw on this graph the straight lines $(\mu - \xi)/\beta v_0$. When

$$\max_{(\mu)} \frac{\partial \tilde{\rho}_h(\mu)}{\partial \mu} \leq \frac{1}{\beta v_0}, \tag{A1}$$

there is only one solution μ for any ξ; otherwise there is a range of values ξ for which three solutions occur. The critical temperature $T_c = (k\beta_c)^{-1}$ is thus determined by

$$\max_{(\mu)} \frac{\partial \tilde{\rho}_h(\mu)}{\partial \mu} \equiv \max_{(y)} \frac{\partial \rho_h(y)}{\partial \ln y} = \frac{1}{\beta_c v_0}. \tag{A2}$$

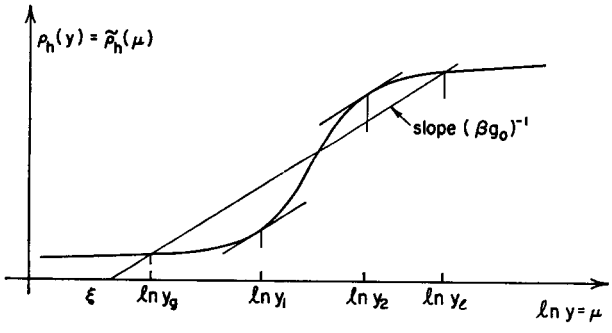

FIG. 5. Maxwell construction (without correction) in the grand canonical formulation.

Taking only the maximum of the integrand in (3.4), we have

$$\beta P(z) = \lim_{V\to\infty} V^{-1} \ln Q(z)$$
$$= \max \left\{ -\frac{c_0^{(0)2}}{2V} + \beta P_h \times \left\{ z \exp\left[-\frac{\beta v(0)}{z} + c_0^{(0)} \left(\frac{\beta v_0}{V}\right)^{\frac{1}{2}} \right] \right\} \right\}. \tag{A3}$$

Since at the extrema

$$\frac{c_0^2}{V} \equiv \frac{(\mu - \xi)^2}{\beta v_0} = \beta v_0 \tilde{\rho}_h^2(\mu), \tag{A4}$$

we have

$$\beta P(z) = \max\left[\beta P_h(y) - \tfrac{1}{2}\beta v_0 \rho_h^2(y)\right], \tag{A5}$$

i.e., the pressure is obtained by taking that solution $\mu^* = \ln y^*$ for which the expression in brackets is largest. One then shows easily that

$$\rho(z) = \frac{\partial \beta P(z)}{\partial \ln z} = \rho_h(y^*). \tag{A6}$$

It is also easily shown that an extremum y_s obtained from (3.12) is a minimum if

$$\left(\frac{\partial \rho_h(y)}{\partial \ln y}\right)_{y_s} > \frac{1}{\beta v_0},$$

so that the intermediate solution in Fig. 5 is never of interest.

The value of z at which the transition occurs can be obtained by application of an equal area construc ion in Fig. 5. The two maxima obtained from solutions y_g and y_l are in fact equal if

$$\int_{\ln y_g}^{\ln y_l} \left[\rho_h(y) - \frac{(\ln y - \ln z e^{-\frac{1}{2}\beta v(0)})}{\beta v_0} \right] d \ln y = 0, \tag{A7}$$

since the integral is equal to

$$\left[\beta P_h(y) - \frac{(\ln y - \ln z e^{-\frac{1}{2}\beta v(0)})}{2\beta v_0} \right]_{y=y_g}^{y=y_l}$$
$$= (\beta P_h(y_l) - \tfrac{1}{2}\beta v_0 \rho_h^2(y_l)) - (\beta P_h(y_g) - \tfrac{1}{2}\beta v_0 \rho_h^2(y_g)). \tag{A8}$$

APPENDIX B

We have seen that we can obtain the order in γ of any diagram if the factors in Eq. (5.4) are finite, even for $V \to \infty$. To establish this, we shall work directly from expression (5.2). First, using (3.27), (3.37), and (3.38), we have

$$\tilde{v}(0) = \frac{1}{V}\sum_\sigma{}' \tilde{v}_\sigma = \frac{1}{V}\sum_\sigma{}' \frac{v_\sigma}{1 - \beta v_\sigma \chi_\sigma} < \frac{\theta^{-1}}{V}\sum_\sigma{}' v_\sigma$$
$$= \theta^{-1}\left(v(0) - \frac{v_0}{V}\right), \tag{B1}$$

where θ is a fixed, positive number. Thus $\tilde{v}(0)$ is finite,

since $v(0)$ is by assumption. Further, since $v(0)/\gamma^\nu$ is nonzero as $\gamma \to 0$, we have

$$\lim_{V\to\infty} \tilde{v}(0) = K\gamma^\nu \tag{B2}$$

with K constant and independent of γ.

We next show that our choice $\chi(\mathbf{r}, \mathbf{r}') = \mu_2(\mathbf{r}, \mathbf{r}' \mid y)$ is permissible. Since $\mu_2(\mathbf{r} \mid y)$ is not everywhere positive, the Fourier components $\mu_{2,\sigma}$ are not necessarily smaller than $\mu_{2,0}$, which is known to be

$$\partial\rho_h(y)/\partial(\ln y) \equiv \dot{\rho}_h(y).$$

Since, however, $\mu_{2,\sigma}$ is of course independent of γ, while v_σ decreases very rapidly in Fourier space for sufficiently small γ, we have

$$|v_\sigma \mu_{2,\sigma}| \leq v_0 \mu_{2,0} \tag{B3}$$

for sufficiently small γ. For the purpose of our estimate, we therefore need only

$$1 - \beta v_0 \mu_{2,0} = 1 - \beta v_0 \dot{\rho}_h(y) \geq \theta. \tag{B4}$$

The critical temperature $T_c = (k\beta_c)^{-1}$ of the van der Waals approximation is given by

$$\beta_c v_0 \max_{(y)} \dot{\rho}_h(y) = 1. \tag{B5}$$

For $T > T_c$ our estimate requires, therefore, only that $T - T_c$ is bounded away from zero. For $T < T_c$, however, an interval of values of y has to be excluded, namely, the interval where

$$\dot{\rho}_h(y) > (1/\beta v_0)(1 - \theta). \tag{B6}$$

Our estimate therefore is valid for $T < T_c$ only if $y \leq y_1 - \epsilon$ or $y \geq y_2 + \epsilon$ with y_1 and y_2 indicated in Fig. 5, and ϵ positive.

Next, we have to show that the factor

$$q \equiv V^{-1} \int_V d^\nu r_1 \cdots d^\nu r_S \prod_{1\leq i<j\leq S} (\tilde{v}(\mathbf{r}_i - \mathbf{r}_j))^{t_{ij}} \tag{B7}$$

is bounded. To do this, we first draw a diagram representing (B7), which consists of points representing $\mathbf{r}_1, \cdots, \mathbf{r}_S$ and lines representing $\tilde{v}(\mathbf{r}_i - \mathbf{r}_j)$. Select a point from this diagram which is not an articulation point. Call this point $\mathbf{r}_1$. Let the points connected to $\mathbf{r}_1$ by $\tilde{v}$ bonds be labeled $\mathbf{r}_{1,a}, \mathbf{r}_{1,b}, \cdots, \mathbf{r}_{1,k}$. Then

$$\begin{aligned}
|q| &\leq V^{-1} \int_V d^\nu r_1 \cdots d^\nu r_S \prod_{1\leq i<j\leq S} |\tilde{v}(\mathbf{r}_i - \mathbf{r}_j)|^{t_{ij}} \\
&= V^{-1} \int_V d^\nu r_2 \cdots d^\nu r_S \prod_{2\leq i<j\leq S} |\tilde{v}(\mathbf{r}_i - \mathbf{r}_j)|^{t_{ij}} \\
&\quad \times \int_V d^\nu r_1 \, |\tilde{v}(\mathbf{r}_1 - \mathbf{r}_{1,a})|^{t_{1,1_a}} \cdots |\tilde{v}(\mathbf{r}_1 - \mathbf{r}_{1,k})|^{t_{1,1_k}} \\
&\leq V^{-1} \int_V d^\nu r_2 \cdots d^\nu r_S \prod_{2\leq i<j\leq S} |\tilde{v}(\mathbf{r}_i - \mathbf{r}_j)|^{t_{ij}} \\
&\quad \times \max_{(r_{1,a},\cdots,r_{1,k})} \int_V d^\nu r_1 \, |\tilde{v}(\mathbf{r}_1 - \mathbf{r}_{1,a})|^{t_{1,1_a}} \cdots \\
&\qquad\qquad |\tilde{v}(\mathbf{r}_1 - \mathbf{r}_{1,k})|^{t_{1,1_k}}.
\end{aligned} \tag{B8}$$

The last factor in (B8) has a finite upper bound which we can establish as follows. Clearly,

$$\begin{aligned}
&\max_{(r_{1,a},\cdots,r_{1,k})} \int_V d^\nu r_1 \, |\tilde{v}(\mathbf{r}_1 - \mathbf{r}_{1,a})|^{t_{1,1_a}} \cdots |\tilde{v}(\mathbf{r}_1 - \mathbf{r}_{1,k})|^{t_{1,1_k}} \\
&\quad < \left[\max_{(\mathbf{r})} |\tilde{v}(\mathbf{r})|\right]^{t_{1,1_a}+\cdots+t_{1,1_k}-1} \int_V d^\nu r \, |\tilde{v}(\mathbf{r})|.
\end{aligned} \tag{B9}$$

From (3.40) and (3.41), with

$$\chi(\mathbf{r}) = \mu_2(\mathbf{r} \mid y) \cong \frac{\partial\rho_h(y)}{\partial \ln y} \delta(\mathbf{r}) = \dot{\rho}_h(y)\delta(\mathbf{r}), \tag{B10}$$

we obtain

$$\begin{aligned}
\tilde{v}_1(\mathbf{r} - \mathbf{r}') &= v(\mathbf{r} - \mathbf{r}') \\
&\quad + \beta\dot{\rho}_h \int_V \tilde{v}_1(\mathbf{r} - \mathbf{r}_1) v(\mathbf{r}_1 - \mathbf{r}') \, d^\nu r_1 .
\end{aligned} \tag{B11}$$

Iterating this equation, we find

$$\tilde{v}_1(\mathbf{r} - \mathbf{r}') = v(\mathbf{r} - \mathbf{r}') + \sum_1^\infty (\beta\dot{\rho}_h)^n v^{(n)}(\mathbf{r} - \mathbf{r}') \tag{B12}$$

with

$$\begin{aligned}
v^{(n)}(\mathbf{r} - \mathbf{r}') &= \int_V d^\nu r_1 \cdots d^\nu r_n \\
&\quad \times v(\mathbf{r} - \mathbf{r}_1) v(\mathbf{r}_1 - \mathbf{r}_2) \cdots v(\mathbf{r}_n - \mathbf{r}').
\end{aligned} \tag{B13}$$

Now using

$$\begin{aligned}
v^{(n)}(\mathbf{r} - \mathbf{r}') &= \int_V d^\nu r_1 v(\mathbf{r} - \mathbf{r}_1) v^{(n-1)}(\mathbf{r}_1 - \mathbf{r}') \\
&\leq \left[\max_{(\mathbf{r})} v^{(n-1)}(\mathbf{r})\right] \int_V d^\nu r_1 v(\mathbf{r} - \mathbf{r}_1) \\
&= \left[\max_{(\mathbf{r})} v^{(n-1)}(\mathbf{r})\right] v_0
\end{aligned} \tag{B14}$$

and

$$\begin{aligned}
v^{(1)}(\mathbf{r} - \mathbf{r}') &= \int_V d^\nu r_1 v(\mathbf{r} - \mathbf{r}_1) v(\mathbf{r}_1 - \mathbf{r}') \\
&\leq \left[\max_{(\mathbf{r})} v(\mathbf{r})\right] \int_V v(\mathbf{r} - \mathbf{r}_1) \, d^\nu r_1 \\
&= \left[\max_{(\mathbf{r})} v(\mathbf{r})\right] v_0,
\end{aligned} \tag{B15}$$

we obtain by induction

$$v^{(n)}(\mathbf{r}) \leq v_0^n \left[\max_{(\mathbf{r})} v(\mathbf{r})\right]. \tag{B16}$$

Then, from (B12),

$$\begin{aligned}
\tilde{v}_1(\mathbf{r}) &\leq v(\mathbf{r}) + \left[\max_{(\mathbf{r})} v(\mathbf{r})\right] \sum_1^\infty (\beta v_0 \dot{\rho}_h)^n \\
&\leq \left[\max_{(\mathbf{r})} v(\mathbf{r})\right] (1 - \beta v_0 \dot{\rho}_h)^{-1}.
\end{aligned} \tag{B17}$$

Next, using

$$\tilde{v}_1(\mathbf{r}) \equiv \tilde{v}(\mathbf{r}) + \tilde{v}_0/V, \tag{B18}$$

we find

$$|\tilde{v}(\mathbf{r})| = \left| \tilde{v}_1(\mathbf{r}) - \frac{\tilde{v}_0}{V} \right| < \tilde{v}_1(\mathbf{r}) + \frac{\tilde{v}_0}{V}$$
$$\leq \left[\max_{(\mathbf{r})} v(\mathbf{r}) + \frac{v_0}{V} \right] (1 - \beta v_0 \dot{\rho}_h)^{-1}. \quad \text{(B19)}$$

We also have

$$\int_V |\tilde{v}(\mathbf{r})|\, d^\nu r \leq \int_V \left[\tilde{v}_1(\mathbf{r}) + \frac{\tilde{v}_0}{V} \right] d^\nu r$$
$$= 2\tilde{v}_0 = 2v_0(1 - \beta v_0 \dot{\rho}_h)^{-1}. \quad \text{(B20)}$$

Inserting (B19) and (B20) into (B9), we obtain

$$\max_{(\mathbf{r}_{1,a}, \cdots, \mathbf{r}_{1,k})} \int_V d^\nu r_1 \, |\tilde{v}(\mathbf{r}_1 - \mathbf{r}_{1,a})|^{t_{1,1_a}} \cdots |\dot{v}(\mathbf{r}_1 - \mathbf{r}_{1,k})|^{t_{1,1_k}}$$
$$\leq \left[\left(\max_{(\mathbf{r})} v(\mathbf{r}) + \frac{v_0}{V} \right) (1 - \beta v_0 \dot{\rho}_h)^{-1} \right]^{t_{1,1_a} + \cdots + t_{1,1_k} - 1}$$
$$\times 2v_0(1 - \beta v_0 \dot{\rho}_h)^{-1}. \quad \text{(B21)}$$

We can now use the bound (B21) in (B8). We continue the process by choosing from among the points $\mathbf{r}_2, \cdots, \mathbf{r}_S$ another point which is not an articulation point of the diagram containing $\mathbf{r}_2, \cdots, \mathbf{r}_S$. In this manner, we successively reduce the number of integrations in (B8) until we have integrated over all points but one. This last integral then gives a factor V. Our final result is then

$$|q| \leq \left(\frac{\max_{(\mathbf{r})} v(\mathbf{r}) + \frac{v_0}{V}}{1 - \beta v_0 \dot{\rho}_h} \right)^{\mathcal{S}} \left(\frac{2v_0}{1 - \beta v_0 \dot{\rho}_h} \right)^{\mathcal{S}-1}, \quad \text{(B22)}$$

where $\mathcal{S}$ here is defined as

$$\mathcal{S} \equiv \sum_{i<j} t_{ij} - S + 1,$$

which shows that $|q|$ is bounded provided $1 - \beta v_0 \dot{\rho}_h > 0$, i.e., provided that we are not at the critical point of the van der Waals approximation.

Taking the limit $V \to \infty$ of (B22) and setting

$$\max_{(\mathbf{r})} v(\mathbf{r}) = K' \gamma^\nu,$$

with K' a finite number independent of γ, we obtain

$$|q| < K'' \gamma^{\nu \mathcal{S}}, \quad \text{(B23)}$$

with $\mathcal{S}$ defined as above. Combining this with the bound (B2) on $\tilde{v}(0)$, we have then that the expression (5.2) is bounded by the quantity

$$K''' \gamma^{\nu \mathcal{S}} = K''' \gamma^{\nu(B-S+1)}, \quad \text{(B24)}$$

with $\mathcal{S}$ as defined after Eq. (B22).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Exact Quantization Conditions. II

J. B. KRIEGER
Physics Department, Polytechnic Institute of Brooklyn, Brooklyn, New York

(Received 31 January 1969)

By employing coordinate transformations, a generalized WKB quantization condition is derived which includes a modified WKB quantization rule and the related higher-order integrals. It is observed that a necessary condition for the first-order integral to give an exact quantization rule is the vanishing of the higher-order integrals. When this condition is satisfied, the resulting quantization condition is of the form of all previously known exact-quantization rules. The higher-order integrals are shown to vanish for certain cases of interest. An error in Paper I [C. Rosenzweig and J. B. Krieger, J. Math. Phys. **9**, 849 (1968)] is noted and an examination of the higher-order integrals shows that a proposed quantization rule given there is not exact.

I. INTRODUCTION

In a previous work[1] (hereafter referred to as I) the method of Froman and Froman[2] was employed to prove rigorously the exactness of WKB or modified WKB quantization conditions for all potentials for which such rules were known to be valid.[3] The advantage of the method of Froman and Froman lies in the fact that exact quantization rules are proved without recourse to a comparison with exact analytic solutions of the Schrödinger equation. Thus this method has the potential of providing exact quantization conditions even for potentials for which no analytic solutions are known to exist. The disadvantage of this method lies in the fact that it requires a detailed examination of the potential in the entire complex plane and, furthermore, it does not provide a means to determine whether a proposed quantization rule is not exact, i.e., either one can prove exactness by some clever choice of contour, or else no definite conclusion can be drawn about the correctness of the proposed quantization condition.

A second method to determine exactness of a quantization condition proceeds by showing that all additional higher-order correction terms to the WKB integral vanish for the given potential.[4,5] This condition is not rigorously sufficient, however, because it is well known that the WKB approximation is valid only asymptotically as $\hbar \to 0$ so that it is still possible to have correction terms of say $O(e^{-|a|/\hbar})$, which would have an asymptotic expansion consisting entirely of zeros. The condition is necessary, however, because if the correction term of $O(\hbar^n)$ is not zero, then the terms having an asymptotic expansion entirely of zeros cannot cancel it. It should also be noted that, although the condition that the higher-order WKB integrals vanish is only a necessary condition for exactness of the WKB quantization condition, there is no known potential for which both the second- and third-order integrals vanish for which the first-order WKB integral does not provide an exact quantization rule. In I, however, it is pointed out that there are known cases for which the WKB integral does not give rise to the exact eigenvalues (the second- and third-order terms are not both zero here), but a modified WKB integral does. In these cases it is not clear what "correction" terms we must show to be actually zero. In this note we report a method which employs a generalized Langer transformation[6] and which gives a necessary condition for the validity of modified WKB quantization rules. We also point out an error in the proof of Case VIII of I and demonstrate that the quantization rule proposed for that case is incorrect.

II. MODIFIED WKB QUANTIZATION CONDITIONS

If E is the energy eigenvalue of an equation of the form

$$-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}X(x) - q^2(x, E)X(x) = 0, \qquad (1)$$

where $X(x) \to 0$ for $x \to \pm\infty$ and X is single-valued and finite, then a generalization of Dunham's[4] derivation yields the result that the WKB quantization condition, through the third-order integral, may be written as[7]

$$\oint q\,dx - \frac{\hbar^2}{64m}\oint\left[\frac{d}{dx}(q^2)\right]^2 q^{-5}\,dx - \frac{\hbar^4}{8192m^2} \times \oint\left\{49\left[\frac{d}{dx}(q^2)\right]^4 - 16\left[\frac{d}{dx}(q^2)\right]\left[\frac{d^3}{dx^3}(q^2)\right]q^{-7}\right\}dx = \frac{(N+\tfrac{1}{2})h}{(2m)^{\frac{1}{2}}}, \qquad (2)$$

[1] C. Rosenzweig and J. B. Krieger, J. Math. Phys. **9**, 849 (1968).
[2] N. Froman and P. O. Froman, *JWKB Approximation; Contribution to the Theory* (North-Holland Publ. Co., Amsterdam, 1965).
[3] P. B. Bailey, J. Math. Phys. **5**, 1293 (1964).
[4] J. L. Dunham, Phys. Rev. **41**, 713 (1932).
[5] P. N. Argyres, Physics **2**, 131 (1965).
[6] R. E. Langer, Phys. Rev. **51**, 669 (1937).
[7] J. B. Krieger and C. Rosenzweig, Phys. Rev. **164**, 171 (1967).

where the integrals are taken about a contour that includes the two real zeros of q and no other singularities. This quantization condition cannot always be directly applied to problems of interest because the boundary condition on X may be different from those required by Dunham's derivation, i.e., in radial problems the radial wavefunction $U(r)$, where $\psi(r) = [U(r)/r]Y_{lm}(\theta, \phi)$, satisfies the boundary conditions $U \to 0$ for $r \to 0$ and $r \to \infty$. One can also consider cases for which the wavefunction vanishes at two points for which x is finite.[1,3] Furthermore, even in those cases where the wavefunction vanishes at $\pm\infty$, Eq. (2) may not be the most convenient form from which to prove exact quantization rules because the higher-order integrals may not vanish.

We consider the following transformation:

$$z = z(x),$$

with a singled-valued inverse,

$$x = x(z),$$

such that, if Eq. (1) is defined in an interval $a \leq x \leq b$ with $X(a) = X(b) = 0$, then $-\infty \leq z \leq \infty$.

Defining

$$X(x) = \phi(z),$$

we immediately obtain the result that ϕ satisfies

$$-\frac{\hbar^2}{2m}\left\{\frac{d^2\phi}{dz^2} + \frac{d\phi}{dz}\left[\frac{d^2z/dx^2}{(dz/dx)^2}\right]\right\} + \frac{(V[x(z)] - E)\phi}{(dz/dx)^2} = 0 \tag{3}$$

with the boundary conditions

$$\phi(z) \to 0, \quad z \to \pm\infty.$$

Equation (3) is not of the same form as Eq. (1), so that it is not yet possible to employ Eq. (2) as the quantization condition in the new coordinate system. However, if

$$\frac{d^2\phi(z)}{dz^2} + p(z)\frac{d\phi}{dz} + r(z)\phi(z) = 0,$$

then, defining

$$\phi(z) = f(z)\exp -\tfrac{1}{2}\int_{z_0}^{z} p(z')\,dz',$$

we find

$$\frac{d^2f(z)}{dz^2} + \left[r(z) - \tfrac{1}{4}p^2(z) - \frac{1}{2}\frac{dp}{dz}\right]f(z) = 0,$$

so that Eq. (3) may be transformed into an equation of the form of Eq. (1). The result is

$$-\frac{\hbar^2}{2m}\frac{d^2f}{dz^2} - q^2(z, E)f = 0,$$

where

$$q^2(z, E) = \frac{(E - V)}{z'^2} - \frac{\hbar^2}{2m}\left[\frac{1}{4}\left(\frac{z''}{z'^2}\right)^2 + \frac{1}{2}\frac{d}{dz}\left(\frac{z''}{z'^2}\right)\right]. \tag{4}$$

Substituting Eq. (4) into Eq. (2) (where the integration is now over z) and transforming back to x as the integration variable, we obtain

$$\begin{aligned}
&\oint [E - V(x) - S(x)z'^2]^{\frac{1}{2}}\,dx \\
&\quad - \frac{\hbar^2}{64m}\oint \left\{\frac{d}{dx}\left[\frac{E - V(x)}{z'^2} - S(x)\right]\right\}^2 z'^4 \\
&\quad \times \{E - V(x) - S(x)z'^2\}^{-\frac{5}{2}}\,dx \\
&\quad - \frac{\hbar^4}{8192m^2}\oint \left\{49\left\{\frac{d}{dx}\left[\frac{E - V}{z'^2} - S(x)\right]\right\}^4\right. \\
&\quad \times \{E - V - S(x)z'^2\}^{-\frac{11}{2}} \\
&\quad - 16\frac{1}{z'}\left\{\frac{d}{dx}\left[\frac{E - V}{z'^2} - S(x)\right]\right. \\
&\quad \left.\left. \times \left[\frac{1}{z'}\left(\frac{d}{dx}\right)\right]^3\left[\frac{E - V}{z'^2} - S(x)\right]\right\} z'^8\right. \\
&\quad \times \{E - V - S(x)z'^2\}^{-\frac{7}{2}}\,dx \\
&\qquad = (N + \tfrac{1}{2})h/(2m)^{\frac{1}{2}},
\end{aligned} \tag{5}$$

where

$$S(x) \equiv \frac{\hbar^2}{2m}\left\{\frac{1}{4}\left[\frac{d}{dx}\left(\frac{1}{z'}\right)\right]^2 - \frac{1}{2z'}\frac{d^2}{dx^2}\left(\frac{1}{z'}\right)\right\}. \tag{6}$$

We note from Eqs. (5) and (6) that if the transformation has made the new second- and third-order integral zero, then the quantization condition becomes

$$\oint \left[\frac{2m(E - V)}{\hbar^2} - \frac{2m}{\hbar^2}Sz'^2\right]^{\frac{1}{2}} dx = (N + \tfrac{1}{2})\pi,$$

where Sz'^2 is a function of x but not of E, which is precisely the form of all known exact quantization conditions given in I. Furthermore, we observe that even if the transformation has not made the second- and third-order integrals identically zero, it will often be possible to reduce their value, leaving the first-order integral as a better approximation to the energy than it would have been if the transformation had not been employed. Finally, although the second- and third-order integrals appear incredibly complicated, we show below that by using the technique of contour integration, it is often easy to demonstrate that they are identically zero.

III. POTENTIALS FOR WHICH UNMODIFIED HIGHER-ORDER INTEGRALS ARE ZERO

The references below to Case I, etc., refer to the case numbers in I. Since the unmodified WKB quantization condition is employed here, $z = x$ and all potentials are defined for $-\infty < x < \infty$.

Case I: $V(x) = \frac{1}{2}kx^2$ (*simple harmonic oscillator*). The higher-order integrals have previously been shown to be zero.[8]

Case VII: $V(x) = Ae^{-2ax} - Be^{-ax}$, $A, B > 0$ (*modified Morse oscillator*). The second-order integral is proportional to

$$\oint \frac{V'^2}{[E-V]^{\frac{5}{2}}}\,dx = \oint \frac{(2aAe^{-2ax} - aBe^{-ax})^2}{[E - Ae^{-2ax} + Be^{-ax}]^{\frac{5}{2}}}\,dx.$$

Making the change of variables

$$e^{-ax} = y,$$

we obtain

$$\oint \frac{V'^2}{[E-V]^{\frac{5}{2}}}\,dx = \oint \frac{(2aAy^2 - aBy)^2}{[E - Ay^2 + By]^{\frac{5}{2}}}\,\frac{dy}{y}.$$

As $y \to 0$, the integrand $\to y$ and thus the only singularities of the integrand occur at the zeros of the denominator which are the two classical turning points. Deforming the contour to a large circle with center at the origin, we see that as $y \to \infty$ the integrand $\to 1/y^2$, and hence evaluating the integral along a circle of radius R gives zero as $R \to \infty$.

Similarly, making the same change of variables in the third-order integral, we find that as $y \to 0$ the integrand $\to y^3$ and hence the only singularities are at the two classical turning points. Furthermore, as $y \to \infty$, the integrand $\to 1/y^4$, and hence evaluating the integral along a circle of radius R again gives zero as $R \to \infty$.

Case VIII: $V(x) = Ae^{2ax} + Be^{-2ax}$. The proof given in I that the unmodified WKB quantization condition is exact for this case is in error. The argument presented there is incorrect because, unlike Case VII, the quantity F_{22} is not real here, since q is real on the contour instead of imaginary as asserted there. Furthermore, the fact that the eigenvalue for the ground state for $A = B$ is the same to three significant figures as the eigenvalues calculated by the quantization rule[1] is not a convincing argument of the exactness of the quantization condition because for a particle in the ground state the potential between the classical turning points is essentially parabolic and hence corresponds to a simple harmonic oscillator for which this quantization rule is known to apply.[9] In fact, expanding the potential about its minimum and keeping terms to $O(x^2)$, we find, by applying the rule in Case I, that the eigenvalue is the same as that given by the result of numerical integration for the exact potential to three significant figures. Thus, from the above considerations it is not clear whether the quantization rule is incorrect or a rigorous proof, different from that previously given, can be constructed to justify it.

However, by investigating the higher-order WKB integrals we have been able to show that the second- and third-order integrals are not zero and thus the proposed quantization rule is not exact.

IV. POTENTIALS FOR WHICH MODIFIED HIGHER-ORDER INTEGRALS ARE ZERO

Case II:

$$V = -\frac{V_0}{r} + \frac{b\hbar^2}{2mr^2} + \frac{l(l+1)\hbar^2}{2mr^2}, \quad 0 < r < \infty.$$

Case III:

$$V = \tfrac{1}{2}kr^2 + \frac{b\hbar^2}{2mr^2} + \frac{l(l+1)\hbar^2}{2mr^2}, \quad 0 < r < \infty.$$

Cases II and III correspond to the effective radial potential when the Schrödinger equation is separated in radial coordinates for the case of the coulomb potential plus degeneracy-breaking term and the harmonic oscillator with degeneracy-breaking term, respectively. The wavefunction $U(r)$, which is the eigenfunction of the effective radial equation, is related to the radial part of the wavefunction $R(r)$ by

$$R(r) = U(r)/r,$$

and thus $U(r) \to 0$ as $r \to 0$ at least as fast as r. If we let[6] $r = e^z$, then $r = 0$ and $r = \infty$ correspond to $z = -\infty$ and $z = +\infty$, respectively, and

$$S(r)z'^2 = \frac{\hbar^2}{2m}\frac{1}{4r^2}.$$

Hence, the quantization rule becomes

$$\oint \left[\frac{2m}{\hbar^2}(E-V) - \frac{1}{4r^2}\right]^{\frac{1}{2}} dr = (N + \tfrac{1}{2})\pi,$$

provided that the second- and third-order integrals are zero. The proofs that these higher-order integrals are zero for Case II and Case III have been given previously[7] for the case $b = 0$. Since letting $b \neq 0$ is equivalent to changing the value of l, the same arguments as in the case $b = 0$ are valid here and will not be repeated.

Finally, we observe from Eqs. (5) and (6) that it is trivial to test whether or not a given transformation suffices to make the transformed higher-order integrals

[8] J. B. Krieger, M. L. Lewis, and C. Rosenzweig, J. Chem. Phys. 47, 2942 (1967).

[9] The above observations are due to Professor G. Wannier (private communication).

zero. However, given a conjectured exact quantization rule, i.e., given $S(x)z'^2$, it is not a simple matter to determine precisely what are the higher-order integrals, since this requires knowing $S(x)$ and z'^2 separately. In fact, given a choice for $S(x)z'^2$, we may obtain z' by solving a second-order nonlinear differential equation obtained by multiplying Eq. (6) by z'^2. We have not yet been successful in finding the appropriate z' for Cases IV, V, and VI, and thus cannot explicitly demonstrate that the higher-order integrals are zero in these cases also.

V. DISCUSSION

Using the method of coordinate transformations, we have derived a generalization of the usual WKB quantization condition through the third-order integral. Since a necessary condition for the first-order integral to give an exact quantization condition is the vanishing of the higher-order integrals, this method leads to the possibility of finding exact quantization conditions in those cases where the usual first-order integral in the original coordinate system is not sufficient. Furthermore, even in those cases where the transformed higher-order integrals are not zero, the technique is still useful in providing a means of reducing the size of these terms and hence increasing the accuracy of eigenvalues computed from the first-order integral alone.

ACKNOWLEDGMENTS

We are indebted to Professor G. Wannier for his stimulating correspondence. We would also like to express our gratitude to C. Rosenzweig and J. Bekenstein for several helpful discussions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

An Exact Quantum Theory of the Time-Dependent Harmonic Oscillator and of a Charged Particle in a Time-Dependent Electromagnetic Field*

H. R. LEWIS, JR., AND W. B. RIESENFELD
Los Alamos Scientific Laboratory, University of California, Los Alamos, New Mexico

(Received 8 July 1968)

The theory of explicitly time-dependent invariants is developed for quantum systems whose Hamiltonians are explicitly time dependent. The central feature of the discussion is the derivation of a simple relation between eigenstates of such an invariant and solutions of the Schrödinger equation. As a specific well-posed application of the general theory, the case of a general Hamiltonian which settles into constant operators in the sufficiently remote past and future is treated and, in particular, the transition amplitude connecting any initial state in the remote past to any final state in the remote future is calculated in terms of eigenstates of the invariant. Two special physical systems are treated in detail: an arbitrarily time-dependent harmonic oscillator and a charged particle moving in the classical, axially symmetric electromagnetic field consisting of an arbitrarily time-dependent, uniform magnetic field, the associated induced electric field, and the electric field due to an arbitrarily time-dependent uniform charge distribution. A class of explicitly time-dependent invariants is derived for both of these systems, and the eigenvalues and eigenstates of the invariants are calculated explicitly by operator methods. The explicit connection between these eigenstates and solutions of the Schrödinger equation is also calculated. The results for the oscillator are used to obtain explicit formulas for the transition amplitude. The usual sudden and adiabatic approximations are deduced as limiting cases of the exact formulas.

I. INTRODUCTION

The use of explicitly time-dependent invariants in applications of quantum theory has received little attention, if any. Presumably, the reason for this lack of attention has been the dearth of examples in which the use of such quantities was both possible and fruitful. Recently, a class of exact invariants for time-dependent harmonic oscillators, both classical and quantum, was reported.[1] The simplicity of the rules for constructing these invariants and the instructive relation of the invariants to the asymptotic expansion of adiabatic invariant theory have stimulated an interest in using the invariants for solving some explicit quantum-mechanical problems. We discuss

* Work performed under the auspices of the U.S. Atomic Energy Commission.

[1] H. R. Lewis, Jr., J. Math. Phys. **9**, 1976 (1968); also, H. R. Lewis, Jr., Phys. Rev. Letters **18**, 510, 636 (1967).

zero. However, given a conjectured exact quantization rule, i.e., given $S(x)z'^2$, it is not a simple matter to determine precisely what are the higher-order integrals, since this requires knowing $S(x)$ and z'^2 separately. In fact, given a choice for $S(x)z'^2$, we may obtain z' by solving a second-order nonlinear differential equation obtained by multiplying Eq. (6) by z'^2. We have not yet been successful in finding the appropriate z' for Cases IV, V, and VI, and thus cannot explicitly demonstrate that the higher-order integrals are zero in these cases also.

V. DISCUSSION

Using the method of coordinate transformations, we have derived a generalization of the usual WKB quantization condition through the third-order integral. Since a necessary condition for the first-order integral to give an exact quantization condition is the vanishing of the higher-order integrals, this method leads to the possibility of finding exact quantization conditions in those cases where the usual first-order integral in the original coordinate system is not sufficient. Furthermore, even in those cases where the transformed higher-order integrals are not zero, the technique is still useful in providing a means of reducing the size of these terms and hence increasing the accuracy of eigenvalues computed from the first-order integral alone.

ACKNOWLEDGMENTS

We are indebted to Professor G. Wannier for his stimulating correspondence. We would also like to express our gratitude to C. Rosenzweig and J. Bekenstein for several helpful discussions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

An Exact Quantum Theory of the Time-Dependent Harmonic Oscillator and of a Charged Particle in a Time-Dependent Electromagnetic Field*

H. R. LEWIS, JR., AND W. B. RIESENFELD
Los Alamos Scientific Laboratory, University of California, Los Alamos, New Mexico

(Received 8 July 1968)

The theory of explicitly time-dependent invariants is developed for quantum systems whose Hamiltonians are explicitly time dependent. The central feature of the discussion is the derivation of a simple relation between eigenstates of such an invariant and solutions of the Schrödinger equation. As a specific well-posed application of the general theory, the case of a general Hamiltonian which settles into constant operators in the sufficiently remote past and future is treated and, in particular, the transition amplitude connecting any initial state in the remote past to any final state in the remote future is calculated in terms of eigenstates of the invariant. Two special physical systems are treated in detail: an arbitrarily time-dependent harmonic oscillator and a charged particle moving in the classical, axially symmetric electromagnetic field consisting of an arbitrarily time-dependent, uniform magnetic field, the associated induced electric field, and the electric field due to an arbitrarily time-dependent uniform charge distribution. A class of explicitly time-dependent invariants is derived for both of these systems, and the eigenvalues and eigenstates of the invariants are calculated explicitly by operator methods. The explicit connection between these eigenstates and solutions of the Schrödinger equation is also calculated. The results for the oscillator are used to obtain explicit formulas for the transition amplitude. The usual sudden and adiabatic approximations are deduced as limiting cases of the exact formulas.

I. INTRODUCTION

The use of explicitly time-dependent invariants in applications of quantum theory has received little attention, if any. Presumably, the reason for this lack of attention has been the dearth of examples in which the use of such quantities was both possible and fruitful. Recently, a class of exact invariants for time-dependent harmonic oscillators, both classical and quantum, was reported.[1] The simplicity of the rules for constructing these invariants and the instructive relation of the invariants to the asymptotic expansion of adiabatic invariant theory have stimulated an interest in using the invariants for solving some explicit quantum-mechanical problems. We discuss

* Work performed under the auspices of the U.S. Atomic Energy Commission.

[1] H. R. Lewis, Jr., J. Math. Phys. **9**, 1976 (1968); also, H. R. Lewis, Jr., Phys. Rev. Letters **18**, 510, 636 (1967).

two systems in detail: the time-dependent harmonic oscillator and a charged particle in a particular type of time-dependent, classical electromagnetic field.

In Sec. II we consider the theory of explicitly time-dependent invariants for a general quantum system whose Hamiltonian operator $H(t)$ is explicitly time-dependent. Of course, such a system is not closed, in the sense that some external influence, which need not be specified, may change the parameters of the system, alter its total energy or angular momentum, etc. The semiclassical theory of radiation provides a well-known example. In that case the quantum system is taken to be an atom or molecule which undergoes radiative transitions, and the explicitly time-dependent term in the Hamiltonian operator is the interaction with the classical radiation field. The usual approximation techniques for treating such a system are time-dependent perturbation theory (in which the time-dependent term is considered small), the adiabatic approximation (in which the time scale of variation of the time-dependent term is long compared to all of the characteristic periods of the system), and the "sudden" approximation (in which the external changes are fast compared to the shortest characteristic period). The results of the adiabatic and "sudden" approximations will be deduced as limiting cases of rigorous results that are presented in this article for the time-dependent harmonic oscillator.

The central feature of our discussion of general systems is the derivation of the relation between eigenstates of an explicitly time-dependent invariant and solutions of the Schrödinger equation. A time-dependent phase transformation can be found for each eigenstate of an invariant such that the eigenstate becomes a solution of the Schrödinger equation, and the phase is determined by solving a simple, first-order differential equation. Later in the article, for the two special systems that we discuss in detail, we derive explicit formulas for the eigenstates and eigenvalues of the invariants and for the phases. Also, in these examples, we evaluate all physically relevant matrix elements elegantly by operator techniques.

To provide a specific well-posed application of these ideas, we consider a Hamiltonian which settles into constant operators in the sufficiently remote past and future, and we assume that each of these two limiting operators has a known complete set of eigenstates and eigenvalues. The time dependence of $H(t)$ for intermediate times is to be at least piecewise continuous, but otherwise arbitrary, and we calculate the transition amplitude connecting any initial state in the remote past to any final state in the remote future.

The first special physical system to which we apply the general results, in Sec. III, is that of a time-dependent harmonic oscillator, that is, a system whose Hamiltonian has the form of the Hamiltonian of a simple harmonic oscillator, but for which the frequency parameter is allowed to vary with time.[2] To begin with, we derive a class of exact invariants for this system by means of a method different than that used previously.[1] Then we calculate the eigenvalues and eigenstates of these invariants, and we also calculate the appropriate time-dependent phase factors that make the eigenstates solutions of the Schrödinger equation. Finally, as in Sec. II, we specialize to the case that the Hamiltonian is a constant operator in the remote past and future and calculate explicit formulas for the transition amplitude between arbitrary states at these times. Using these exact formulas, we discuss the adiabatic and "sudden" approximations and deduce the usual formulas for those limiting cases.

In Sec. IV we consider a charged particle in the classical, axially symmetric electromagnetic field consisting of an arbitrarily time-dependent, uniform magnetic field, the associated induced electric field, and the electric field due to an arbitrarily time-dependent, uniform charge distribution. The dynamical variables of this system are simply related to those of the time-dependent harmonic oscillator by a noncanonical transformation. We use this noncanonical transformation to derive a class of invariants for the particle system from the invariants for the oscillator. These invariants for the particle system are not Hermitian. However, it turns out to be possible to derive from them a class of Hermitian invariants that are formally identical to the Hamiltonian for a particle in a uniform, time-independent magnetic field. Using operator techniques, we derive the eigenstates and eigenvalues of this class of Hermitian invariants, and we find the phases for which the eigenstates are solutions of the Schrödinger equation. The results are a generalization of the solution[3,4,5] for a particle in a uniform, time-independent magnetic field.

---

[2] The two special systems that we consider in Secs. III and IV have been treated along different lines by M. Kolsrud: (a) "Exact Quantum Dynamical Solutions for Oscillator-Like Systems," Institute for Theoretical Physics, University of Oslo (Norway), Institute Report No. 28 (1965); (b) Kgl. Norske Videnskab. Selskabs Forh. **31**, No. 5 (1958); (c) Phys. Rev. **104**, 1186 (1956).

[3] L. Landau, Z. Physik **64**, 629 (1930).

[4] R. B. Dingle, Proc. Roy. Soc. (London) **A211**, 500 (1952).

[5] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-relativistic Theory* (Addison-Wesley Publ. Co., Inc., Reading, Mass., 1965), 2nd. ed., p. 426. There is an error in this derivation: the wavefunction is assumed proportional to $e^{im\phi}$, but the subsequent formulas are derived for $e^{-im\phi}$.

II. EXPLICITLY TIME-DEPENDENT INVARIANTS AND THEIR RELATION TO SOLUTIONS OF THE SCHRÖDINGER EQUATION

We consider a system whose Hamiltonian operator $H(t)$ is an explicit function of time, and we assume the existence of another explicitly time-dependent nontrivial Hermitian operator $I(t)$, which is an invariant. That is, $I(t)$ satisfies the conditions

$$\frac{dI}{dt} \equiv \frac{\partial I}{\partial t} + \frac{1}{i\hbar}[I, H] = 0 \tag{1}$$

and

$$I^\dagger = I. \tag{2}$$

The equation determining the time-dependent Schrödinger state vector $|\ \rangle$ is

$$i\hbar \frac{\partial}{\partial t}|\ \rangle = H(t)|\ \rangle. \tag{3}$$

By operating with the left-hand side of Eq. (1) on $|\ \rangle$ and using Eq. (3), we obtain the relation

$$i\hbar \frac{\partial}{\partial t}(I|\ \rangle) = H(I|\ \rangle), \tag{4}$$

which implies that the action of the invariant operator on a Schrödinger state vector produces another solution of the Schrödinger equation. This result is valid for any invariant, even if the latter involves the operation of time differentiation. If the invariant does *not* involve time differentiation, then we are able to derive a simple and explicit rule for choosing the phases of the eigenstates of $I(t)$ such that these states themselves satisfy the Schrödinger equation. In what follows, we assume that $I(t)$ does not involve time differentiation. The invariants with which we treat the time-dependent harmonic oscillator, described in Sec. III, and the motion of a charged particle, described in Sec. IV, satisfy this requirement.

We assume that the invariant operator is one of a complete set of commuting observables, so that there is a complete set of eigenstates of I. We denote the eigenvalues of I by λ, and the orthonormal eigenstates associated with a given λ by $|\lambda, \kappa\rangle$, where the label κ represents all of the quantum numbers other than λ that are necessary for specifying the eigenstates:

$$I(t)|\lambda, \kappa\rangle = \lambda|\lambda, \kappa\rangle, \tag{5a}$$

$$\langle\lambda', \kappa'|\lambda, \kappa\rangle = \delta_{\lambda'\lambda}\delta_{\kappa'\kappa}. \tag{5b}$$

The eigenvalues λ are real by virtue of Eq. (2). They are also time independent, as we can deduce in the following simple way. By differentiating Eq. (5a) with respect to time, we obtain

$$\frac{\partial I}{\partial t}|\lambda, \kappa\rangle + I\frac{\partial}{\partial t}|\lambda, \kappa\rangle = \frac{\partial \lambda}{\partial t}|\lambda, \kappa\rangle + \lambda\frac{\partial}{\partial t}|\lambda, \kappa\rangle. \tag{6}$$

We also operate with the left-hand side of Eq. (1) on $|\lambda, \kappa\rangle$ to obtain

$$i\hbar \frac{\partial I}{\partial t}|\lambda, \kappa\rangle + IH|\lambda, \kappa\rangle - \lambda H|\lambda, \kappa\rangle = 0. \tag{7}$$

The scalar product of Eq. (7) with a state $|\lambda', \kappa'\rangle$ is

$$i\hbar\langle\lambda', \kappa'|\frac{\partial I}{\partial t}|\lambda, \kappa\rangle + (\lambda' - \lambda)\langle\lambda', \kappa'|H|\lambda, \kappa\rangle = 0, \tag{8}$$

implying

$$\langle\lambda, \kappa'|\frac{\partial I}{\partial t}|\lambda, \kappa\rangle = 0. \tag{9}$$

Now taking the scalar product of Eq. (6) with $|\lambda, \kappa\rangle$, we obtain

$$\frac{\partial \lambda}{\partial t} = \langle\lambda, \kappa|\frac{\partial I}{\partial t}|\lambda, \kappa\rangle = 0. \tag{10}$$

Since the eigenvalues are time independent, it is clear that the eigenstates must be time dependent.

In order to investigate the connection between eigenstates of I and solutions of the Schrödinger equation, we first write the equation of motion of $|\lambda, \kappa\rangle$, starting from Eq. (6) and using Eq. (10):

$$(\lambda - I)\frac{\partial}{\partial t}|\lambda, \kappa\rangle = \frac{\partial I}{\partial t}|\lambda, \kappa\rangle. \tag{11}$$

By taking the scalar product with $|\lambda', \kappa'\rangle$ and using Eq. (8) to eliminate

$$\langle\lambda', \kappa'|\frac{\partial I}{\partial t}|\lambda, \kappa\rangle,$$

we get

$$i\hbar(\lambda - \lambda')\langle\lambda', \kappa'|\frac{\partial}{\partial t}|\lambda, \kappa\rangle = (\lambda - \lambda')\langle\lambda', \kappa'|H|\lambda, \kappa\rangle. \tag{12}$$

From this, for $\lambda' \neq \lambda$, we infer

$$i\hbar\langle\lambda', \kappa'|\frac{\partial}{\partial t}|\lambda, \kappa\rangle = \langle\lambda', \kappa'|H|\lambda, \kappa\rangle. \tag{13}$$

Equation (12) does not imply

$$i\hbar\langle\lambda, \kappa'|\frac{\partial}{\partial t}|\lambda, \kappa\rangle = \langle\lambda, \kappa'|H|\lambda, \kappa\rangle.$$

If Eq. (13) held for $\lambda' = \lambda$ as well as for $\lambda' \neq \lambda$, then we would immediately deduce that $|\lambda, \kappa\rangle$ satisfies the Schrödinger equation, i.e., is a special solution for $|\ \rangle$.

The phase of $|\lambda, \kappa\rangle$ has not been fixed by our definitions. We assume that some definite phase has been chosen, but we are still free to multiply $|\lambda, \kappa\rangle$ by an arbitrarily time-dependent phase factor. That is, we can define a new set of eigenvectors of $I(t)$ related

to our initial set by a time-dependent gauge transformation

$$|\lambda, \kappa\rangle_\alpha = e^{i\alpha_{\lambda\kappa}(t)} |\lambda, \kappa\rangle, \tag{14}$$

where the $\alpha_{\lambda\kappa}(t)$ are arbitrary real functions of time. Because $I(t)$ is assumed not to contain time-derivative operators, the $|\lambda, \kappa\rangle_\alpha$ are orthonormal eigenstates of $I(t)$ just as are the $|\lambda, \kappa\rangle$. For $\lambda' \neq \lambda$, Eq. (13) also holds for matrix elements taken with respect to the new eigenstates. Each of the new eigenstates will satisfy the Schrödinger equation if we choose the phases $\alpha_{\lambda\kappa}(t)$ such that Eq. (13) holds for $\lambda' = \lambda$. This requirement is equivalent to the following first-order differential equation for the $\alpha_{\lambda\kappa}(t)$:

$$\hbar\delta_{\kappa\kappa'} \frac{d\alpha_{\lambda\kappa}}{dt} = \langle \lambda, \kappa' | \, i\hbar \frac{\partial}{\partial t} - H \, |\lambda, \kappa\rangle.$$

In order to satisfy this equation, the states $|\lambda, \kappa\rangle$ must be chosen such that the right-hand side vanishes for $\kappa' \neq \kappa$. This diagonalization is always possible because the operator $i\hbar(\partial/\partial t) - H$ is Hermitian. Once the states have been so chosen, the phases $\alpha_{\lambda\kappa}(t)$ are chosen to satisfy the simple equation

$$\hbar \frac{d\alpha_{\lambda\kappa}}{dt} = \langle \lambda, \kappa | \, i\hbar \frac{\partial}{\partial t} - H \, |\lambda, \kappa\rangle. \tag{15}$$

Since each of the new set of eigenstates of $I(t)$, $|\lambda, \kappa\rangle_\alpha$, satisfies the Schrödinger equation, the general solution is

$$|t\rangle = \sum_{\lambda,\kappa} c_{\lambda\kappa} e^{i\alpha_{\lambda\kappa}(t)} |\lambda, \kappa; t\rangle, \tag{16}$$

where the $c_{\lambda\kappa}$ are time-independent coefficients. All of the state vectors with which we have dealt so far are time dependent, and we have revised the notation in Eq. (16) slightly by indicating the dependences on time explicitly. Now the Schrödinger state vector is denoted by $|t\rangle$ and the eigenstates of the invariant by $|\lambda, \kappa; t\rangle$.

We now assume that in the remote past the Hamiltonian $H(t)$ settled into a constant operator $H(-\infty)$ having a complete, orthonormal set of time-independent eigenstates $|n; i\rangle$, n being a label for all relevant quantum numbers including the energy eigenvalue and i standing for "initial state." Similarly, we assume that the Hamiltonian settles into a constant operator $H(\infty)$ in the distant future and that it possesses time-independent eigenstates $|m; f\rangle$, m labeling the quantum numbers and f standing for "final state." The explicit time variation of $H(t)$ for intermediate times is arbitrary except for piecewise continuity; in particular, we do not exclude the possibility of variations rapid enough to render an analysis in terms of quasi-stationary states of $H(t)$ impossible. Our aim is to calculate the transition amplitude $T(n \to m)$ connecting an initial state $|n; i\rangle$ to a final state $|m; f\rangle$. Thus we consider the case in which the Schrödinger state vector $|-\infty\rangle$ in the remote past corresponds to an eigenstate $|n; i\rangle$ and, after tracing the exact time evolution of $|t\rangle$ into the distant future, we compute the overlap of $|\infty\rangle$ with the desired final state $|m; f\rangle$ to obtain the exact transition amplitude. The superposition coefficients of Eq. (16) for this problem are given by

$$c_{\lambda\kappa} = e^{-i\alpha_{\lambda\kappa}(-\infty)} \langle \lambda, \kappa; -\infty \,|\, n; i\rangle, \tag{17}$$

from which we obtain

$$|t\rangle = \sum_{\lambda,\kappa} \exp\{i[\alpha_{\lambda\kappa}(t) - \alpha_{\lambda\kappa}(-\infty)]\} \, |\lambda, \kappa; t\rangle \times \langle \lambda, \kappa; -\infty \,|\, n; i\rangle. \tag{18}$$

The transition amplitude is therefore given by

$$\begin{aligned} T(n \to m) &= \langle m; f \,|\, \infty\rangle \\ &= \sum_{\lambda,\kappa} \exp\{i[\alpha_{\lambda\kappa}(\infty) - \alpha_{\lambda\kappa}(-\infty)]\} \\ &\quad \times \langle m; f \,|\, \lambda, \kappa; \infty\rangle \langle \lambda, \kappa; -\infty \,|\, n; i\rangle. \end{aligned} \tag{19a}$$

Our discussion of the properties of $I(t)$ applies equally well to any operator that is an invariant corresponding to a given $H(t)$. In general, for a system of f degrees of freedom, there is an infinite family of such invariants, the members of which are functions of a set of f independent invariants. Two invariants $I_1(t)$ and $I_2(t)$ will, in general, have different eigenstates, different time derivatives, and different commutators with the Hamiltonian. In Secs. III and IV we give examples of this by constructing families of invariants for our two special systems in detail. Of course, we must obtain the same physical results no matter what invariant we use and, therefore, the choice of which particular invariant to use may be made on the basis of mathematical convenience. In order to illustrate explicitly that the physical results do not depend on our choice of invariant, we give a simple and direct proof that a transition amplitude like that of Eq. (19a) is indeed independent of our choice of invariant.

Suppose that we have two complete orthonormal sets of states, $|v; t\rangle$ and $|w; t\rangle$, all of which satisfy the time-dependent Schrödinger equation; and suppose that the states $|v; t\rangle$ are eigenstates of one set of operators, whose eigenvalues are labeled by v, and that the states $|w; t\rangle$ are eigenstates of a different set of operators, whose eigenvalues are labeled by w. The transition amplitude $T(n \to m)$ can be expressed as

$$T(n \to m) = \sum_v \langle m; f \,|\, v; \infty\rangle \langle v; -\infty \,|\, n; i\rangle \tag{19b}$$

or as

$$T(n \to m) = \sum_w \langle m; f \,|\, w; \infty\rangle\langle w; -\infty \,|\, n; i\rangle. \quad (19c)$$

We want to show directly that these two expressions are the same. The completeness of the states $|w; t\rangle$ requires

$$|v; t\rangle = \sum_w |w; t\rangle\langle w; t \,|\, v; t\rangle. \quad (20)$$

Operating on this equation with $(i\hbar(\partial/\partial t) - H)$, and using the facts that all of the states satisfy the Schrödinger equation and that the states $|w; t\rangle$ are orthogonal, we obtain

$$\frac{\partial}{\partial t}\langle w; t \,|\, v; t\rangle = 0. \quad (21)$$

Thus the quantity $\langle w; t \,|\, v; t\rangle$ is independent of time. We now use the completeness of the states $|v; t\rangle$ and $|w; t\rangle$, Eq. (21), and the orthonormality of the states $|w; t\rangle$ to rewrite Eq. (19b) as

$$\begin{aligned} T(n \to m) &= \sum_v \sum_w \sum_{w'} \langle m; f \,|\, w; \infty\rangle\langle w; \infty \,|\, v; \infty\rangle \\ &\quad \times \langle v; -\infty \,|\, w'; -\infty\rangle\langle w'; -\infty \,|\, n; i\rangle \\ &= \sum_v \sum_w \sum_{w'} \langle m; f \,|\, w; \infty\rangle\langle w; -\infty \,|\, v; -\infty\rangle \\ &\quad \times \langle v; -\infty \,|\, w'; -\infty\rangle\langle w'; -\infty \,|\, n; i\rangle \\ &= \sum_w \langle m; f \,|\, w; \infty\rangle\langle w; -\infty \,|\, n; i\rangle. \end{aligned} \quad (22)$$

Thus, Eqs. (19b) and (19c) are the same, as asserted.

We have used the $t \to \pm\infty$ limits in the above expressions as if the limits exist. In fact, however, the factors entering Eq. (19a) are generally undamped oscillating quantities for $t \to \pm\infty$ [for example, see Eq. (62), which gives the form of $\alpha_{\lambda\kappa}(t)$ for a time-dependent harmonic oscillator]. Nevertheless, this circumstance generates no difficulties in the calculation of transition probabilities in this limit. If t_1 and t_2 are finite times in the sufficiently remote past and future, respectively, then it is easily shown with the argument leading to Eq. (21) that the dependence of the transition amplitude of Eq. (19) on t_1 and t_2 is only $\exp[i(E_n t_1 - E_m t_2)/\hbar]$, where E_n and E_m are the initial and final state energies, respectively. The transition probability does not involve this phase factor, and therefore we shall continue to use the limits $t \to \pm\infty$.

Suppose for simplicity that the eigenstates of I are nondegenerate, so that the eigenvalue of I is the only quantum number required for describing the system. When this is so, as it is in our discussion of the time-dependent harmonic oscillator, then it is particularly convenient to choose an invariant having the property that it becomes time-independent as $t \to -\infty$ so that the commutator $[I(-\infty), H(-\infty)]$ vanishes. Then the normalized eigenvectors of $H(-\infty)$ and $I(-\infty)$ are identical to within arbitrary constant phase factors. Consequently, we may choose the initial state $|n; i\rangle$ simply to be a given eigenstate of $I(-\infty)$, say $|\lambda_n; -\infty\rangle$. Equation (19a) then reduces to

$$T(n \to m) = \exp\{i[\alpha_n(\infty) - \alpha_n(-\infty)]\}\langle m; f \,|\, \lambda_n; \infty\rangle, \quad (23)$$

and the transition probability is given by

$$\begin{aligned} P_{nm} &= |T(n \to m)|^2 \\ &= |\langle m; f \,|\, \lambda_n; \infty\rangle|^2. \end{aligned} \quad (24)$$

As $t \to \infty$, the invariant operator $I(t)$ in general remains time dependent and does not commute with the Hamiltonian. Therefore, the state $|\lambda_n; \infty\rangle$ in Eq. (24) is a superposition of eigenstates of $H(\infty)$; this is another expression of the fact that energy is not conserved in our system.

From the structure of Eq. (19a), it is apparent that we may express the transition amplitude as a matrix element of an S matrix by writing [keeping in mind the comment following Eq. (22)]

$$S = \sum_{\lambda,\kappa} e^{i\alpha_{\lambda\kappa}(\infty)} |\lambda, \kappa; \infty\rangle\langle\lambda, \kappa; -\infty| \, e^{-i\alpha_{\lambda\kappa}(-\infty)},$$

$$T(n \to m) = \langle m; f| \, S \, |n; i\rangle. \quad (25)$$

It is easily verified that this operator is unitary:

$$S^\dagger S = SS^\dagger = 1. \quad (26)$$

In the special case that the Hamiltonian operators in the remote past and distant future are identical, $H(-\infty) = H(\infty)$, so that the initial and final states are the same set, we may define an elastic scattering operator R in the standard fashion:

$$S = 1 + 2\pi i R. \quad (27)$$

The operator R describes the nondiagonal transitions just as S does, but subtracts a noninteracting part from the diagonal amplitudes so that $\langle n| \, R \, |n\rangle$ represents a "forward reaction amplitude" from the state $|n\rangle$ to the same state. The unitarity of the S matrix implies

$$\sum_m |\langle m| \, R \, |n\rangle|^2 = \frac{1}{\pi} \operatorname{Im}(\langle n| \, R \, |n\rangle), \quad (28)$$

which is a statement of the optical theorem: the total reaction probability is proportional to the imaginary part of the forward reaction amplitude.

III. APPLICATION TO TIME-DEPENDENT HARMONIC OSCILLATORS

A. A Family of Invariant Operators for a Time-Dependent Harmonic Oscillator

A time-dependent, one-dimensional harmonic oscillator is a system whose Hamiltonian operator is of the

form

$$H(t) = (1/2M)[p^2 + \Omega^2(t)q^2], \tag{29}$$

where q is a canonical coordinate, p is its conjugate momentum, $\Omega(t)$ is an arbitrary, piecewise-continuous function of time, and M is a real, positive mass parameter. The variables q and p satisfy the canonical commutation relation

$$[q, p] = i\hbar, \tag{30}$$

and the canonical equations of motion are

$$\dot{q} = \frac{1}{i\hbar}[q, H] = \frac{1}{M}p,$$

$$\dot{p} = \frac{1}{i\hbar}[p, H] = -\frac{1}{M}\Omega^2(t)q, \tag{31}$$

where the dots denote time derivative operators. To obtain the simple harmonic oscillator in the limit that $\Omega(t)$ is time independent, we would have to require that Ω be a real function. However, our discussion is equally valid if Ω is imaginary; all that is necessary is that Ω^2 be real, either positive or negative. Therefore we allow Ω^2 to be a positive or negative real function. In order for the usual adiabatic approximation to be applicable, M and $\Omega(t)$ must satisfy the criterion

$$\frac{1}{M} \gg \frac{1}{\Omega^2}\left|\frac{d\Omega}{dt}\right| \tag{32}$$

for all t. However, except where we discuss the adiabatic approximation specifically, we do not impose such a restriction on M and $\Omega(t)$.

For such oscillator systems, a convenient representation has been derived for the class of invariants which are homogeneous, quadratic expressions in the dynamical variables p and q.[1] This representation was constructed as a result of an examination of the classical trajectories, and the invariants were normalized in such a way as to reduce to the usual adiabatic invariant (energy divided by frequency) in the limit that the inequality (32) is satisfied. Here we present a purely quantum-mechanical derivation of this representation of the quadratic invariants.

We assume the existence of a Hermitian invariant of the homogeneous, quadratic form

$$I(t) = \tfrac{1}{2}[\alpha(t)q^2 + \beta(t)p^2 + \gamma(t)\{q, p\}_+], \tag{33}$$

where α, β, and γ are real functions of time, the multiplicative numerical factor has been chosen for convenience, and we have used the conventional anticommutator notation $\{q, p\}_+ \equiv qp + pq$. The time derivative of $I(t)$ is given by

$$\dot{I} = \frac{\partial I}{\partial t} + \frac{1}{i\hbar}[I, H]$$
$$= \frac{1}{2}\left[\left(\dot{\alpha} - \frac{2\Omega^2}{M}\gamma\right)q^2 + \left(\dot{\beta} + \frac{2}{M}\gamma\right)p^2 + \left(\dot{\gamma} + \frac{1}{M}\alpha - \frac{\Omega^2}{M}\beta\right)\{q, p\}_+\right]. \tag{34}$$

In order to satisfy Eq. (1), we demand

$$\dot{\alpha} = \frac{2\Omega^2}{M}\gamma,$$

$$\dot{\beta} = -\frac{2}{M}\gamma, \tag{35}$$

$$\dot{\gamma} = -\frac{1}{M}\alpha + \frac{\Omega^2}{M}\beta.$$

It is convenient to introduce another function $\sigma(t)$, defined by

$$\beta(t) = \sigma^2(t), \tag{36}$$

where $\sigma^2(t)$ is a real function of time. The second of Eqs. (35) then becomes

$$\gamma = -M\sigma\dot{\sigma}, \tag{37}$$

and the third equation yields

$$\alpha = M^2(\dot{\sigma}^2 + \sigma\ddot{\sigma}) + \Omega^2\sigma^2. \tag{38}$$

The first of Eqs. (35) imposes a constraint on $\sigma(t)$ which may be expressed in the form

$$\sigma\frac{d}{dt}(M^2\ddot{\sigma} + \Omega^2\sigma) + 3\dot{\sigma}(M^2\ddot{\sigma} + \Omega^2\sigma) = 0. \tag{39}$$

A first integral of Eq. (39) may immediately be written in the form

$$M^2\ddot{\sigma} + \Omega^2\sigma = c/\sigma^3, \tag{40}$$

where c is an arbitrary real constant of integration. Then Eq. (38) becomes

$$\alpha = M^2\dot{\sigma}^2 + c/\sigma^2. \tag{41}$$

The invariant may therefore be expressed in the form

$$I = \tfrac{1}{2}[(c/\sigma^2)q^2 + (\sigma p - M\dot{\sigma}q)^2], \tag{42}$$

with Eq. (40) as a subsidiary condition. The arbitrariness implied by the presence of the constant c is illusory, as may be verified by making the scale transformation

$$\sigma(t) = c^{\frac{1}{2}}\rho(t), \tag{43}$$

$\rho(t)$ being a new auxiliary function of time. After discarding a constant multiplicative factor $c^{\frac{1}{2}}$, we may

write Eq. (42) in the form

$$I = \tfrac{1}{2}[(1/\rho^2)q^2 + (\rho p - M\dot{\rho}q)^2], \tag{44}$$

and the auxiliary condition given by Eq. (40) becomes

$$M^2\ddot{\rho} + \Omega^2(t)\rho - 1/\rho^3 = 0. \tag{45}$$

In order to make $I(t)$ Hermitian, we choose only the real solutions of this equation.

Any particular solution of Eq. (45) may be used to construct an invariant operator of the form given by Eq. (44). We thus have obtained a family of operators which is in one-to-one correspondence with the family of solutions of the nonlinear differential equation (45). Later in this section we shall consider the special case of a system for which Ω is a constant function in the remote past and in the remote future, and we shall calculate transition amplitudes connecting states at these two times. In obtaining these transition amplitudes, we shall need the general solution of Eq. (45) for constant Ω, which we now derive.[6] The problem is not quite so trivial as it appears at first glance because the obvious time-independent, real solution,

$$\rho = \pm\,|\Omega|^{-\frac{1}{2}}, \tag{46}$$

is by no means the most general solution. According to the discussion of Sec. II, we are free to choose the solution given by Eq. (46) for $t \to -\infty$ if we like. But then we shall find that the time dependence of $\Omega(t)$ produces a more general solution for $\rho(t)$ as $t \to \infty$. The above choice for ρ as $t \to -\infty$ leads to the condition $[I(-\infty), H(-\infty)] = 0$, which, according to the discussion preceding Eq. (23), is a particularly convenient choice.

To find the general solution of Eq. (45) for constant Ω, we note that $\dot{\rho}$ is an integrating factor of this equation and immediately obtain the first integral

$$M^2\dot{\rho}^2 + \Omega^2\rho^2 + 1/\rho^2 = 2\,|\Omega|\cosh\delta, \tag{47}$$

where δ is an arbitrary real constant. The right-hand side of Eq. (47) is the integration constant, which, it turns out, must be greater than or equal to $2\,|\Omega|$ if ρ is to be real. The integration constant has been written in this way so that ρ will be real for all values of the real parameter δ. Solution of Eq. (47) is straightforward and leads to the result

$$\rho(t) = \gamma_1\,|\Omega|^{-\frac{1}{2}}\,[\cosh\delta + \gamma_2\sinh\delta\,\sin((2\Omega/M)t + \varphi)]^{\frac{1}{2}}, \tag{48}$$

where γ_1 and γ_2 can each independently assume the values ± 1, and φ is a real phase constant. The special solution of Eq. (46) corresponds to the case $\delta = 0$. Whenever $\Omega(t)$ becomes constant, the solution for $\rho(t)$ is necessarily of the form given by Eq. (48). Therefore, the transition amplitudes that we shall calculate are completely determined by the parameters in this expression that are appropriate to the limit $t \to \infty$, no matter how complicated or violent the time dependence of $\Omega(t)$ for earlier times. We shall express the transition amplitudes in terms of these parameters in Part C of this Section.

B. Eigenstates and Eigenvalues of $I(t)$ and the Phases

The eigenstates and eigenvalues of the invariant operator $I(t)$ may be found by an operator technique that is completely analogous to the method introduced by Dirac[7] for diagonalizing the Hamiltonian of a constant-frequency harmonic oscillator. Thus we define time-dependent canonical lowering and raising operators a and $a^\dagger$ by the relations

$$a = (2\hbar)^{-\frac{1}{2}}[(1/\rho)q + i(\rho p - M\dot{\rho}q)],$$
$$a^\dagger = (2\hbar)^{-\frac{1}{2}}[(1/\rho)q - i(\rho p - M\dot{\rho}q)]. \tag{49}$$

These operators satisfy the canonical commutation rule

$$[a, a^\dagger] = 1, \tag{50}$$

so that the operator $a^\dagger a$ is a number operator with nonnegative integer eigenvalues. The invariant operator given by Eq. (44) can be written in terms of a and $a^\dagger$ as

$$I = \hbar(a^\dagger a + \tfrac{1}{2}), \tag{51}$$

from which it follows[8] that the normalized eigenstates $|\lambda\rangle$ of I are the same as the normalized eigenstates $|s\rangle$ of $a^\dagger a$:

$$a^\dagger a\,|s\rangle = s\,|s\rangle, \quad s = 0, 1, 2, \cdots. \tag{52}$$

We specify the relative phases of these normalized eigenstates $|s\rangle$ by requiring the standard lowering and raising relations:

$$a\,|s\rangle = s^{\frac{1}{2}}\,|s-1\rangle,$$
$$a^\dagger\,|s\rangle = (s+1)^{\frac{1}{2}}\,|s+1\rangle. \tag{53}$$

The eigenvalue spectrum of I is given by

$$\lambda_s = (s + \tfrac{1}{2})\hbar, \quad s = 0, 1, 2, \cdots. \tag{54}$$

To effect the transformation of Eqs. (14) and (15) we need to calculate the diagonal matrix elements of the operators H and $\partial/\partial t$. The former are obtained by

[6] A method for expressing the general solution of Eq. (45) for arbitrary $\Omega(t)$ in terms of independent solutions of the equations for a classical oscillator has been described in Ref. 1.

[7] P. A. M. Dirac, *The Principles of Quantum Mechanics* (Clarendon Press, Oxford, 1947), 3rd ed. Also see A. Messiah, *Quantum Mechanics* (Interscience Publishers, New York, 1962), Vol. I.

[8] For the present we are omitting the time label t in our notation for these eigenstates. When it is required for clarity, we shall replace $|s\rangle$ by $|s; t\rangle$ to denote an eigenstate at time t.

using Eqs. (49) to express H in terms of a and $a^\dagger$ and then applying Eqs. (53):

$$\langle s|\, H\, |s\rangle = \frac{\hbar}{4M}\left(M^2\dot\rho^2 + \Omega^2\rho^2 + \frac{1}{\rho^2}\right)\langle s|\, \{a, a^\dagger\}_+\, |s\rangle$$
$$= \frac{1}{2M}\left(M^2\dot\rho^2 + \Omega^2\rho^2 + \frac{1}{\rho^2}\right)(s + \tfrac{1}{2})\hbar. \quad (55)$$

The Hamiltonian, of course, also has nondiagonal matrix elements since the representation defined by Eqs. (52) and (53) does not diagonalize this operator.

To evaluate the diagonal matrix elements of $\partial/\partial t$, we take the partial derivative of the second of Eqs. (53) with respect to time, and then take the appropriate scalar product, obtaining

$$\langle s|\frac{\partial}{\partial t}|s\rangle = \langle s-1|\frac{\partial}{\partial t}|s-1\rangle + s^{-\frac{1}{2}}\langle s|\frac{\partial a^\dagger}{\partial t}|s-1\rangle. \quad (56)$$

The expression for $\partial a^\dagger/\partial t$ in terms of a and $a^\dagger$ is

$$\frac{\partial a^\dagger}{\partial t} = \frac{1}{2}\left\{\left[-\frac{2\dot\rho}{\rho} + iM(\rho\ddot\rho - \dot\rho^2)\right]a + iM(\rho\ddot\rho - \dot\rho^2)a^\dagger\right\}, \quad (57)$$

so that Eq. (56) becomes

$$\langle s|\frac{\partial}{\partial t}|s\rangle = \langle s-1|\frac{\partial}{\partial t}|s-1\rangle + i\frac{M}{2}(\rho\ddot\rho - \dot\rho^2)$$
$$= \langle 0|\frac{\partial}{\partial t}|0\rangle + i\frac{s}{2}M(\rho\ddot\rho - \dot\rho^2). \quad (58)$$

It is clear that the anti-Hermitian character of $\partial/\partial t$ requires all diagonal matrix elements of $\partial/\partial t$ to be purely imaginary. However, no further information about $\langle 0|\,\partial/\partial t\,|0\rangle$ can be determined from Eq. (58); indeed, the choice of relative phases given by Eqs. (53) leaves the phase of a given state, say the state $|0\rangle$, undetermined. This time-dependent state can, in general, have a time-dependent phase factor, the choice of which is arbitrary. A convenient choice, which we now adopt, is one which makes $\langle 0|\,\partial/\partial t\,|0\rangle$ vanish in the limit that ρ becomes a constant, and which makes a "zero-point" contribution to Eq. (58):

$$\langle 0|\frac{\partial}{\partial t}|0\rangle = i\frac{M}{4}(\rho\ddot\rho - \dot\rho^2). \quad (59)$$

With this convention we can now write the general diagonal matrix element of $\partial/\partial t$ as

$$\langle s|\frac{\partial}{\partial t}|s\rangle = i\frac{M}{2}(\rho\ddot\rho - \dot\rho^2)(s + \tfrac{1}{2}). \quad (60)$$

The phases required for carrying out the transformation of Eq. (14) may be calculated by substituting Eqs. (55) and (60) into Eq. (15) to give

$$\frac{d\alpha_s}{dt} = -\frac{1}{2M}\left[M^2(\rho\ddot\rho - \dot\rho^2) + M^2\dot\rho^2 + \Omega^2\rho^2 + \frac{1}{\rho^2}\right](s + \tfrac{1}{2})$$
$$= -\frac{1}{M}(s + \tfrac{1}{2})\frac{1}{\rho^2}, \quad (61)$$

where we have made use of the subsidiary condition of Eq. (45). Thus the phase functions may be written in the form

$$\alpha_s(t) = -\frac{1}{M}(s + \tfrac{1}{2})\int^t dt'\,\frac{1}{\rho^2(t')}. \quad (62)$$

It is interesting to note that these phases are closely related to a quantity that occurs in the analysis of classical time-dependent harmonic oscillators.[1] In the classical case, the invariant I can be chosen as a generalized canonical momentum and the corresponding cyclic canonical coordinate is then equal to $-\alpha_s/(s + \frac{1}{2})$.

The off-diagonal matrix elements of H and $\partial/\partial t$, though not required for the purpose of the present discussion, are straightforward to compute and are given for completeness. The expression of H in terms of the raising and lowering operators immediately yields

$$\langle s'|\, H\, |s\rangle$$
$$= (\hbar/4)\{[M(\dot\rho^2 - \rho\ddot\rho) - 2i(\dot\rho/\rho)][s(s-1)]^{\frac{1}{2}}\delta_{s'+2,s}$$
$$+ [M(\dot\rho^2 - \rho\ddot\rho) + 2i(\dot\rho/\rho)][(s+1)(s+2)]^{\frac{1}{2}}\delta_{s',s+2}\}, \quad s' \neq s; \quad (63a)$$

and from Eq. (13) we obtain

$$\langle s'|\frac{\partial}{\partial t}|s\rangle = \frac{1}{i\hbar}\langle s'|\, H\, |s\rangle, \quad s' \neq s. \quad (63b)$$

C. Calculation of the Transition Probability

We assume that the Hamiltonian operator in the remote past, $H(-\infty)$, corresponds to a harmonic oscillator whose frequency parameter Ω_1 is constant and positive, and we choose the convenient form of $I(t)$ that leads to Eq. (23) by taking

$$\rho(-\infty) = \Omega_1^{-\frac{1}{2}}, \quad (64a)$$

from which follows

$$I(-\infty) = (M/\Omega_1)H(-\infty), \quad (64b)$$

so that

$$[I(-\infty), H(-\infty)] = 0. \quad (64c)$$

In the distant future the Hamiltonian is to settle into a harmonic oscillator Hamiltonian $H(\infty)$, with a

constant and positive frequency parameter Ω_2. The form of the invariant at any time is given by Eq. (44) or Eq. (51), and, as $t \to \infty$, the auxiliary function ρ necessarily satisfies Eq. (48) with $|\Omega|$ replaced by Ω_2. The detailed dynamics of the time variation determines the parameters δ and φ. We assume these parameters to be known, and we shall express the transition amplitude of Eq. (23) in terms of them. In general, specific numerical values for δ and φ can only be obtained by integrating Eq. (45) numerically.

Let us first suppose that the initial state $|n; i\rangle$ is the ground state of $H(-\infty)$. From Eq. (64b) it is clear that this state, apart from an arbitrary phase factor, is the same as the "ground state" $|0; -\infty\rangle$ of $I(-\infty)$. The Schrödinger state vector of the system at all later times is $e^{i\alpha_0(t)}\,|0; t\rangle$, where $\alpha_0(t)$ is given by Eq. (62), and this state vector is at all times an eigenvector (corresponding to the "ground state") of $I(t) = \hbar[a^\dagger(t)a(t) + \frac{1}{2}]$. We seek the transition amplitude to an eigenstate $|m; f\rangle$ of the final Hamiltonian, and, according to Eq. (23), this transition amplitude is given by

$$T(0 \to m) = \exp\{i[\alpha_0(\infty) - \alpha_0(-\infty)]\}\langle m; f \,|\, 0; \infty\rangle. \tag{23'}$$

The final Hamiltonian may be written in the form

$$H(\infty) = (\hbar\Omega_2/M)(b^\dagger b + \tfrac{1}{2}), \tag{65}$$

where

$$b = \left(\frac{\Omega_2}{2\hbar}\right)^{\frac{1}{2}}\left(q + i\frac{1}{\Omega_2}p\right),$$

$$b^\dagger = \left(\frac{\Omega_2}{2\hbar}\right)^{\frac{1}{2}}\left(q - i\frac{1}{\Omega_2}p\right), \quad [b, b^\dagger] = 1. \tag{66}$$

The lowering and raising operators of the invariant, $a(\infty)$ and $a^\dagger(\infty)$, may be expressed in terms of the lowering and raising operators of the final Hamiltonian, b and $b^\dagger$, by use of Eqs. (49):

$$a(\infty) = \eta(\infty)b + \zeta(\infty)b^\dagger,$$

$$a^\dagger(\infty) = \zeta^*(\infty)b + \eta^*(\infty)b^\dagger, \tag{67a}$$

where

$$\eta(t) = (4\Omega_2)^{-\frac{1}{2}}((1/\rho) + \Omega_2\rho - iM\dot{\rho}),$$

$$\zeta(t) = (4\Omega_2)^{-\frac{1}{2}}((1/\rho) - \Omega_2\rho - iM\dot{\rho}), \tag{67b}$$

and ρ is given by the final state form of Eq. (48). The condition that the transformation of Eq. (67) satisfy Eq. (50),

$$|\eta|^2 - |\zeta|^2 = 1,$$

clearly is satisfied.

To calculate the matrix element $\langle m; f \,|\, 0; \infty\rangle$, we expand the state $|0; \infty\rangle$ in terms of the eigenstates $|n; f\rangle$ of $H(\infty)$:

$$|0; \infty\rangle = \sum_{n=0}^{\infty} |n; f\rangle\langle n; f \,|\, 0; \infty\rangle. \tag{68}$$

By applying the lowering operator $a(\infty)$ to the left-hand side of this equation, we obtain zero, while the right-hand side may be transformed by use of Eq. (67). After regrouping terms, the resulting equation may be written as

$$0 = |0; f\rangle\,\eta\langle 1; f \,|\, 0; \infty\rangle + \sum_{n=1}^{\infty} |n; f\rangle[\eta(n+1)^{\frac{1}{2}}\langle n+1; f \,|\, 0; \infty\rangle + \zeta n^{\frac{1}{2}}\langle n-1; f \,|\, 0; \infty\rangle]. \tag{69}$$

In this equation and in what follows, by η and ζ we mean $\eta(\infty)$ and $\zeta(\infty)$. The orthonormality of the eigenvectors $|n; f\rangle$ then yields the recursion relations

$$\langle 1; f \,|\, 0; \infty\rangle = 0,$$

$$\langle n+1; f \,|\, 0; \infty\rangle = -\frac{\zeta}{\eta}\left(\frac{n}{n+1}\right)^{\frac{1}{2}}\langle n-1; f \,|\, 0; \infty\rangle, \tag{70}$$

which have the solution

$$\langle 2r+1; f \,|\, 0; \infty\rangle = 0,$$

$$\langle 2r; f \,|\, 0; \infty\rangle = \left(-\frac{\zeta}{\eta}\right)^r \frac{[(2r)!]^{\frac{1}{2}}}{2^r r!}\,\langle 0; f \,|\, 0; \infty\rangle, \tag{71}$$

where r is an integer. The first of Eqs. (71) expresses the usual parity selection rule: states of negative parity have vanishing overlap with a state of positive parity. By combining Eqs. (71) with the expansion of Eq. (68) and imposing the normalization requirement on the state $|0; \infty\rangle$, we obtain

$$|\langle 0; f \,|\, 0; \infty\rangle|^2 = \left[\sum_{r=0}^{\infty}\left|\frac{\zeta}{\eta}\right|^{2r}\frac{(2r)!}{2^{2r}(r!)^2}\right]^{-1} = \left(1 - \left|\frac{\zeta}{\eta}\right|^2\right)^{\frac{1}{2}} = \frac{1}{|\eta|}, \tag{72}$$

where we have used the summation formula

$$\sum_{r=0}^{\infty}\frac{(2r)!}{2^{2r}(r!)^2}x^r = (1-x)^{-\frac{1}{2}}.$$

Therefore,

$$|T(0 \to 0)| = |\langle 0; f \,|\, 0; \infty\rangle| = |\eta|^{-\frac{1}{2}} = (4\Omega_2)^{\frac{1}{4}}\left[\left(\frac{1}{\rho} + \Omega_2\rho\right)^2 + M^2\dot{\rho}^2\right]^{-\frac{1}{4}} = \left(\frac{2}{1+\cosh\delta}\right)^{\frac{1}{4}}, \tag{73}$$

where we have used Eqs. (23′), (47), and (67b). This corresponds to a transition probability

$$P_{00} = \left(\frac{2}{1+\cosh\delta}\right)^{\frac{1}{2}}. \tag{74}$$

Similarly, Eqs. (71) lead to the more general transition probability

$$P_{0m} = \frac{m!}{2^m[(m/2)!]^2}\left(\frac{\cosh\delta - 1}{\cosh\delta + 1}\right)^{m/2}\left(\frac{2}{\cosh\delta+1}\right)^{\frac{1}{2}}, \quad m \text{ even},$$
$$= 0, \quad m \text{ odd}. \tag{75}$$

By repeatedly applying the raising operator $a^\dagger(\infty)$ to Eq. (68), we may express any eigenstate $|s;\infty\rangle$ of I in terms of the states $|m;f\rangle$, and thus we may compute the general amplitude $T(s\to m)$. The resulting transition probability is, according to Eq. (24), given by

$$\begin{aligned} P_{sm} &= |\langle m;f\,|\,s;\infty\rangle|^2 \\ &= \frac{1}{s!}\left|\sum_{n=0}^{\infty}\langle m;f|\,(\zeta^* b + \eta^* b^\dagger)^s\,|n;f\rangle\langle n;f\,|\,0;\infty\rangle\right|^2 \\ &= \frac{1}{s!}\left(\frac{2}{\cosh\delta+1}\right)^{\frac{1}{2}} \\ &\quad\times\left|\sum_{r=0}^{\infty}\left(-\frac{\zeta}{2\eta}\right)^r\frac{[(2r)!]^{\frac{1}{2}}}{r!}\right. \\ &\quad\left.\times\,\langle m;f|\,(\zeta^* b + \eta^* b^\dagger)^s\,|2r;f\rangle\right|^2. \end{aligned} \tag{76}$$

These probabilities, of course, obey the sum rule $\sum_m P_{sm} = 1$. This is easily verified by direct summation of Eq. (76):

$$\begin{aligned} \sum_{m=0}^{\infty} P_{sm} &= \frac{1}{s!}\left(\frac{2}{\cosh\delta+1}\right)^{\frac{1}{2}}\sum_{r=0}^{\infty}\sum_{r'=0}^{\infty}\left(-\frac{\zeta^*}{2\eta^*}\right)^{r'}\left(-\frac{\zeta}{2\eta}\right)^r \\ &\quad\times\frac{[(2r)!\,(2r')!]^{\frac{1}{2}}}{r!\,r'!} \\ &\quad\times\langle 2r';f|\,(\eta b + \zeta b^\dagger)^s(\zeta^* b + \eta^* b^\dagger)^s\,|2r;f\rangle. \end{aligned}$$

On the other hand, to within a phase factor Eq. (68) leads to

$$|s;\infty\rangle = (s!)^{-\frac{1}{2}}\left(\frac{2}{\cosh\delta+1}\right)^{\frac{1}{4}} \times\sum_{r=0}^{\infty}\left(-\frac{\zeta}{2\eta}\right)^r\frac{[(2r)!]^{\frac{1}{2}}}{r!}(\zeta^* b + \eta^* b^\dagger)^s\,|2r;f\rangle,$$

so that we obtain

$$\begin{aligned} \sum_{m=0}^{\infty} P_{sm} &= \langle s;\infty\,|\,s;\infty\rangle \\ &= 1. \end{aligned}$$

Of course, P_{sm} vanishes unless the initial and final states have the same parity.

The case $\delta = 0$ corresponds to a situation in which $\lim_{t\to\infty}\rho(t)$ is equal to the constant $\Omega_2^{-\frac{1}{2}}$, so that $\eta = 1$, and $\zeta = 0$. Eq. (76) then yields

$$P_{sm}\,(\delta = 0) = \delta_{sm}, \tag{77}$$

which also is the result given by the adiabatic approximation. We conclude that the rigorous transition probability coincides with the adiabatic transition probability whenever the continuous time evolution of the auxiliary function $\rho(t)$ leads to a final form $\rho(\infty) = \Omega_2^{-\frac{1}{2}}$, starting from the initial form $\rho(-\infty) = \Omega_1^{-\frac{1}{2}}$. It is clear that only a restricted class of $\Omega(t)$ functions will produce such a result, but the members of this class need by no means satisfy any adiabaticity requirement. The time evolution of such systems, while leading to Eq. (77), will in general be nonadiabatic.

D. The Adiabatic and Sudden Approximations

In the adiabatic limit,

$$\frac{M}{\Omega^2(t)}\frac{d\Omega}{dt} \equiv \theta(t), \quad |\theta(t)| \ll 1, \tag{78}$$

it has been shown for the classical theory[1] that the leading term in the expansion in powers of θ of the invariant of Eq. (44) is the usual adiabatic invariant, energy divided by frequency. In the quantum theory this statement becomes the assertion that the quantum number remains constant, implying Eq. (77). This equation, of course, holds independently of any particular representation for $\rho(t)$; however, the choice $\rho(-\infty) = \Omega_1^{-\frac{1}{2}}$ is especially convenient, and with this assumption the adiabatic condition implies $\delta = 0$. For the sake of completeness we furnish an outline of a simple proof of the adiabatic theorem.

We let the frequency parameter $\Omega(t)$ evolve continuously from an initial value Ω_1 in the remote past to a final value Ω_2 in the distant future, such that Eq. (78) remains valid for all times $-\infty < t < \infty$. Since, according to Eq. (78), the frequency cannot change sign, we take Ω_1 and Ω_2 both to be positive constants. Eq. (45) may be formally integrated to yield

$$\begin{aligned} M^2\dot\rho^2 + \Omega^2(t)\rho^2 + \frac{1}{\rho^2} &= 2\Omega_1 + 2\int_{-\infty}^{t} dt'\,\rho^2(t')\Omega(t')\frac{d\Omega}{dt'} \\ &= 2\Omega_1 + \frac{2}{M}\int_{-\infty}^{t} dt'\,\rho^2(t')\Omega^3(t')\theta(t'), \end{aligned} \tag{79}$$

provided that $\rho(-\infty) = \Omega_1^{-\frac{1}{2}}$. We make the ansatz

$$\rho(t) = \Omega^{-\frac{1}{2}}(t)[1 + \nu(t)],$$
$$\nu(-\infty) = 0. \tag{80}$$

In the limit that θ vanishes so that $\Omega(t)$ becomes constant, the function $\nu(t)$ also must vanish. Hence, if θ is an infinitesimal quantity, so is ν. The essence of the adiabatic theorem is that ν is a higher-order infinitesimal than θ. By differentiating Eq. (80) with respect to time, we obtain

$$\dot{\rho} = -(1/2M)\Omega^{\frac{1}{2}}\theta(1+\nu) + \Omega^{-\frac{1}{2}}\dot{\nu}, \tag{81}$$

which implies that $\dot{\nu}$ is an infinitesimal of the same order as θ. To see that ν is of higher order than θ, we substitute the ansatz of Eq. (80) into Eq. (79), retaining terms only up to first order in θ:

$$\Omega\left[(1+\nu)^2 + \frac{1}{(1+\nu)^2}\right] = 2\Omega_1 + 2\int_{-\infty}^{t} dt' \frac{d\Omega}{dt'}[1+\nu(t')]^2. \tag{82}$$

Since the left-hand side of this integral equation has no first-order contributions in ν, it follows that the solution to first order in θ consistent with the condition $\Omega(-\infty) = \Omega_1$ is

$$\nu(t) = 0, \tag{83}$$

establishing the theorem. Equation (80) then asserts $\rho(\infty) = \Omega_2^{-\frac{1}{2}}$, leading to $\delta = 0$ and the result of Eq. (77).

The sudden approximation is used to describe the time evolution of systems in which the Hamiltonian operator experiences a rapid change during a time interval which is short compared to the characteristic periods of the system. Such a change might be represented by a jump discontinuity in the function $\Omega(t)$ at a specified instant of time. The sudden theory asserts that the state vector remains constant across such a discontinuity; the transition amplitudes bridging the discontinuity are therefore given by simple overlap integrals of eigenstates of the Hamiltonian just before and after the discontinuity. This result, which is rigorous for instantaneous discontinuities, is the basis of the sudden approximation for "fast," but not instantaneous, changes in the Hamiltonian operator.

The rigorous transition amplitude of Eq. (19) contains all the features of the sudden theory for $\Omega(t)$ whose time histories involve jump discontinuities. We may easily derive the result of the sudden theory for the harmonic oscillator on the basis of a simple example. The continuity of $I(t)$ as an explicit function of time is guaranteed by requiring ρ and $\dot{\rho}$ to be continuous. The possibility of jump discontinuities in the piecewise-continuous Hamiltonian operator $H(t)$ is retained by making $\ddot{\rho}$ discontinuous.

Our example is conveniently based on a representation of ρ in the form

$$\begin{aligned}\rho(t) &= G_1(t)\Omega_1^{-\frac{1}{2}}, \quad t < 0,\\ &= G_2(t)\gamma_1\Omega_2^{-\frac{1}{2}}\\ &\quad \times [\cosh\delta + \gamma_2 \sinh\delta \sin((2\Omega_2/M)t + \varphi)]^{\frac{1}{2}}, \quad t > 0,\end{aligned} \tag{84}$$

where $G_1(t)$ and $G_2(t)$ are continuous functions of time with continuous derivatives, possessing the limits

$$\begin{aligned}\lim_{t\to-\infty} G_1(t) &= 1,\\ \lim_{t\to\infty} G_2(t) &= 1,\end{aligned} \tag{85}$$

and the remaining symbols are defined as in Eq. (48). Thus $\rho(t)$ corresponds to a time history with constant, positive frequency parameters Ω_1 and Ω_2 in the remote past and future, respectively, and arbitrary behavior at intermediate times. We demand that ρ and $\dot{\rho}$ be continuous at $t = 0$ and retain the possibility that $\ddot{\rho}$, and thus $\Omega(t)$, experience a jump discontinuity at this point that depends on the behavior of the G functions. For compactness we introduce the notation

$$\begin{aligned} g_{1,2} &\equiv G_{1,2}(0),\\ \Delta &\equiv \begin{vmatrix} G_1(0) & G_2(0) \\ \dfrac{dG_1}{dt}(0) & \dfrac{dG_2}{dt}(0) \end{vmatrix}.\end{aligned} \tag{86}$$

The continuity conditions at $t = 0$ determine the parameters φ and δ, the latter of which, after some algebra, may be written in the form

$$\cosh\delta = \frac{1}{2}\left(\frac{g_1^2\Omega_2}{g_2^2\Omega_1} + \frac{g_2^2\Omega_1}{g_1^2\Omega_2} + \frac{M^2\Delta^2}{g_2^4\Omega_1\Omega_2}\right). \tag{87}$$

It is easily verified that the right-hand side of Eq. (87) is greater than unity for positive frequencies. Substitution into the transition probability of Eq. (75) yields

$$P_{0,2n} = \frac{(2n)!}{2^{2n}(n!)^2}\left[\frac{(g_1^2\Omega_2 - g_2^2\Omega_1)^2 + M^2\Delta^2\left(\frac{g_1}{g_2}\right)^2}{(g_1^2\Omega_2 + g_2^2\Omega_1)^2 + M^2\Delta^2\left(\frac{g_1}{g_2}\right)^2}\right]^n \times \left[\frac{4g_1^2g_2^2\Omega_1\Omega_2}{(g_1^2\Omega_2 + g_2^2\Omega_1)^2 + M^2\Delta^2\left(\frac{g_1}{g_2}\right)^2}\right]^{\frac{1}{2}}. \tag{88}$$

The special case $G_1(t) = 1$, $G_2(t) = 1$ corresponds to a step-function discontinuity in the frequency Ω at $t = 0$, from a constant value Ω_1 to a constant value Ω_2. In this case we have $\Delta = 0$, and the exact formula of Eq. (88) reduces to the usual result of the sudden

theory:

$$P_{0,2n}(\text{sudden}) = \frac{(2n)!}{2^{2n}(n!)^2}\left(\frac{\Omega_2 - \Omega_1}{\Omega_2 + \Omega_1}\right)^{2n}\frac{2(\Omega_1\Omega_2)^{\frac{1}{2}}}{\Omega_2 + \Omega_1}. \quad (89)$$

The right-hand side of Eq. (89) is, of course, simply the modulus squared of the overlap between the ground state of the initial oscillator Hamiltonian and the $2n$th state of the final oscillator Hamiltonian, in accordance with the fundamental assertion of the sudden theory. The more general transition probability of Eq. (76) can be calculated similarly.

IV. APPLICATION TO CHARGED PARTICLE MOTION IN A TIME-DEPENDENT ELECTROMAGNETIC FIELD

A. The Physical System

We consider a particle of mass M and charge e moving in a classical, axially symmetric electromagnetic field defined by the vector potential

$$\mathbf{A} = \tfrac{1}{2}B(t)\mathbf{k} \times \mathbf{r} \quad (90a)$$

and the scalar potential

$$\varphi = \frac{1}{2}\frac{e}{Mc^2}\eta(t)r^2 = \frac{1}{2}\frac{e}{Mc^2}\eta(t)(x^2 + y^2), \quad (90b)$$

where $\mathbf{r}$ is the position vector, $\mathbf{k}$ is a unit vector along the symmetry axis, r is perpendicular distance from the symmetry axis, x and y are Cartesian coordinates perpendicular to the symmetry axis, $B(t)$ and $\eta(t)$ are arbitrary piecewise-continuous functions of time, and c is the speed of light. The potential φ corresponds to an axially symmetric, time-dependent uniform charge density equal to $-(1/2\pi)(e/Mc^2)\eta(t)$. The electric and magnetic fields are

$$\begin{aligned}\mathbf{E} &= -\nabla\varphi - \frac{1}{c}\dot{\mathbf{A}} \\ &= -\frac{e}{Mc^2}\eta(t)(x\mathbf{i} + y\mathbf{j}) - \frac{1}{2c}\dot{B}(t)\mathbf{k} \times \mathbf{r} \quad (91)\end{aligned}$$

and

$$\mathbf{B} = \nabla \times \mathbf{A} = B(t)\mathbf{k},$$

where $\mathbf{i}$ and $\mathbf{j}$ are unit vectors along the positive x and y directions, respectively, and $\mathbf{k} = \mathbf{i} \times \mathbf{j}$. Since the axial motion of a particle in these fields is trivial, we shall ignore it and treat only the motion perpendicular to the symmetry axis. The Hamiltonian for this system is

$$\begin{aligned}H &= \frac{1}{2M}\left(\mathbf{p} - \frac{e}{c}\mathbf{A}\right)^2 + e\varphi \\ &= \frac{1}{2M}(p_x^2 + p_y^2) + \frac{e^2}{2Mc^2}\left(\frac{B^2}{4} + \eta\right) \\ &\quad \times (x^2 + y^2) + \frac{eB}{2Mc}(yp_x - xp_y), \quad (92)\end{aligned}$$

where the operator $\mathbf{p} = \mathbf{i}p_x + \mathbf{j}p_y$ is the canonical momentum of the particle. The only nonvanishing commutators between the coordinates and momenta are, as usual,

$$[x, p_x] = [y, p_y] = i\hbar. \quad (93)$$

We introduce cylindrical coordinates r and θ and their conjugate momenta p_r and p_θ by the definitions

$$\begin{aligned}r &= (x^2 + y^2)^{\frac{1}{2}}, \\ \theta &= \tan^{-1}\left(\frac{y}{x}\right), \\ p_r &= \frac{1}{2}\left(\frac{x}{r}p_x + p_x\frac{x}{r} + \frac{y}{r}p_y + p_y\frac{y}{r}\right) \quad (94) \\ &= \frac{1}{r}(xp_x + yp_y) - \frac{i\hbar}{2}\frac{1}{r}, \\ p_\theta &= xp_y - yp_x.\end{aligned}$$

These operators are Hermitian, and the only nonvanishing commutators between them are

$$[r, p_r] = [\theta, p_\theta] = i\hbar. \quad (95)$$

Expressed in terms of these variables, the Hamiltonian given by Eq. (92) is

$$\begin{aligned}H = \frac{1}{2M}\Bigg[p_r^2 + \frac{\left(p_\theta - \frac{\hbar}{2}\right)\left(p_\theta + \frac{\hbar}{2}\right)}{r^2}\Bigg] \\ + \frac{e^2}{2Mc^2}\left(\frac{B^2}{4} + \eta\right)r^2 - \frac{eB}{2Mc}p_\theta. \quad (96)\end{aligned}$$

Because of the axial symmetry, p_θ is a constant of the motion, as is clearly evident from this form of the Hamiltonian. The usual wave equation[5] can be obtained from Eq. (96) by substituting

$$p_r \to -i\hbar\left(\frac{\partial}{\partial r} + \frac{1}{2r}\right),$$

which is the standard coordinate representation of the *Hermitian* operator p_r.

B. Connection with the Time-Dependent Harmonic Oscillator

The Cartesian operator variables for the particle can be related to variables that satisfy the same equations of motion as time-dependent harmonic oscillator variables by means of the *noncanonical* transformation

$$\begin{aligned}Q &= (x + iy)\exp\left[i\frac{e}{2Mc}\int^t B(t')\,dt'\right], \\ P &= \frac{c}{e}(p_x + ip_y)\exp\left[i\frac{e}{2Mc}\int^t B(t')\,dt'\right]. \quad (97)\end{aligned}$$

It is easily verified that the variables Q and P satisfy

$$\dot{Q} = \frac{e}{Mc} P,$$

$$\dot{P} = -\frac{e}{Mc} \Omega^2(t) Q, \tag{98}$$

where $\Omega^2(t)$ is defined by

$$\Omega^2(t) = \tfrac{1}{4}B^2(t) + \eta(t). \tag{99}$$

Equations (98) are identical in form to Eqs. (31) for the time-dependent harmonic oscillator. However, we emphasize that the transformation given by Eqs. (97) is *not canonical*, because Q and P satisfy the commutation relation $[Q, P] = 0$. Nevertheless, the transformation can be used to obtain an invariant for the charged particle because the I defined by Eq. (44) for the oscillator is an invariant as long as q and p satisfy Eqs. (31). It is not necessary that the canonical commutation relation, Eq. (30), be satisfied, nor that q and p be Hermitian. The invariant that we obtain from Eq. (44) (and also denote by I) is

$$I(t) = \frac{1}{2}\left\{\rho^{-2}(x + iy)^2 + \left(\frac{Mc}{e}\right)^2 \times \left[\frac{1}{M}\rho(p_x + ip_y) - \dot{\rho}(x + iy)\right]^2\right\} \times \exp\left[i\frac{e}{Mc}\int^t B(t')\,dt'\right], \tag{100}$$

where ρ is *any* particular solution of

$$\left(\frac{Mc}{e}\right)^2 \ddot{\rho} + \Omega^2(t)\rho - \rho^{-3} = 0, \tag{101}$$

with $\Omega^2(t)$ given by Eq. (99). It is easily verified by direct computation that the I defined by Eq. (100) satisfies

$$\dot{I} \equiv \frac{\partial I}{\partial t} + \frac{1}{i\hbar}[I, H] = 0$$

with H given by Eq. (92). The invariant I is neither Hermitian nor anti-Hermitian, but we shall derive from it a Hermitian invariant to which the theory discussed in Sec. II is applicable.

C. Derivation of a Hermitian Invariant[9]

We introduce time-dependent Cartesian coordinates and momenta such that the explicit time dependence of $I(t)$ is contained solely in a phase factor. These coordinates and momenta are defined by

$$X = \frac{1}{\rho}x, \qquad Y = \frac{1}{\rho}y,$$
$$P_X = \rho p_x - M\dot{\rho}x, \quad P_Y = \rho p_y - M\dot{\rho}y. \tag{102}$$

In order that the new variables be Hermitian, we choose ρ to be a *real* solution of Eq. (101). The transformation is canonical because the only nonvanishing commutators between the variables are

$$[X, P_X] = [Y, P_Y] = i\hbar. \tag{103}$$

Expressed in terms of the new variables, $I(t)$ can be written as

$$I(t) = \frac{1}{2}\left\{(X + iY)^2 + \left(\frac{c}{e}\right)^2 (P_X + iP_Y)^2\right\} \times \exp\left[i\frac{e}{Mc}\int^t B(t')\,dt'\right]. \tag{104}$$

Because of the axial symmetry, it is also convenient to introduce cylindrical operators associated with X, Y, P_X, and P_Y:

$$R = (X^2 + Y^2)^{\frac{1}{2}} = \frac{1}{\rho}r,$$

$$\theta = \tan^{-1}\left(\frac{Y}{X}\right) = \tan^{-1}\left(\frac{y}{x}\right),$$

$$P_R = \frac{1}{2}\left(\frac{X}{R}P_X + P_X\frac{X}{R} + \frac{Y}{R}P_Y + P_Y\frac{Y}{R}\right) = \frac{1}{R}(XP_X + YP_Y) - \frac{i\hbar}{2}\frac{1}{R} = \rho p_r - M\dot{\rho}r, \tag{105}$$

$$p_\theta = XP_Y - YP_X = xp_y - yp_x.$$

This transformation is also canonical because, as before, the only nonvanishing commutators are

$$[R, P_R] = [\theta, p_\theta] = i\hbar. \tag{106}$$

In terms of these operators, $I(t)$ can be written as

$$I(t) = \frac{1}{2}\left\{\exp\left[2i\left(\theta + \frac{e}{2Mc}\int^t B(t')\,dt'\right)\right]\right\}(C + iD), \tag{107}$$

where C and D are Hermitian operators given by

$$C = R^2 + \left(\frac{c}{e}\right)^2\left(P_R^2 - \frac{(p_\theta + \hbar)^2 - \frac{3}{4}\hbar^2}{R^2}\right),$$

$$D = \left(\frac{c}{e}\right)^2(p_\theta + \hbar)\left(P_R\frac{1}{R} + \frac{1}{R}P_R\right). \tag{108}$$

Equation (107) can also be rewritten in the similar form

$$I(t) = \tfrac{1}{2}(C_1 + iD_1)\exp\left[2i\left(\theta + \frac{e}{2Mc}\int^t B(t')\,dt'\right)\right], \tag{109}$$

where C_1 and D_1 are obtained from C and D, respectively, by replacing $(p_\theta + \hbar)$ by $(p_\theta - \hbar)$.

From Eqs. (107) and (109) it follows that the Hermitian operators $I^\dagger I$ and $II^\dagger$ are invariants that are independent of θ and differ from each other only by a constant operator depending on p_θ. Motivated by results for the corresponding classical system,[9] we seek to construct a θ-independent invariant of biquadratic form in R, P_R, and R^{-1} which can be written as a linear function of $I^\dagger I$:

$$\left[R^2 + \left(\frac{c}{e}\right)^2 P_R^2 + \frac{\beta}{R^2}\right]^2 = 4I^\dagger I + \gamma, \quad (110)$$

where β and γ are Hermitian constant operators that may depend on p_θ but not on θ, P_R, or R. The numerical factor multiplying $I^\dagger I$ in Eq. (110) is immediately obtained from the normalization of I given by Eqs. (107) and (108). Use of the operator $II^\dagger$ in Eq. (110) would merely have changed the value of γ. The solution of Eq. (110) is

$$\beta = \left(\frac{c}{e}\right)^2 (p_\theta - \tfrac{1}{2}\hbar)(p_\theta + \tfrac{1}{2}\hbar),$$
$$\gamma = 4\left(\frac{c}{e}\right)^2 (p_\theta + \hbar)^2. \quad (111)$$

Therefore the operator

$$I_1^2 = \left\{R^2 + \left(\frac{c}{e}\right)^2 \left[P_R^2 + \frac{(p_\theta - \tfrac{1}{2}\hbar)(p_\theta + \tfrac{1}{2}\hbar)}{R^2}\right]\right\}^2 \quad (112)$$

is the desired biquadratic invariant. Moreover, direct calculation shows that the operator I_1 [defined as the inside of the curly brackets of Eq. (112)] is itself a Hermitian invariant, i.e., it satisfies Eqs. (1) and (2) for the Hamiltonian of Eq. (96).

Finally, instead of working with I_1 directly, we define another Hermitian invariant K by

$$K = \tfrac{1}{4}(|e|/c)I_1 - \tfrac{1}{2}sp_\theta$$
$$= \frac{c}{4|e|}\left[P_R^2 + \frac{(p_\theta - \tfrac{1}{2}\hbar)(p_\theta + \tfrac{1}{2}\hbar)}{R^2}\right]$$
$$+ \frac{|e|}{4c}R^2 - \tfrac{1}{2}sp_\theta, \quad (113)$$

where

$$s = e/|e|. \quad (114)$$

The form of K is identical to the form of the Hamiltonian for a particle moving in a *time-independent* magnetic field [see Eq. (96)]. The general theory developed in Sec. II is applicable to K, and the eigenvalues and eigenvectors of K can be found elegantly by operator methods.

[9] This derivation is closely related to a derivation of an analogous invariant for the corresponding classical system. The treatment of the classical system is given in H. R. Lewis, Jr., Phys. Rev. **172**, 1313 (1968).

D. Eigenvalues and Eigenstates of K

Since the invariant K has the form of the Hamiltonian for a particle moving in a time-independent magnetic field, its eigenvalues and eigenvectors are known. The usual derivation[4,5] is in terms of confluent hypergeometric functions. However, it is possible to derive the eigenvalues and eigenvectors by purely operator techniques. This derivation, which we present here, was motivated by the work of Infeld,[10] although his method differs somewhat in detail from ours.

We define operators a, $a^\dagger$, b, and $b^\dagger$ by

$$b = \frac{1}{2}\left(\frac{sc}{|e|\hbar}\right)^{\frac{1}{2}}\left\{P_R - i\left[\frac{e}{c}R + \frac{(p_\theta + \tfrac{1}{2}\hbar)}{R}\right]\right\},$$
$$b^\dagger = \frac{1}{2}\left[\left(\frac{sc}{|e|\hbar}\right)^{\frac{1}{2}}\right]^*\left\{P_R + i\left[\frac{e}{c}R + \frac{(p_\theta + \tfrac{1}{2}\hbar)}{R}\right]\right\}, \quad (115)$$
$$a = be^{-i\theta}, \quad a^\dagger = e^{i\theta}b^\dagger.$$

The commutator of a with $a^\dagger$ is

$$[a, a^\dagger] = s. \quad (116)$$

In terms of a and $a^\dagger$, the expression for K can be written as

$$K = \hbar aa^\dagger - s(p_\theta + \tfrac{1}{2}\hbar)$$
$$= \hbar a^\dagger a - s(p_\theta - \tfrac{1}{2}\hbar). \quad (117)$$

The commutators of K with a and $a^\dagger$ are

$$[K, a] = [K, a^\dagger] = 0, \quad (118)$$

which implies that operation with a or $a^\dagger$ on an eigenstate of K produces another eigenstate of K with the same eigenvalue. The commutators of p_θ with a and $a^\dagger$ are

$$[p_\theta, a] = -\hbar a$$

and

$$[p_\theta, a^\dagger] = \hbar a^\dagger. \quad (119)$$

Therefore, a and $a^\dagger$ are, respectively, lowering and raising operators for the eigenvalues of p_θ.

Since K and p_θ commute, we can define simultaneous eigenstates of these two operators. Let $|j, n\rangle$ denote a normalized eigenstate for which the eigenvalue of K is $(j + \frac{1}{2})\hbar$ by definition, and the eigenvalue of p_θ is $n\hbar$, where n is an integer:

$$\langle j, n | j, n\rangle = 1,$$
$$K|j, n\rangle = (j + \tfrac{1}{2})\hbar\,|j, n\rangle, \quad (120)$$
$$p_\theta|j, n\rangle = n\hbar\,|j, n\rangle.$$

[10] L. Infeld, Phys. Rev. **59**, 737 (1941). Later developments of this interesting method can be found in the following references: (a) T. Inui, Progr. Theoret. Phys. (Kyoto) **3**, 168, 244 (1948); (b) L. Infeld and T. E. Hull, Rev. Mod. Phys. **23**, 21 (1951); (c) A. Joseph, *ibid.* **39**, 829 (1967); (d) C. A. Coulson and A. Joseph, *ibid.* **39**, 838 (1967).

Because a and $a^\dagger$ are lowering and raising operators for the eigenvalues of p_θ, the states $|j, n-1\rangle$ and $|j, n+1\rangle$ are proportional to $a|j, n\rangle$ and $a^\dagger|j, n\rangle$, respectively. Therefore, the requirement that all admissible states be normalizable means that the matrix elements $\langle j, n|\, a^\dagger a\,|j, n\rangle$ and $\langle j, n|\, aa^\dagger\,|j, n\rangle$ must be nonnegative:

$$\begin{aligned}\langle j, n|\, a^\dagger a\,|j, n\rangle &= \frac{1}{\hbar}\langle j, n|\, K + s(p_\theta - \tfrac{1}{2}\hbar)\,|j, n\rangle \\ &= (j + \tfrac{1}{2}) + (n - \tfrac{1}{2})s \geq 0, \\ \langle j, n|\, aa^\dagger\,|j, n\rangle &= \frac{1}{\hbar}\langle j, n|\, K + s(p_\theta + \tfrac{1}{2}\hbar)\,|j, n\rangle \\ &= (j + \tfrac{1}{2}) + (n + \tfrac{1}{2})s \geq 0. \quad (121)\end{aligned}$$

From this we immediately conclude that j must be an integer; otherwise, by repeated application of a or $a^\dagger$ we could obtain an unnormalizable state from an admissible state. The first of the inequalities (121) is the more restrictive for $s = 1$ and the second is the more restrictive for $s = -1$. Therefore we can replace these inequalities by the single inequality

$$j + ns \geq 0. \quad (122)$$

We can restrict the values of j further by expressing K in terms of X, Y, P_X, and P_Y as

$$K = \frac{c}{4|e|}\left[\left(P_X + \frac{e}{c}Y\right)^2 + \left(P_Y - \frac{e}{c}X\right)^2\right]. \quad (123)$$

Thus we see that

$$(j + \tfrac{1}{2})\hbar = \langle j, n|\, K\,|j, n\rangle$$

is the expectation value of the sum of the squares of two Hermitian operators and therefore cannot be negative.

With these results we can write the allowable solutions of the inequalities (121) as

$$j = l + \tfrac{1}{2}(|n| - sn), \quad (124)$$

where l is an integer that can assume any nonnegative value $(0, 1, 2, \cdots)$. For fixed j and $s = 1$ the minimum value of n that is allowed is $-j$, whereas the maximum value of n allowed for fixed j and $s = -1$ is j. Therefore the state $|j, -j\rangle$ for $s = 1$ and the state $|j, j\rangle$ for $s = -1$ are determined by

$$\left.\begin{aligned} p_\theta\,|j, -j\rangle &= -j\hbar\,|j, -j\rangle, \\ a\,|j, -j\rangle &= 0, \end{aligned}\right\} \text{ for } s = 1, \quad (125a)$$

and

$$\left.\begin{aligned} p_\theta\,|j, j\rangle &= j\hbar\,|j, j\rangle, \\ a^\dagger\,|j, j\rangle &= 0, \end{aligned}\right\} \text{ for } s = -1. \quad (125b)$$

All other admissible states are obtained by repeated operation with $a^\dagger$ on $|j, -j\rangle$ (for $s = 1$) or with a on $|j, j\rangle$ (for $s = -1$). Finally, by using Eqs. (121) and making a suitable choice of relative phases for the states, we obtain the following recursion formulas for the admissible normalized eigenstates:

$$\left.\begin{aligned} |j, n+1\rangle &= (j + n + 1)^{-\frac{1}{2}}a^\dagger\,|j, n\rangle \quad \text{for } s = 1, \\ |j, n-1\rangle &= (j - n + 1)^{-\frac{1}{2}}a\,|j, n\rangle \quad \text{for } s = -1. \end{aligned}\right\} \quad (126)$$

E. Calculation of the Phases

The matrix element $\langle j, n'|\, i\hbar(\partial/\partial t) - H\,|j, n\rangle$ vanishes for $n' \neq n$ because p_θ commutes with $\partial/\partial t$ and with H. Therefore the state $e^{i\alpha_{jn}}|j, n\rangle$ will be a normalized solution of the Schrödinger equation if we choose $\alpha_{jn}(t)$ as a solution of Eq. (15):

$$\hbar\frac{d\alpha_{jn}}{dt} = \langle j, n|\, i\hbar\frac{\partial}{\partial t} - H\,|j, n\rangle. \quad (15')$$

We begin by finding a recursion formula for the right-hand side of Eq. (15′) for $s = 1$ and for $s = -1$. For $s = 1$ we have

$$\begin{aligned}&\langle j, n|\, i\hbar\frac{\partial}{\partial t} - H\,|j, n\rangle \\ &= \frac{1}{j+n}\langle j, n-1|\, a\left(i\hbar\frac{\partial}{\partial t} - H\right)a^\dagger\,|j, n-1\rangle \\ &= \frac{1}{j+n}\langle j, n-1|\left\{\left(i\hbar\frac{\partial}{\partial t} - H\right)a\right. \\ &\quad \left.+ \left[a, i\hbar\frac{\partial}{\partial t} - H\right]\right\}a^\dagger\,|j, n-1\rangle \\ &= \langle j, n-1|\, i\hbar\frac{\partial}{\partial t} - H\,|j, n-1\rangle \\ &\quad + \frac{1}{j+n}\langle j, n-1|\left[a, i\hbar\frac{\partial}{\partial t} - H\right]a^\dagger\,|j, n-1\rangle. \end{aligned} \quad (127)$$

Similarly, for $s = -1$ we obtain

$$\begin{aligned}&\langle j, n|\, i\hbar\frac{\partial}{\partial t} - H\,|j, n\rangle \\ &= \langle j, n+1|\, i\hbar\frac{\partial}{\partial t} - H\,|j, n+1\rangle \\ &\quad + \frac{1}{j-n}\langle j, n+1|\left[a^\dagger, i\hbar\frac{\partial}{\partial t} - H\right]a\,|j, n+1\rangle. \end{aligned} \quad (128)$$

We only need calculate the commutator $[a, i\hbar(\partial/\partial t) - H]$ that appears in Eq. (127) because the commutator in Eq. (128) is related to it by

$$\left[a^\dagger, i\hbar\frac{\partial}{\partial t} - H\right] = -\left[a, i\hbar\frac{\partial}{\partial t} - H\right]^\dagger.$$

To evaluate $[a, i\hbar(\partial/\partial t) - H]$ we use the expressions for H and a in terms of r, p_r, and p_θ. The expression for H is given by Eq. (96), and the expression for a is

$$a = \frac{1}{2}\left(\frac{sc}{|e|\hbar}\right)^{\frac{1}{2}} \times \left\{\rho p_r - M\dot{\rho}r - i\left[\frac{e}{c}\frac{r}{\rho} + \rho\frac{(p_\theta + \frac{1}{2}\hbar)}{r}\right]\right\}e^{-i\theta}. \tag{129}$$

The commutator may be expressed in the form

$$\left[a, i\hbar\frac{\partial}{\partial t} - H\right] = -i\hbar\dot{a}, \tag{130}$$

where the right-hand side refers to the total time derivative operator. The evaluation is straightforward and particularly simple if one uses the Heisenberg equations of motion in evaluating $\dot{a}$. The result is

$$\left[a, i\hbar\frac{\partial}{\partial t} - H\right] = \frac{e\hbar}{Mc}\left(\frac{B}{2} - \frac{1}{\rho^2}\right)a$$

and

$$\left[a^\dagger, i\hbar\frac{\partial}{\partial t} - H\right] = -\frac{e\hbar}{Mc}\left(\frac{B}{2} - \frac{1}{\rho^2}\right)a^\dagger. \tag{131}$$

We now substitute the commutators given by Eqs. (131) into Eqs. (127) and (128) and use Eqs. (121) to obtain

$$\langle j, n|\, i\hbar\frac{\partial}{\partial t} - H\,|j, n\rangle = \langle j, n-1|\, i\hbar\frac{\partial}{\partial t} - H\,|j, n-1\rangle + \frac{e\hbar}{Mc}\left(\frac{B}{2} - \frac{1}{\rho^2}\right) \quad \text{for } s = 1 \tag{132}$$

and

$$\langle j, n|\, i\hbar\frac{\partial}{\partial t} - H\,|j, n\rangle = \langle j, n+1|\, i\hbar\frac{\partial}{\partial t} - H\,|j, n+1\rangle - \frac{e\hbar}{Mc}\left(\frac{B}{2} - \frac{1}{\rho^2}\right) \quad \text{for } s = -1. \tag{133}$$

We are still free to choose the phase of $|j, -j\rangle$ for $s = 1$ and the phase of $|j, j\rangle$ for $s = -1$ arbitrarily. We choose these phases in such a way that the solution of Eqs. (132) and (133) is

$$\langle j, n|\, i\hbar\frac{\partial}{\partial t} - H\,|j, n\rangle = [n + (j + \tfrac{1}{2})s]\frac{e\hbar}{Mc}\left(\frac{B}{2} - \frac{1}{\rho^2}\right). \tag{134}$$

The expression for the phase $\alpha_{jn}(t)$ that we obtain by substituting the matrix element given by Eq. (134) into Eq. (15′) is

$$\alpha_{jn}(t) = [n + (j + \tfrac{1}{2})s]\frac{e}{Mc}\int^t dt'[\tfrac{1}{2}B(t') - \rho^{-2}(t')]. \tag{135}$$

Using Eq. (135), we may construct the time-dependent Schrödinger state vector according to the prescription of Eq. (16), and hence we may compute transition probabilities for processes analogous to those treated in Sec. III. It should be pointed out that the definition of Eq. (99) may lead to both positive and negative values of $\Omega^2(t)$. The latter situation arises when the sign of the particle charge is the same as the sign of the background charge density, provided that the instantaneous Larmor frequency $|eB|/2Mc$ is less than the "electrostatic oscillation frequency" $(2\pi e\sigma/M)^{\frac{1}{2}}$, where σ is the background charge density. Under these circumstances the asymptotic form for real $\rho(t)$ given by Eq. (48) has to be modified appropriately. The transition probability formalism of Sec. III is directly transcribable to the present case of charged particle motion only if $\Omega^2(\pm\infty)$ is positive.

Mth Power of an $N \times N$ Matrix and Its Connection with the Generalized Lucas Polynomials

RICHARD BARAKAT*
Division of Engineering and Applied Physics, Harvard University, Cambridge, Massachusetts
AND
ELIZABETH BAUMANN
Itek Corporation, Lexington, Massachusetts

(Received 2 December 1968)

The Mth power of an $N \times N$ matrix is expressed via the Cayley–Hamilton theorem as a linear combination of the lower powers of the matrix. The polynomial coefficients of the lower powers of the matrix are expressed in terms of polynomials in N variables, termed the generalized Lucas polynomials. The independent variables in the generalized Lucas polynomials are the traces of the lower powers of the matrix.

1. INTRODUCTION

Many problems in applied mathematics require a knowledge of the Mth power of an $N \times N$ matrix. If purely numerical results are desired, then the well-known Sylvester interpolation scheme[1] is probably the most elegant and efficient algorithm. Unfortunately, the scheme requires a knowledge of the eigenvalues of the matrix under consideration and thus is not particularly useful for obtaining explicit algebraic results in terms of the matrix elements. An alternative method is to employ the Caley–Hamilton theorem, along with continued matrix multiplication, in order to exhibit the Mth power of the matrix as a linear combination of lower powers of the matrix.

We will show in the present paper that the coefficients of the lower powers of the matrix are polynomials in the traces of the powers of the matrix. Furthermore, these polynomials (in many variables) are natural generalizations of the Lucas polynomials employed in number theory. Various properties of these polynomials are listed, and the Mth power of an $N \times N$ matrix is explicitly written in terms of these generalized Lucas polynomials.

2. CAYLEY–HAMILTON THEOREM

If X be an $N \times N$ matrix, the Cayley–Hamilton theorem states that X satisfies its own characteristic equation

$$\sum_{l=0}^{N} (-1)^l \phi_l^N \mathsf{X}^{l-1} = 0, \quad \mathsf{X}^0 \equiv \mathsf{I}, \tag{1}$$

where $\phi_l^{(N)}$ is the lth symmetric polynomial formed out of the eigenvalues of X. Note that $\phi_l^{(N)} \equiv \det \mathsf{X}$.

Rather than deal with the eigenvalues of X, let us work with the traces of X^l, $\sigma_l^{(N)} \equiv \mathrm{Tr}\,(\mathsf{X}^l)$, where N denotes the size of the matrix. It is a well-known fact that if X is $N \times N$, then a set of invariants of X is $\sigma_l^{(N)}$ $(l = 1, 2, \cdots, N)$. Any value of $\sigma_l^{(N)}$ for $l > N$ can be expressed in terms of the N invariants $\sigma_1^{(N)}, \cdots, \sigma_N^{(N)}$. The connection between ϕ_l and σ_l is outlined in Ref. 2 and is given by the following sequence of relations:

$$\begin{aligned}
\phi_1 &= \sigma_1, \\
2\phi_2 &= \phi_1\sigma_1 - \sigma_2, \\
3\phi_3 &= \phi_2\sigma_1 - \phi_1\sigma_2 + \sigma_3, \\
4\phi_4 &= \phi_3\sigma_1 - \phi_2\sigma_2 + \phi_1\sigma_3 - \sigma_4, \\
5\phi_5 &= \phi_4\sigma_1 - \phi_3\sigma_2 + \phi_2\sigma_3 - \phi_1\sigma_4 + \sigma_5, \\
&\;\;\vdots
\end{aligned}$$

These equations can be sequentially solved for $\sigma_1, \sigma_2, \cdots$; the final result is

$$\begin{aligned}
\phi_1 &= \sigma_1, \\
\phi_2 &= \tfrac{1}{2}(\sigma_1^2 - \sigma_2), \\
\phi_3 &= \tfrac{1}{6}(\sigma_1^3 - 3\sigma_1\sigma_2 + 2\sigma_3), \\
\phi_4 &= \tfrac{1}{48}(2\sigma_1^4 - 9\sigma_1^2\sigma_2 + 16\sigma_1\sigma_3 + 3\sigma_2^2 - 12\sigma_4), \\
\phi_5 &= \tfrac{1}{240}(2\sigma_1^5 - 17\sigma_1^3\sigma_2 + 40\sigma_1^2\sigma_3 + 27\sigma_1\sigma_2^2 \\
&\qquad - 60\sigma_1\sigma_4 - 40\sigma_2\sigma_3 + 48\sigma_5), \\
&\;\;\vdots
\end{aligned}$$

We employ Eq. (1) in our quest for determining the Nth power of an $M \times M$ matrix, by continued post multiplication of Eq. (1) by X. If, for example,

* On leave of absence from Itek Corporation, Lexington, Massachusetts.

[1] F. R. Ganthmacher, *The Theory of Matrices* (Chelsea Publishing Co., New York, 1959), Vols. 1 and 2.

[2] H. Weyl, *The Classical Groups* (Princeton University Press, Princeton, N.J., 1946).

$N = 3$, we easily obtain the sequence

$$\mathsf{X}^3 = \phi_1\mathsf{X}^2 - \phi_2\mathsf{X} + \phi_3\mathsf{I},$$
$$\mathsf{X}^4 = (\phi_1^2 - \phi_2)\mathsf{X}^2 + (\phi_3 - \phi_1\phi_2)\mathsf{X} + \phi_1\phi_3\mathsf{I},$$
$$\mathsf{X}^5 = (\phi_1^3 - 2\phi_1\phi_2 + \phi_3)\mathsf{X}^2 + (-\phi_1^2\phi_2 + \phi_1\phi_3 + \phi_2^2)\mathsf{X} + \phi_3(\phi_1^2 - \phi_2)\mathsf{I},$$
$$\vdots$$

We can always express the Mth power of an $N \times N$ matrix ($M \geq N$) in terms of a set of polynomials in the N variables $\phi_1, \phi_2, \cdots, \phi_N$ multiplied by X^{N-1}, $\mathsf{X}^{N-2}, \cdots, \mathsf{X}, \mathsf{I}$.

The case $N = 2$ frequently occurs in a variety of physical problems (iterated networks, thin films, etc.) and the special case of unit determinant (i.e., $\phi_2^{(2)} = 1$) has been exhaustively studied.[3] The general case, $\phi_2^{(2)} \neq 0$, was treated independently by Herpin[4] and Barakat[5]; they showed that the appropriate polynomials in two variables are the Lucas polynomials. Extension of the analysis to $N > 2$ is the subject of the present note.

3. GENERALIZED LUCAS POLYNOMIALS

Lucas,[6] in his work on number theory, studied a class of polynomials in two variables. The polynomial is termed $U_n(P, Q)$ and is defined by the recurrence relation

$$U_{n+2}(P, Q) = PU_{n+1}(P, Q) - QU_{n-1}(P, Q) \quad (2)$$

together with the initial conditions $U_0 = 0$, $U_1 = 1$. In the special case $Q = 1$, these polynomials reduce to the Chebyschev polynomials of the second kind $S_n(P)$, as defined in Lanczos[7]; in fact,

$$U_n(P, 1) = S_{n-1}(P) = \begin{cases} \sin n\theta/\sin\theta, & P \equiv 2\cos\theta, \\ \sinh n\phi/\sinh\phi, & P \equiv 2\cosh\phi, \end{cases} \quad n \geq 1. \quad (3)$$

The former expression holds for $|P| \leq 2$, the latter for $|P| \geq 2$. With $Q = -1$, the Lucas polynomials are equivalent to a set of polynomials, termed the Fibonnacci polynomials,[8,9] which have found some application in the theory of ladder networks.

An explicit expression for $U_n(P, Q)$ is

$$U_{n+1}(P, Q) = P^n - \binom{n-1}{1}P^{n-2}Q + \binom{n-2}{2}P^{n-4}Q^2 + \cdots, \quad (4)$$

the series terminating when the exponent of P or Q is negative.

Motivated by the discussion contained in the previous section, we introduce a generalization of the Lucas polynomials to a class of polynomials (hereafter termed generalized Lucas polynomials) in N variables, namely $U_n^{(N)}(\phi_1, \phi_2, \cdots, \phi_N)$. This function is defined by the recurrence relation

$$U_{n+N}^{(N)}(\phi_1, \phi_2, \cdots, \phi_N) = \phi_1 U_{n+N-1}^{(N)} - \phi_2 U_{n+N-2}^{(N)} + \phi_3 U_{n+N-3}^{(N)} - \phi_4 U_{n+N-4}^{(N)} + \cdots \pm \phi_N U_n^{(N)} \quad (5)$$

together with the initial conditions:

$$U_0^{(N)} = U_1^{(N)} = \cdots = U_{N-2}^{(N)} = 0, \quad U_{N-1}^{(N)} = 1.$$

$N = 2$, of course, is the usual Lucas polynomial situation. The first members of the polynomials $U_n^{(3)}$, $U_n^{(4)}$, and $U_n^{(5)}$ are listed in Table I. Examination of Table I and simple induction yields the reduction formulas

$$U_n^{(N)}(\phi_1, \phi_2, \cdots, \phi_{N-1}, 0) = U_{n-1}^{(N-1)}(\phi_1, \phi_2, \cdots, \phi_{N-1}),$$
$$U_n^{(N)}(\phi_1, \phi_2, \cdots, \phi_{N-2}, 0, 0) = U_{n-2}^{(N-2)}(\phi_1, \phi_2, \cdots, \phi_{N-2}), \quad (6)$$
$$\vdots$$
$$U_n^{(N)}(\phi_1, 0, \cdots, 0, 0) = \phi_1^{n-N+1}.$$

Although we were unable to obtain explicit expressions for general N, we found that $U^{(3)}$ could be expressed in closed form as

$$U_{n+2}^{(3)}(\phi_1, \phi_2, \phi_3) = U_{n+1}^{(2)}(\phi_1, \phi_2) + \sum_{k=1}\sum_{l=k}(-1)^{l-k}\binom{l}{k}\binom{n-l-k}{l}\phi_1^{n-2l-k}\phi_2^{l-k}\phi_3^k, \quad (7)$$

subject to the convention that if any exponents of ϕ_1, ϕ_2, ϕ_3 are negative, then the term is taken to be zero.

A closed form for $U^{(N)}$ is desirable, but it is just as easy to generate the polynomial via Eq. (5). In fact, Eq. (5) is ideally suited for programming on a computer using algebraic routines.

---

[3] V. Moweny, IRE Trans. Circuit Theory **11**, 232 (1964).
[4] A. Herpin, Compt. Rend. Acad. Sci. Paris **225**, 17 (1947).
[5] R. Barakat, J. Math. & Phys. **43**, 332 (1964).
[6] E. Lucas, *Theorie des nombres* (Gauthier-Villars, Paris, 1891).
[7] C. Lanczos, *Applied Analysis* (Prentice-Hall, Inc., Englewood Cliffs, N.J., 1956).
[8] E. Jacobsthal, Sitzungsber. Berliner Math. Ges. **17**, 43 (1919).
[9] S. Basin, Math. Mag. **37**, 83 (1964).

TABLE I. First members of the polynomials $U_n^{(3)}$, $U_n^{(4)}$, and $U_n^{(5)}$.

$n = 3, \cdots, 8$
$U_3^{(3)} = \phi_1$
$U_4^{(3)} = \phi_1^2 - \phi_2$
$U_5^{(3)} = \phi_1^3 - 2\phi_1\phi_2 + \phi_3$
$U_6^{(3)} = \phi_1^4 - 3\phi_1^2\phi_2 + 2\phi_1\phi_3 + \phi_2^2$
$U_7^{(3)} = \phi_1^5 - 4\phi_1^3\phi_2 + 3\phi_1^2\phi_3 + 3\phi_1\phi_2^2 - 2\phi_2\phi_3$
$U_8^{(3)} = \phi_1^6 - 5\phi_1^4\phi_2 + 4\phi_1^3\phi_3 + 6\phi_1^2\phi_2^2 - 6\phi_1\phi_2\phi_3 - \phi_2^3 + \phi_3^2$
.
.
.

$n = 4, \cdots, 9$
$U_4^{(4)} = \phi_1$
$U_5^{(4)} = \phi_1^2 - \phi_2$
$U_6^{(4)} = \phi_1^3 - 2\phi_1\phi_2 + \phi_3$
$U_7^{(4)} = \phi_1^4 - 3\phi_1^2\phi_2 + 2\phi_1\phi_3 + \phi_2^2 - \phi_4$
$U_8^{(4)} = \phi_1^5 - 4\phi_1^3\phi_2 + 3\phi_1^2\phi_3 + 3\phi_1\phi_2^2 - 2\phi_1\phi_4 - 2\phi_2\phi_3$
$U_9^{(4)} = \phi_1^6 - 5\phi_1^4\phi_2 + 4\phi_1^3\phi_3 + 6\phi_1^2\phi_2^2 - 3\phi_1^2\phi_4 - 6\phi_1\phi_2\phi_3 - \phi_2^3 + 2\phi_2\phi_4 + \phi_3^2$
.
.
.

$n = 5, \cdots, 10$
$U_5^{(5)} = \phi_1$
$U_6^{(5)} = \phi_1^2 - \phi_2$
$U_7^{(5)} = \phi_1^3 - 2\phi_1\phi_2 + \phi_3$
$U_8^{(5)} = \phi_1^4 - 3\phi_1^2\phi_2 + 2\phi_1\phi_3 + \phi_2^2 - \phi_4$
$U_9^{(5)} = \phi_1^5 - 4\phi_1^3\phi_2 + 3\phi_1^2\phi_3 + 3\phi_1\phi_2^2 - 2\phi_1\phi_4 - 2\phi_2\phi_3 + \phi_5$
$U_{10}^{(5)} = \phi_1^6 - 5\phi_1^4\phi_2 + 4\phi_1^3\phi_3 + 6\phi_1^2\phi_2^2 - 3\phi_1^2\phi_4 - 6\phi_1\phi_2\phi_3 + 2\phi_1\phi_5 - \phi_2^3 + \phi_3^2 + 2\phi_2\phi_4$
.
.
.

4. Mth POWER OF MATRIX

Armed with a knowledge of the generalized Lucas polynomials, Eq. (1), and a large measure of patience, we proceeded, by a tedious induction, to prove that the Mth power of an $N \times N$ matrix ($M \geq N$) is given by

$$\begin{aligned}\mathsf{X}^M = {} & U_M^{(N)}\mathsf{X}^{N-1} + [U_{M+1}^{(N)} - \phi_1 U_M^{(N)}]\mathsf{X}^{N-2} \\ & + [\phi_3 U_{M-1}^{(N)} - \phi_4 U_{M-2}^{(N)} \\ & + \cdots (\pm)\phi_N U_{M-N+2}^{(N)}]\mathsf{X}^{N-3} \\ & + [-\phi_4 U_{M-1}^{(N)} + \phi_5 U_{M-2}^{(N)} \\ & - \cdots (\pm)\phi_N U_{M-N+3}^{(N)}]\mathsf{X}^{N-4} \\ & + [\phi_5 U_{M-1}^{(N)} - \phi_6 U_{M-2}^{(N)} \\ & + \cdots (\pm)\phi_N U_{M-N+4}^{(N)}]\mathsf{X}^{N-5} \\ & + \cdots + [\phi_N U_{M-1}^{(N)}]\mathsf{I}. \end{aligned} \tag{8}$$

We have thus expressed the polynomial coefficients of the lower powers of X in terms of linear combinations of the appropriate generalized Lucas polynomials.

In the special cases $N = 3, 4$, Eq. (8) becomes

$$\mathsf{X}^M = U_M^{(3)}\mathsf{X}^2 + [U_{M+1}^{(3)} - \phi_1 U_M^{(3)}]\mathsf{X} + \phi_3 U_{M-1}^{(3)}\mathsf{I}, \tag{9}$$

$$\begin{aligned}\mathsf{X}^M = {} & U_M^{(4)}\mathsf{X}^3 + [U_{M+1}^{(4)} - \phi_1 U_M^{(4)}]\mathsf{X}^2 \\ & + [\phi_3 U_{M-1}^{(4)} - \phi_4 U_{M-2}^{(4)}]\mathsf{X} - \phi_4 U_{M-1}^{(4)}\mathsf{I}. \end{aligned} \tag{10}$$

As an application of these formulas, we have employed the case $N = 3$ in a study of the coupled linear differential equations describing the traveling-wave-tube parametric amplifier.[10] The case $N = 4$ has been applied to some problems in the theory of non-image-forming optical instruments described by Mueller matrices.[11]

[10] To be published.
[11] To be published.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Clebsch–Gordan Coefficients and Functions over Cosets

WILLIAM H. KLINK
Department of Physics and Astronomy, The University of Iowa, Iowa City, Iowa

(Received 1 December 1968)

Functions over cosets are used to obtain Clebsch–Gordan coefficients of induced unitary representations of noncompact groups. In particular, the Clebsch–Gordan coefficients of the Poincaré group are obtained using functions over cosets. It is also shown how the Clebsch–Gordan coefficients of other groups having induced irreducible representations, such as the Lorentz group, may be obtained.

INTRODUCTION

One of the problems which has arisen in the applications of noncompact groups to elementary particle physics is the explicit construction of the Clebsch–Gordan or coupling coefficients defined relative to some complete set of commuting observables. In higher-symmetry schemes, such as $U(6, 6)$ and *In* $SL(6, C)$,[1] the Clebsch–Gordan coefficients could be obtained because finite-dimensional representations of these groups were used; these groups, however, were plagued, among other things, by the lack of unitarity arising from the finite-dimensional representations. In working with higher symmetry schemes it is thus important to obtain the Clebsch–Gordan coefficients for the unitary (and therefore infinite-dimensional) representations of noncompact groups.

There are instances where the Clebsch–Gordan coefficients of the unitary representations of noncompact groups have been obtained. Jacob and Wick[2] obtained the Clebsch–Gordan coefficients for the Poincaré group, and since their work appeared many authors[3] have obtained the Clebsch–Gordan coefficients of the Poincaré group using different techniques than those of Jacob and Wick. The Clebsch–Gordan coefficients for the homogeneous Lorentz group have also recently been obtained,[4] at least for the so-called principal series of representations.

In both of these groups, the unitary irreducible representations can be written as induced representations, in the sense in which Mackey[5] has defined this notion. Mackey has also shown how it is possible to decompose the tensor products of induced representations into direct integrals over double cosets of groups having induced irreducible representations. One of the difficulties in utilizing Mackey's theory in elementary particle physics is that the spaces on which the induced representations are defined are fairly abstract as far as the physicist is concerned. In this paper it will be shown how the use of functions over cosets enables one to make a connection between the abstract notion of induced representations and the Clebsch–Gordan coefficients of the group. Roughly speaking, the irreducible induced representations of noncompact groups are defined on function spaces whose functions have as arguments coset labels of the group. Let $[\chi]$ be a set of irreducible representation labels of a group G and let $\{x\}$ be a set of eigenvalues chosen from a complete set of commuting observables of G. Then the quantity

$$D^{[\chi]}_{x'x}(G) \equiv \langle [\chi]x' |\, U(G)\, |[\chi]x\rangle$$

can be thought of as a function over the group; here, $U(G)$ is a unitary irreducible representation of G.

For example,

$$D^{[j]}_{m'm}(R) = \langle [j]m' |\, U(R)\, |[j]m\rangle$$

is a function over the three-dimensional rotation group, with R a rotation, and $D^{[j]}_{m'm}(R)$ a "matrix element" or a Wigner function.[6]

In particular, $D^{[\chi]}_{x'x}(g_c)$ can be thought of as a function over a coset; here g_c, an element of G, labels right cosets of G/H, with H a subgroup of G to be specified later. If the direct-integral decomposition of tensor products of functions defined over cosets is known, then the tensor-product decomposition of $D^{[\chi]}_{x'x}(g_c)$ can be obtained. But it will be shown that $D^{[\chi]}_{x'x}(g_c)$ can be thought of as a concrete realization of the basis elements $|[\chi]x\rangle$ and the Clebsch–Gordan coefficients are precisely the coefficients which reduce the tensor products of basis elements $|[\chi_1]x_1\rangle\, |[\chi_2]x_2\rangle$.

In Sec. I, the main ideas needed for getting the Clebsch–Gordan coefficients via functions over cosets

[1] A. Pais, Rev. Mod. Phys. **38**, 215 (1966); T. Fulton and G. Wess, Phys. Letters **15**, 177 (1965); A. Salam, R. Delbourgo, and J. Strathdee, Proc. Roy. Soc. (London) **A284**, 146 (1965).

[2] M. Jacob and G.-C. Wick, Ann. Phys. (Paris) **7**, 404 (1959); G.-C. Wick, Ann. Phys. (Paris) **18**, 65 (1962).

[3] See P. Moussa and R. Stora [*Lectures in Theoretical Physics* (University of Colorado Press, Boulder, Colorado, 1964), Vol. VIIa] for a list of references preceding their work. See also M. Kummer, J. Math. Phys. **7**, 997 (1966).

[4] R. Anderson, L. R. Raczka, M. A. Rashid, and P. Winternitz, "Clebsch–Gordan Coefficients for the Lorentz Group. I," IAEA Preprint, IC/67/50, Trieste, Italy, 1967.

[5] G. W. Mackey, *The Theory of Group Representations* (Dept. of Mathematics, The University of Chicago, Chicago, Ill., 1955).

[6] E. P. Wigner, *Group Theory* (Academic Press Inc., New York, 1959).

will be discussed; in order to present these ideas it will be necessary briefly to review the notions of induced representations, tensor products, and double-coset decompositions. As a simple example, the ideas sketched in Sec. I will be applied in Sec. II to the group P_2, the Poincaré group involving only one spatial dimension.

In Sec. III, the Poincaré group involving three spatial dimensions will be discussed; in particular, its unitary irreducible representations will be defined on functions over its cosets. The tensor-product decomposition for positive-mass representations of the Poincaré group will be given in Sec. IV. In this section, our work will draw on that of Moussa and Stora[3] and Rideau,[7] who have used Mackey's induced-representation theory to obtain the tensor-product decomposition of the Poincaré group. Finally, in Sec. V, the Clebsch–Gordan coefficients of the Lorentz group will be briefly discussed.

I. REVIEW OF INDUCED REPRESENTATIONS, TENSOR PRODUCTS, AND DOUBLE COSETS; THE CLEBSCH-GORDAN COEFFICIENTS

In this section the relevant parts of Mackey's theory of induced representations[5] will be briefly summarized, in order to see how this theory can be used to obtain Clebsch–Gordan coefficients. Consider a group G and let G be decomposed into right cosets with respect to a subgroup H:

$$G = \bigcup_c Hg_c \tag{1}$$

with g_c an element of G labeling the right cosets. Denote a representation of H by $\mathcal{K}$, defined on a vector space $\mathfrak{U}(\mathcal{K})$, and consider that class of functions f mapping G into $\mathfrak{U}(\mathcal{K})$ which have the property $f(hg) = \mathcal{K}(h)f(g)$, with h an element of H and g an element of G; this class of functions forms a new vector space

$$\mathfrak{V}(\mathcal{K}) = \{f \,|\, f(g) \in \mathfrak{U}(\mathcal{K}), f(hg) = \mathcal{K}(h)f(g), \text{ for all } h \text{ in } H, g \text{ in } G\} \tag{2}$$

on which a representation called the induced representation is defined:

$$U(g_0)f(g) = f(gg_0). \tag{3}$$

Notice that the condition $f(hg) = \mathcal{K}(h)f(g)$ means that it is possible to think of $f(g)$ as a function over a coset, since $f(g) = \mathcal{K}(h)f(g_c)$. Further, by restricting $f(g_c)$ to those functions which are square integrable over g_c, $\mathfrak{V}(\mathcal{K})$ becomes a Hilbert space and $U(g_0)$ becomes a unitary representation.[5] For the groups considered in this paper, the measures needed to get the norms of $f(g_c)$,[8]

$$\|f\|^2 = \int dg_c \, |f(g_c)|^2 < \infty, \tag{4}$$

are given in Refs. 7 and 9.

Consider now induced representations on two Hilbert spaces $\mathfrak{V}(\mathcal{K}_1)$ and $\mathfrak{V}(\mathcal{K}_2)$. The tensor-product space $\mathfrak{V}(\mathcal{K}_1) \otimes (\mathcal{K}_2)$ is the Hilbert space of functions $F(g_{c_1}, g_{c_2})$ with norm

$$\|F\|^2 = \int dg_{c_1} \, dg_{c_2} \, |F(g_{c_1}, g_{c_2})|^2 < \infty \tag{5}$$

and the induced representation on this tensor-product space is

$$U(g_0)F(g_{c_1}, g_{c_2}) = F(g_{c_1}g_0, g_{c_2}g_0). \tag{6}$$

Assume that $\mathfrak{V}(\mathcal{K}_1)$ and $\mathfrak{V}(\mathcal{K}_2)$ are irreducible spaces of G; the problem is to reduce the reducible tensor-product space into a direct integral of subspaces, each subspace being an irreducible subspace. Several techniques will be used in this paper to effect the direct-integral decomposition; in particular, Mackey has shown[5] that a partial decomposition of the tensor-product space can be obtained via a double-coset decomposition.

However, assume for the moment that the direct-integral decomposition is known and that $f_{[\chi,\eta]}(g_c)$ is an element of the irreducible subspace, with norm

$$\|f_{[\chi,\eta]}\|^2 = \int dg_c \, |f_{[\chi\eta]}(g_c)|^2 < \infty;$$

$[\chi, \eta]$ are the irreducible-representation labels and the degeneracy labels, respectively; η is needed whenever the multiplicity is greater than unity. Then,

$$\|F\|^2 = \int d[\chi, \eta] \, \|f_{[\chi,\eta]}\|^2 \tag{7}$$

is a direct integral over $[\chi, \eta]$ with weight function $d[\chi, \eta]$.

Now let S (the notation is that of Ref. 10) be the operator carrying functions from the tensor-product

[7] G. Rideau, Ann. Inst. Henri Poincaré **3**, 339 (1965).

[8] Actually, the norm of f,

$$\|f\|^2 = \int dg_c \, |f(g_c)|^2,$$

is only correct when one-dimensional representations of the inducing subgroup are being considered; in cases where higher-dimensional representations of the inducing subgroup are used, the definition of the norm is modified to be

$$\int dg_c \, \|f(g_c)\|^2_{\mathcal{K}},$$

where the subscript $\mathcal{K}$ refers to the length of the vector f in $\mathfrak{V}(\mathcal{K})$. In the case of the Poincaré group, the representations of $SU(2)$ are of dimension $2S + 1$ and, thus, it is necessary to use the more general definition of the norm of f.

[9] M. A. Naimark, *Linear Representations of the Lorentz Group* (The Macmillan Company, New York, 1964).

[10] M. A. Naimark, Am. Math. Soc. Trans. (2) **36**, 101 (1964).

space $\mathfrak{U}(\mathcal{H}_1) \otimes \mathfrak{U}(\mathcal{H}_2)$ to the direct-integral subspaces, that is,

$$S: F(g_{c_1}, g_{c_2}) \to f_{[\chi,\eta]}(g_c) = SF(g_{c_1}, g_{c_2}). \tag{8}$$

Then, since $f_{[\chi,\eta]}(g_c)$ is an element of the Hilbert space (4), one can consider the inner product

$$\begin{aligned}(f'_{[\chi]}, f_{[\chi,\eta]}) &= \int dg_c f'^{*}_{[\chi]}(g_c) f_{[\chi,\eta]}(g_c) \\ &= \int dg_c f'^{*}_{[\chi]}(g_c)[SF(g_{c_1}, g_{c_2})]. \end{aligned} \tag{9}$$

Up to this point, the functions $f(g_c)$ have been square integrable. Consider, however, as functions over cosets (which in general will not be square integrable),

$$D^{[\chi]}_{x'x}(g_c) = \langle [\chi]x' |\, U(g_c)\, | [\chi]x \rangle, \tag{10}$$

where $[\chi]$ denotes a set of irreducible representation labels and x', x are eigenvalues of a complete set of commuting observables. Choose $f'_{[\chi]}(g_c)$ to be $D^{[\chi]}_{x'x}(g_c)$ and $F(g_{c_1}, g_{c_2})$ to be $D^{[\chi_1]}_{x_1'x_1}(g_{c_1}) D^{[\chi_2]}_{x_2'x_2}(g_{c_2})$. Since $D^{[\chi]}_{x'x}(g_c)$ can be thought of as a concrete realization of $|[\chi], x\rangle$, the Clebsch–Gordan coefficients are proportional to the "overlap" between $D^{[\chi]}_{x'x}(g_c)$ and $S\{D^{[\chi_1]}_{x_1'x_1}(g_{c_1}) D^{[\chi_2]}_{x_2'x_2}(g_{c_2})\}$:

$$\begin{aligned}&\langle [\chi]x; \eta \mid [\chi_1]x_1; [\chi_2]x_2 \rangle \\ &\quad \times \alpha(D^{[\chi]}_{x'x}(g_c), S\{D^{[\chi_1]}_{x_1'x_1}(g_{c_1}) D^{[\chi_2]}_{x_2'x_2}(g_{c_2})\}) \\ &\quad = N \int dg_c D^{[\chi]*}_{x'x}(g_c) S D^{[\chi_1]}_{x_1'x_1}(g_{c_1}) D^{[\chi_2}_{x_2'x_2}(g_{c_2}), \end{aligned} \tag{11}$$

with N a normalization factor. Thus, if it is possible to find S and calculate $D^{[\chi]}_{x'x}(g_c)$, it should be possible to get the Clebsch–Gordan coefficients from Eq. (11).

To find the operator S we first decompose $F(g_{c_1}, g_{c_2})$ into a direct integral over double cosets. That is, we consider the outer product group[5] of G, the set of ordered pairs of elements $\{(g_1, g_2)\} = (G_1, G_2)$, with g_1 and g_2 arbitrary elements of G. This outer-product group can be decomposed into right cosets with respect to the H_1 and H_2 subgroups:

$$(G_1, G_2) = \bigcup_{c_1, c_2} (H_1, H_2)(g_{c_1}, g_{c_2}). \tag{12}$$

Now a subgroup of (G_1, G_2) is the "diagonal" group G itself, consisting of elements (g, g) with g an arbitrary element of G. We wish to write a double-coset decomposition[11] of (G_1, G_2) with respect to (H_1, H_2) and (G, G):

$$(G_1, G_2) = \bigcup_{D} (H_1, H_2)(e, g_D)(G, G), \tag{13}$$

where e is the identity element of G and g_D is the se of elements of G labeling the double cosets.

Mackey has shown that the induced representatio defined on the tensor-product space (6) is equivalen to induced representations defined on subspace labeled by the double cosets (13). The inducin subgroup is given by

$$\begin{aligned} H_D &= (e, g_D)^{-1}(H_1, H_2)(e, g_D) \cap (G, G) \\ &= (H_1, g_D^{-1} H_2 g_D) \cap (G, G) \\ &= H_1 \cap g_D^{-1} H_2 g_D \end{aligned} \tag{14}$$

and functions $f_D(G/H_D)$ carry the induced representations. Further, Mackey shows what the measure is or the double-coset labels so that

$$\|F\|^2 = \int d(D) \|f_D\|^2. \tag{15}$$

Usually the double-coset decomposition does not result in irreducible subspaces and it is necessary to decompose further the double-coset subspaces. The techniques for doing so will vary from group to group, but the end result will be to have a direct-integral decomposition of the kind given in Eq. (7).

II. THE CLEBSCH-GORDAN COEFFICIENTS OF THE TWO-DIMENSIONAL POINCARÉ GROUP

As a simple application of the technique outlined in the previous section, we will obtain the Clebsch–Gordan coefficients for the two-dimensional Poincaré group consisting of transformations

$$\begin{pmatrix} z' \\ t' \end{pmatrix} = \begin{pmatrix} \cosh\beta & \sinh\beta \\ \sinh\beta & \cosh\beta \end{pmatrix} \begin{pmatrix} z \\ t \end{pmatrix} + \begin{pmatrix} a_z \\ a_t \end{pmatrix}, \tag{16}$$

where $\binom{a_z}{a_t}$ are arbitrary space–time translations. Equation (16) can be rewritten as

$$\begin{pmatrix} z' \\ t' \\ 1 \end{pmatrix} = \begin{pmatrix} \cosh\beta & \sinh\beta & a_z \\ \sinh\beta & \cosh\beta & a_t \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} z \\ t \\ 1 \end{pmatrix} \tag{17}$$

and, finally, as a matrix group

$$P_2 = \left\{ \begin{pmatrix} \Lambda & a \\ 0 & 1 \end{pmatrix} \right\}, \quad \Lambda = \begin{pmatrix} \cosh\beta & \sinh\beta \\ \sinh\beta & \cosh\beta \end{pmatrix}, \quad a = \begin{pmatrix} a_z \\ a_t \end{pmatrix}. \tag{18}$$

All of the unitary irreducible representations of P_2 can be written as induced representations,[12] induced by the subgroup

$$H = \left\{ \begin{pmatrix} I & a \\ 0 & 1 \end{pmatrix} \right\} \tag{19}$$

[11] A double coset is a subset of a group G of the form $H_1 g H_2$, with H_1 and H_2 subgroups of G and g an element of G.

[12] See Ref. 5, p. 165ff.

(I is the identity matrix of Λ) having representations

$$\mathcal{K}(a) = e^{ip\cdot a} \tag{20}$$

defined on the one-dimensional vector space $\mathfrak{V}(p)$. $p \cdot a$ is the Lorentz scalar product $-p_z a_z + Ea_t$, with p_z and E having the usual interpretation of momentum and energy, respectively.

Mackey shows that all momentum vectors $p = \binom{p_z}{E}$ having the same mass generate equivalent representations of P_2 so that it is only necessary to consider a standard vector chosen to be

$$\hat{p} = \begin{pmatrix} 0 \\ M \end{pmatrix}; \tag{21}$$

that is, $\hat{p}$ is the "rest-frame" momentum vector.

The right-coset decomposition of P_2 with respect to H of Eq. (19) can be written as

$$\begin{pmatrix} \Lambda & a \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} I & a \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \Lambda & 0 \\ 0 & 1 \end{pmatrix}, \tag{22}$$

so that we consider functions in $\mathfrak{V}(p)$, Eq. (2), with the property that

$$\begin{aligned} f\begin{pmatrix} \Lambda & a \\ 0 & 1 \end{pmatrix} &= f\left[\begin{pmatrix} I & a \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \Lambda & 0 \\ 0 & 1 \end{pmatrix}\right] \\ &= e^{i\hat{p}\cdot a} f(\Lambda). \end{aligned} \tag{23}$$

The functions over cosets $f(\Lambda) \equiv f(\beta)$ have norm

$$\|f\|^2 = \int_{-\infty}^{+\infty} d\beta\, |f(\beta)|^2 < \infty. \tag{24}$$

The induced representation is

$$\begin{aligned} U(\Lambda_0, a_0) f(\Lambda) &= f\left[\begin{pmatrix} \Lambda & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \Lambda_0 & a_0 \\ 0 & 1 \end{pmatrix}\right] \\ &= f\begin{bmatrix} \Lambda\Lambda_0 & \Lambda a_0 \\ 0 & 1 \end{bmatrix} \\ &= f\left[\begin{pmatrix} I & \Lambda a_0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \Lambda\Lambda_0 & 0 \\ 0 & 1 \end{pmatrix}\right] \\ &= e^{i\hat{p}\cdot\Lambda a_0} f(\Lambda\Lambda_0) \end{aligned} \tag{25}$$

and it is easily seen that U is a unitary representation.

The tensor-product space consists of functions $F(\Lambda_1, \Lambda_2) \equiv F(\beta_1, \beta_2)$ having norm

$$\|F\|^2 = \int_{-\infty}^{+\infty} d\beta_1\, d\beta_2\, |F(\beta_1, \beta_2)|^2 < \infty; \tag{26}$$

the arguments of F are obtained from the right-coset decomposition of the outer-product group:

$$\begin{aligned} \left[\begin{pmatrix} \Lambda_1 & a_1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} \Lambda_2 & a_2 \\ 0 & 1 \end{pmatrix}\right] &= \left[\begin{pmatrix} I & a_1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} I & a_2 \\ 0 & 1 \end{pmatrix}\right] \\ &\times \left[\begin{pmatrix} \Lambda_1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} \Lambda_2 & 0 \\ 0 & 1 \end{pmatrix}\right] \end{aligned} \tag{27}$$

with the inducing subgroup

$$\left[\begin{pmatrix} I & a_1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} I & a_2 \\ 0 & 1 \end{pmatrix}\right]$$

having representations

$$\mathcal{K}_1 \otimes \mathcal{K}_2 = e^{i\hat{p}_1\cdot a_1} e^{i\hat{p}_2\cdot a_2} \tag{28}$$

on the vector space $\mathfrak{U}(\hat{p}_1) \otimes \mathfrak{U}(\hat{p}_2)$.

The induced *reducible* unitary representation on the tensor-product space is

$$\begin{aligned} &U(\Lambda_0, a_0) F(\Lambda_1, \Lambda_2) \\ &= F\left\{\left[\begin{pmatrix} \Lambda_1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} \Lambda_2 & 0 \\ 0 & 1 \end{pmatrix}\right]\left[\begin{pmatrix} \Lambda_0 & a_0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} \Lambda_0 & a_0 \\ 0 & 1 \end{pmatrix}\right]\right\} \\ &= F\left\{\left[\begin{pmatrix} I & \Lambda_1 a_0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} I & \Lambda_2 a_0 \\ 0 & 1 \end{pmatrix}\right]\left[\begin{pmatrix} \Lambda_1\Lambda_0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} \Lambda_2\Lambda_0 & 0 \\ 0 & 1 \end{pmatrix}\right]\right\} \\ &= e^{i\hat{p}_1\cdot\Lambda_1 a_0} e^{i\hat{p}_2\cdot\Lambda_2 a_0} F(\Lambda_1\Lambda_0, \Lambda_2\Lambda_0). \end{aligned} \tag{29}$$

Instead of the right-coset decomposition (27), it is possible to write a double-coset decomposition with respect to the diagonal group:

$$\begin{aligned} \left[\begin{pmatrix} \Lambda_1 & a_1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} \Lambda_2 & a_2 \\ 0 & 1 \end{pmatrix}\right] &= \left[\begin{pmatrix} I & a_1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} I & a_2 \\ 0 & 1 \end{pmatrix}\right] \\ &\times \left[\begin{pmatrix} I & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} \Lambda_D & 0 \\ 0 & 1 \end{pmatrix}\right]\left[\begin{pmatrix} \Lambda & a \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} \Lambda & a \\ 0 & 1 \end{pmatrix}\right], \end{aligned} \tag{30}$$

where

$$\Lambda_D = \begin{pmatrix} \cosh D & \sinh D \\ \sinh D & \cosh D \end{pmatrix}$$

is a convenient choice of double cosets.

Multiplying out Eq. (30) gives

$$\begin{aligned} \left[\begin{pmatrix} \Lambda_1 & a_1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} \Lambda_2 & a_2 \\ 0 & 1 \end{pmatrix}\right] &= \left[\begin{pmatrix} I & a_1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} I & a_2 \\ 0 & 1 \end{pmatrix}\right] \\ &\times \left[\begin{pmatrix} \Lambda & a \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} \Lambda_D\Lambda & \Lambda_D a \\ 0 & 1 \end{pmatrix}\right] \end{aligned} \tag{31}$$

so that, with the substitutions

$$\begin{aligned} \Lambda_1 &= \Lambda, \\ \Lambda_2 &= \Lambda_D\Lambda, \end{aligned} \tag{32}$$

it is possible to define a new function

$$f_D(\Lambda) = F(\Lambda_1, \Lambda_2). \tag{32'}$$

The fact that D is subscripted rather than being included in the argument of f is meant to indicate that it remains invariant under an arbitrary Poincaré

transformation. To see this, note that

$$\begin{aligned} U(\Lambda_0, a_0)f_D(\Lambda) &\equiv U(\Lambda_0, a_0)F(\Lambda_1, \Lambda_2) \\ &= e^{i\hat{p}_1\cdot\Lambda_1 a_0}e^{i\hat{p}_2\cdot\Lambda_2 a_0}F(\Lambda_1\Lambda_0, \Lambda_2\Lambda_0) \\ &= e^{i[\hat{p}_1\cdot\Lambda a_0+\hat{p}_2\cdot\Lambda_D\Lambda a_0]}F(\Lambda\Lambda_0, \Lambda_D\Lambda\Lambda_0) \\ &= e^{i[\hat{p}_1+\Lambda_D^{-1}\hat{p}_2]\cdot\Lambda a_0}f_D(\Lambda\Lambda_0). \end{aligned} \tag{33}$$

Defining

$$p_D = \hat{p}_1 + \Lambda_D^{-1}\hat{p}_2 = \begin{pmatrix} -M_2 \sinh D \\ M_1 + M_2 \cosh D \end{pmatrix} \tag{34}$$

gives

$$U(\Lambda_0, a_0)f_D(\Lambda) = e^{ip_D\cdot\Lambda a_0}f_D(\Lambda\Lambda_0), \tag{35}$$

which has exactly the same structure as Eq. (25) so that Eq. (35) gives a unitary irreducible representation of P_2 induced by the subgroup H of Eq. (19) with representation $e^{ip_D\cdot a}$ on $\mathfrak{U}(p_D)$.

Notice that

$$M^2 = p_D^2 = M_1^2 + M_2^2 + 2M_1M_2 \cosh D, \tag{36}$$

so that the subscript D actually gives the "mass" of the subspace $f_D(\Lambda)$ for arbitrary but fixed values of M_1 and M_2. In fact D labels the direct-integral decomposition of $F(\Lambda_1, \Lambda_2)$, since

$$\begin{aligned} \|F\|^2 &= \int_{-\infty}^{+\infty} d\beta_1\, d\beta_2\, |F(\beta_1\beta_2)|^2 \\ &= \int_{-\infty}^{+\infty} dD\, d\beta\, |f_D(\beta)|^2 \\ &= \int_{-\infty}^{+\infty} dD\, \|f_D\|^2, \end{aligned} \tag{37}$$

where

$$\|f_D\|^2 = \int_{-\infty}^{+\infty} d\beta\, |f_D(\beta)|^2.$$

Finally, to get the Clebsch–Gordan coefficients of P_2, it is necessary to know the functions $D^{[M]}_{p'p}(\Lambda)$ of Eq. (10). But it is well known that states $|p, E\rangle \equiv |[M], p\rangle$, with $M^2 = -p^2 + E^2$, transform under P_2 as

$$U(\Lambda_0, a_0)\,|[M], p\rangle = e^{i\Lambda_0 p\cdot a_0}\,|[M], \Lambda_0 p\rangle, \tag{38}$$

so that

$$\begin{aligned} D^{[M]}_{p'p}(\Lambda) &= \langle [M], p'|\, U(\Lambda, 0)\,|[M]p\rangle \\ &= \langle [M], \Lambda^{-1}p'\,|[M], p\rangle \\ &= E\delta(\Lambda^{-1}p' - p), \end{aligned} \tag{39}$$

where $E = (M^2 + p^2)^{\frac{1}{2}}$ is a standard normalization factor.[13]

We must check that $D^{[M]}_{p'p}(\Lambda)$ is indeed a function of Λ with the correct transformation properties under P_2, namely, those given in Eq. (25):

$$\begin{aligned} &U(\Lambda_0, a_0)D^{[M]}_{p'p}(\Lambda) \\ &\quad\equiv D^{[M]}_{p'p}\left[\begin{pmatrix}\Lambda & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}\Lambda_0 & a_0\\ 0 & 1\end{pmatrix}\right] \\ &\quad= \langle [M], p'|\, U(I, \Lambda a_0)U(\Lambda\Lambda_0, 0)\,|[M], p\rangle \\ &\quad= e^{ip'\cdot\Lambda a_0}D^{[M]}_{p'p}(\Lambda\Lambda_0). \end{aligned} \tag{40}$$

In order that (40) agree with (25), p' must be set equal to $\hat{p}$. But $D^{[M]}_{\hat{p}p}(\Lambda)$ can be thought of as a concrete realization of $|[M], p\rangle$, since $\hat{p}$ is now a fixed vector.[14]

Thus, in Eq. (9) and Eq. (10), we choose

$$\begin{aligned} f'_D(\Lambda) &= D^{[M]}_{p_Dp}(\Lambda), \\ F(\Lambda_1, \Lambda_2) &= D^{[M_1]}_{\hat{p}_1p_1}(\Lambda_1)D^{[M_2]}_{\hat{p}_2p_2}(\Lambda_2) \end{aligned} \tag{41}$$

and proceed to change variables in $F(\Lambda_1, \Lambda_2)$:

$$\begin{aligned} F(\Lambda_1, \Lambda_2) &= D^{[M_1]}_{\hat{p}_1p_1}(\Lambda_1)D^{[M_2]}_{\hat{p}_2p_2}(\Lambda_2) \\ &= D^{[M_1]}_{\hat{p}_1p_1}(\Lambda)D^{[M_2]}_{\hat{p}_2p_2}(\Lambda_D\Lambda) \\ &= D^{[M_1]}_{\hat{p}_1p_1}(\Lambda)D^{[M_2]}_{(p_D-\hat{p}_2)p_2}(\Lambda), \end{aligned} \tag{41'}$$

where $p_D = \hat{p}_1 + \Lambda_D^{-1}\hat{p}_2$ was defined in Eq. (34). Equation (41′) shows that

$$f_D(\Lambda) = D^{[M_1]}_{\hat{p}_1p_1}(\Lambda)D^{[M_2]}_{(p_D-\hat{p}_1)p_2}(\Lambda). \tag{42}$$

The Clebsch–Gordan coefficients are, according to Eq. (11),

$$\begin{aligned} &\langle [M]p \mid [M_1]p_1; [M_2]p_2\rangle \\ &\quad= N(D^{[M]}_{p_Dp}(\Lambda),\, D^{[M_1]}_{\hat{p}_1p_1}(\Lambda)D^{[M_2]}_{(\hat{p}_D-\hat{p}_1)p_2}(\Lambda)) \\ &\quad= N\int d\beta D^{[M]*}_{p_Dp}(\Lambda)D^{[M_1]}_{\hat{p}_1p_1}(\Lambda)D^{[M_2]}_{(p_D-\hat{p}_1)p_2}(\Lambda) \\ &\quad= N\int d\beta E\delta(\Lambda^{-1}p_D - p)E_1\delta(\Lambda^{-1}\hat{p}_1 - p_1) \\ &\qquad\times E_2\delta(\Lambda^{-1}(p_D - \hat{p}_1) - p_2). \end{aligned} \tag{43}$$

We wish to show that the δ functions in (43) can be written as

$$\delta(p - p_1 - p_2)\delta\left(\cosh D - \frac{p_1}{M_1}\cdot\frac{p_2}{M_2}\right).$$

If this is the case, then in the second δ function we will have the usual expression for the mass $M^2 = (p_1 + p_2)^2$, since, according to Eq. (36),

$$\begin{aligned} M &= M_1^2 + M_2^2 + 2M_1M_2 \cosh D \\ &= M_1^2 + M_2^2 + 2M_1M_2\, p_1p_2/M_1M_2 \\ &= (p_1 + p_2)^2. \end{aligned} \tag{44}$$

[13] See, for example, J. Werle, *Relativistic Theory of Reactions* (John Wiley & Sons, Inc., New York, 1966).

[14] It is not difficult to see that $D^{[M]}_{\hat{p}p}(\Lambda)$ is a concrete realization of $|[M]p\rangle$, for $D^{[M]}_{\hat{p}p}(\Lambda)$ transforms in the same way that $|[M]p\rangle$ does and further has the same normalization as $\langle [M]p' \mid [M]p\rangle$, so that

$$\langle [M]p' \mid [M]p\rangle = \int d\beta D^{[M]*}_{\hat{p}p'}(\Lambda)D^{[M]}_{\hat{p}p}(\Lambda)$$

(up to a possible factor involving M).

Now,

$$\begin{aligned}
&\int d\beta\delta(\Lambda^{-1}p_D - p)\delta(\Lambda^{-1}\hat{p}_1 - p_1)\delta(\Lambda^{-1}(p_D - \hat{p}_1) - p_2)\\
&= \delta(p - p_1 - p_2)\int d\beta\delta(M_1\sinh\beta + p_1)\\
&\quad\times \delta(M_2\sinh D\cosh\beta\\
&\quad + (M_1 + M_2\cosh D)\sinh\beta - M_1\sinh\beta + p_2)\\
&= \delta(p - p_1 - p_2)\delta(\sinh(D+\beta) - \sinh\sinh^{-1}p_2/M_2)\\
&= \delta(p - p_1 - p_2)\delta(D - (\sinh^{-1}p_2/M_2 - \beta))\\
&= \delta(p - p_1 - p_2)\\
&\quad\times \delta(\cosh D - (E_1E_2/M_1M_2 - p_1p_2/M_1M_2))\\
&= \delta(p - p_1 - p_2)\delta(M^2 - (p_1 + p_2)^2); \qquad (45)
\end{aligned}$$

mass and energy terms preceding the δ functions have been dropped in these equations. Also, $p = \binom{p}{E}$ so that p means both the vector and the space component of the vector; however, the usage should be clear from the context.

Finally, then, the Clebsch–Gordan coefficients are

$$\begin{aligned}
\langle[M]p \mid [M_1]p_1;\,[M_2]p_2\rangle &= N\delta(p - p_1 - p_2)\\
&\quad\times \delta(M^2 - (p_1 + p_2)^2) \qquad (46)
\end{aligned}$$

and gives, as would be expected, conservation of energy and momentum along one space axis.

III. REPRESENTATIONS AND FUNCTIONS OVER COSETS FOR THE POINCARÉ GROUP

The way in which one obtains the irreducible unitary representations of the Poincaré group using Mackey's theory of induced representations has been given by Moussa and Stora.[3] In this section, the inducing subgroups and their representations will be written in such a way as easily to obtain the functions over cosets needed to obtain the Clebsch–Gordan coefficients in Sec. IV.

Usually one writes the Poincaré group as

$$x' = \Lambda x + a, \qquad (47)$$

where x' and x are space–time vectors, Λ is a Lorentz transformation, and a is a space–time translation. Equation (47) can be rewritten as

$$\begin{pmatrix} x' \\ 1 \end{pmatrix} = \begin{pmatrix} \Lambda & a \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ 1 \end{pmatrix}$$

and thus as the matrix

$$P \equiv \left\{\begin{pmatrix} \Lambda & a \\ 0 & 1 \end{pmatrix}\right\}. \qquad (48)$$

However, in this paper we will be dealing almost exclusively with the covering group of P, denoted by $\bar{P}$. Now, it is well known[9] that the covering group of $\{\Lambda\}$ is $SL(2, C)$, the set of 2×2 complex unimodular matrices. Further, the space–time translation a can be written as a 2×2 Hermitian matrix $H(a)$ with

$$H(a) = \begin{pmatrix} a_t + a_z & a_x - ia_y \\ a_x + ia_y & a_t - a_z \end{pmatrix}, \qquad (49)$$

where $(a_x a_y a_z a_t)$ are the components of a. Under a Lorentz transformation which sends a into $a' = \Lambda a$, $H(a)$ is sent into

$$H(a') = \bar{\Lambda}H(a)\bar{\Lambda}^{+}, \qquad (50)$$

where $\bar{\Lambda}$ is that element of $SL(2, C)$ which corresponds to Λ.[9] Further, the length of a, i.e., a^2, is readily seen from Eq. (49) to be $|H(a)|$.

The question is then: what does $\bar{P}$ look like as a matrix? It is not hard to see that a suitable choice is

$$\bar{P} \equiv \left\{\begin{pmatrix} \bar{\Lambda} & H(a)\bar{\Lambda}^{-1+} \\ 0 & \bar{\Lambda}^{-1+} \end{pmatrix}\right\}, \qquad (51)$$

where $\bar{\Lambda}^{-1}$ is the inverse matrix and $\bar{\Lambda}^{+}$ the adjoint matrix of $\bar{\Lambda}$. The matrix given in (51) combines elements in the same way that P in Eq. (48) does. Thus, if

$$\begin{aligned}
\begin{pmatrix} \Lambda_1 & a_1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \Lambda_2 & a_2 \\ 0 & 1 \end{pmatrix} &= \begin{pmatrix} \Lambda_1\Lambda_2 & \Lambda_1 a_2 + a_1 \\ 0 & 1 \end{pmatrix}\\
&= \begin{pmatrix} \Lambda_3 & a_3 \\ 0 & 1 \end{pmatrix}, \qquad (52)
\end{aligned}$$

then

$$\begin{aligned}
&\begin{pmatrix} \bar{\Lambda}_1, & H(a_1)\bar{\Lambda}_1^{-1+} \\ 0 & \bar{\Lambda}_1^{-1+} \end{pmatrix}\begin{pmatrix} \bar{\Lambda}_2, & H(a_2)\bar{\Lambda}_2^{-1+} \\ 0 & \bar{\Lambda}_2^{-1+} \end{pmatrix}\\
&\quad= \begin{pmatrix} \bar{\Lambda}_1\bar{\Lambda}_2, & \bar{\Lambda}_1H(a_2)\bar{\Lambda}_2^{-1+} + H(a_1)\bar{\Lambda}_1^{-1+}\bar{\Lambda}_2^{-1+} \\ 0 & [\bar{\Lambda}_1\bar{\Lambda}_2]^{-1+} \end{pmatrix}\\
&\quad= \begin{pmatrix} \bar{\Lambda}_1\bar{\Lambda}_2, & [\bar{\Lambda}_1H(a_2)\bar{\Lambda}_1^{+} + H(a_1)][\bar{\Lambda}_1\bar{\Lambda}_2]^{-1+} \\ 0 & [\bar{\Lambda}_1\bar{\Lambda}_2]^{-1+} \end{pmatrix}. \qquad (53)
\end{aligned}$$

But, according to (50), $\bar{\Lambda}_1H(a_2)\bar{\Lambda}_1^{+}$ corresponds to $\Lambda_1 a_2$ so that $\bar{\Lambda}_1H(a_2)\bar{\Lambda}^{+} + H(a_1)$ corresponds to $\Lambda_1 a_2 + a_1$ as in (52).

In order to write out the induced representations of $\bar{P}$, it is first necessary to find the inducing subgroups. This is done with the help of little-group theory,[15] which says that the inducing subgroups which generate irreducible representations of $\bar{P}$ consist of all translations $\binom{I\ H(a)}{0\ I}$ plus those Lorentz transformations satisfying $\Lambda p = p$ or, what is equivalent,

$$\bar{\Lambda}H(p)\bar{\Lambda}^{+} = H(p). \qquad (54)$$

p is a "momentum–energy" vector and arises as an irreducible representation label $e^{ip\cdot a}$ of the translations mentioned above ($p \cdot a = -\mathbf{p}\cdot\mathbf{a} + Ea_t$).

[15] E. P. Wigner, Ann. Math. **40**, 149 (1939); J. S. Lomont, *Applications of Finite Groups* (Academic Press Inc., New York, 1959).

It is well known[15] that there are essentially three classes of momentum–energy vectors p satisfying (54):

(A) $p^2 = p \cdot p > 0$ or $|H(p)| > 0$; then a convenient choice of p is $\hat{p} = \begin{pmatrix} 0 \\ M \end{pmatrix}$, which corresponds to $H(\hat{p}) = M\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = MI$. Substituting into (54) gives

$$\bar{\Lambda}(MI)\bar{\Lambda}^{+} = MI, \quad \bar{\Lambda}\bar{\Lambda}^{+} = I, \tag{55}$$

which is the definition of $SU(2)$. All other vectors p satisfying $p^2 > 0$ generate equivalent representations of $\bar{P}$.

The inducing subgroup for $p^2 > 0$ is thus

$$H = \begin{pmatrix} SU(2) & \{H(a)\}SU(2) \\ 0 & SU(2) \end{pmatrix}, \tag{56}$$

which has representations

$$\mathcal{H}_1 = e^{i\hat{p}\cdot a} D^{[S]}_{j'j}(R), \quad R \in SU(2), \tag{57}$$

where $D^{[S]}_{j'j}(R)$ are the irreducible representation matrices of $SU(2)$ (Wigner functions).

(B) $p^2 = 0$; choose

$$\hat{p} = \begin{pmatrix} 0 \\ 0 \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$$

as a standard momentum–energy vector for this class. Then, $H(\hat{p}) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $\bar{\Lambda}H(\hat{p})\bar{\Lambda}^{+} = H(\hat{p})$, since

$$\bar{\Lambda} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, \quad \alpha\delta - \beta\gamma = 1,$$

give

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \alpha^* & \gamma^* \\ \beta^* & \delta^* \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

$$\begin{pmatrix} |\alpha|^2 & \alpha\gamma^* \\ (\alpha\gamma^*)^* & |\delta|^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \tag{58}$$

so that $\gamma = 0$, $\alpha = 1$, and thus the set of elements leaving $H(\hat{p})$ invariant is of the form $\begin{pmatrix} \alpha/|\alpha| & \beta \\ 0 & \alpha/|\alpha|^* \end{pmatrix}$, which can also be written as

$$\bar{E}_2 = \begin{pmatrix} e^{i\varphi/2} & \tilde{\beta}e^{-i\varphi/2} \\ 0 & e^{-i\varphi/2} \end{pmatrix} \tag{59}$$

with $\varphi/2 = \alpha/|\alpha|$ and $\beta = \tilde{\beta}e^{-i\varphi/2}$. In this form it is not hard to see that $\bar{E}_2$ is the covering group of the two-dimensional Euclidean group, that is, the group of transformations

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \beta_x \\ \beta_y \end{pmatrix},$$

with

$$\tilde{\beta} = \beta_x + i\beta_y.$$

The inducing subgroup for $\hat{p}$-type representations is

$$H = \begin{pmatrix} \bar{E}_2 & \{H(\hat{p})\}\bar{E}_2^{-1+} \\ 0 & \bar{E}_2^{-1+} \end{pmatrix}.$$

The representations of $\bar{E}_2$ are themselves induced representations, but, in this paper only, the positive-mass representations will be considered in regard to the tensor-product decomposition.

(C) $p^2 < 0$; choose

$$\hat{p} = \begin{pmatrix} 0 \\ 0 \\ M \\ 0 \end{pmatrix};$$

then

$$H(\hat{p}) = M\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

and

$$\bar{\Lambda}M\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\bar{\Lambda}^{+} = M\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

shows that the set of Lorentz transformations leaving (54) invariant is $SU(1, 1)$.

Returning to the positive-mass representations in order to define induced representations on functions over cosets, it is necessary to decompose $\bar{P}$ with respect to the inducing subgroup H of Eq. (56). This is equivalent to decomposing the Lorentz group $SL(2, C)$ into right cosets with respect to $SU(2)$:

$$SL(2, C) = \bigcup_c SU(2)\Lambda_c. \tag{60}$$

The set $\{\Lambda_c\}$ are elements of $SL(2, C)$ labeling right cosets; they can be chosen in many different ways, for example, as rotationless Lorentz transformations[7]

$$\Lambda_c = \Lambda_c^{+}. \tag{61}$$

Once the right-coset labels are chosen, it is possible to write

$$\bar{P} = \bigcup_c \begin{pmatrix} SU(2) & \{H(a)\}SU(2) \\ 0 & SU(2) \end{pmatrix}\begin{pmatrix} \Lambda_c & 0 \\ 0 & \Lambda_c^{-1+} \end{pmatrix}. \tag{62}$$

In the following discussion, the "bar" notation for the covering groups will be omitted and, in fact,

$$\begin{pmatrix} \Lambda & a \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \Lambda & H(a)\Lambda^{-1+} \\ 0 & \Lambda^{-1+} \end{pmatrix}$$

will denote the same element unless there is a possibility of ambiguity. The functions over right cosets are $f_j(\Lambda_c)$ with norm

$$\|f_j\|^2 = \sum_{j=-S}^{+S} \int d\Lambda_c \, |f_j(\Lambda_c)|^2 < \infty. \tag{63}$$

The summation over j comes from the fact that the representations of H are themselves of dimension $2S+1$ (see Refs. 7 and 8). The induced unitary irreducible positive-mass representations are

$$\begin{aligned} U(\Lambda_0, a_0) f_j(\Lambda_c) &= f_j\left[\begin{pmatrix}\Lambda_c & 0\\ 0 & \Lambda_c^{-1+}\end{pmatrix}\begin{pmatrix}\Lambda_0 & H(a_0)\Lambda_0^{-1+}\\ 0 & \Lambda_0^{-1+}\end{pmatrix}\right] \\ &= f_j\left[\begin{pmatrix}R & \Lambda_c H(a_0)\Lambda_c^{+}R\\ 0 & R\end{pmatrix}\begin{pmatrix}\Lambda_{c'} & 0\\ 0 & \Lambda_{c'}^{-1+}\end{pmatrix}\right] \\ &= e^{i\hat{p}\cdot\Lambda_c a_0}\sum_{j'=-S}^{+S} D^{[S]}_{jj'}(R) f_{j'}(\Lambda_{c'}) \end{aligned} \tag{64}$$

with $R \in SU(2)$ and $\Lambda_{c'}$ defined by $\Lambda_c\Lambda_0 = R\Lambda_{c'}$.

Finally, in this section, we calculate the functions over cosets which will be used in the calculation of the Clebsch–Gordan coefficients. We must calculate

$$D^{[MS]}_{p'jp\sigma}(\Lambda_c) = \langle [MS]\mathbf{p}'j|\, U(\Lambda_c)\, |[MS]\mathbf{p}, \sigma\rangle, \tag{65}$$

where $|[MS]\mathbf{p}, \sigma\rangle$ is a nonnormalizable basis element having mass M, spin S, momentum $\mathbf{p}$, and "spin component" σ. Now, the transformation properties of $|[M, S]\mathbf{p}\sigma\rangle$ are[16]

$$\begin{aligned} &U(\Lambda_0, a_0)\, |[M, S]\mathbf{p}\sigma\rangle \\ &\quad = e^{i\Lambda_0 p\cdot a_0}\sum_{\sigma'=-S}^{+S} D^{[S]}_{\sigma'\sigma}(\mathbf{p}, \Lambda_0)\, |[M, S]\Lambda_0 p, \sigma'\rangle, \end{aligned} \tag{66}$$

where $(\mathbf{p}, \Lambda_0)$ is a "Wigner rotation" which must be carefully defined.

Generally one defines $(\mathbf{p}, \Lambda_0)$ as the rotation $O^{-1}(\Lambda_0\mathbf{p})\Lambda_0 O(\mathbf{p})$, where $O(\mathbf{p})$ is a Lorentz transformation from the rest frame of the particle to the frame where it has momentum $\mathbf{p}$. There are many ways to choosing $O(\mathbf{p})$, but all choices are equivalent to a choice of coset representatives. However, because of the way in which Mackey has defined an induced representation $U(g_0)f(g) = f(gg_0)$ with g_0 acting to the right of g, the connection between $O(\mathbf{p})$ and Λ_c is

$$O(\mathbf{p}) = \Lambda_c^{-1}(\mathbf{p}), \tag{67}$$

where $\Lambda_c(\mathbf{p})$ is that right-coset element which carries the particle from the frame, where it has momentum p, to its rest frame, where it has momentum $\hat{p}$. This means that in covering-group form

$$\begin{aligned} H(p) &= \Lambda_c^{-1}(p)H(\hat{p})\Lambda_c^{-1+}(p) \\ &= M\Lambda_c^{-1}(p)\Lambda_c^{-1+}(p), \\ (p, \Lambda_0) &= \Lambda_c(\Lambda_0 p)\Lambda_0\Lambda_c^{-1}(p). \end{aligned} \tag{68}$$

[16] See, for example, Refs. 3 or 13.

From Eq. (66) we get that

$$\begin{aligned} &D^{[MS]}_{p'jp\sigma}(\Lambda_0, a_0) \\ &\quad = \langle [M, S]\mathbf{p}'j|\, e^{i\Lambda_0 p\cdot a_0}\sum_{\sigma'=-S}^{+S} D^{[S]}_{\sigma'\sigma}(p, \Lambda_0)\, |[M, S]\Lambda_0\mathbf{p}, \sigma'\rangle \\ &\quad = E' e^{i\Lambda_0 p\cdot a_0} D^{[S]}_{j\sigma}(p, \Lambda_0)\delta^3(\mathbf{p}' - \Lambda_0\mathbf{p}). \end{aligned} \tag{69}$$

Then $f_j(\Lambda_c)$ is chosen to be $D^{[M,S]}_{\hat{p}jp\sigma}(\Lambda_c)$ and we must check whether $D^{[M,S]}_{\hat{p}jp\sigma}(\Lambda_c)$ transforms properly under an arbitrary Poincaré transformation:

$$\begin{aligned} &U(\Lambda_0, a_0) D^{[M,S]}_{\hat{p}jp\sigma}(\Lambda_c) \\ &\quad = \langle [M, S]\hat{p}j|\, U(\Lambda_c, 0)U(\Lambda_0, a_0)\, |[M, S]p\sigma\rangle \\ &\quad = \langle [M, S]\hat{p}j|\, U(R, \Lambda_c a_0)U(\Lambda_{c'})\, |[M, S]p\sigma\rangle \\ &\quad = e^{i\hat{p}\cdot\Lambda_c a_0}\sum_{j'=-S}^{+S} D^{[S]}_{jj'}(\hat{p}, R) D^{[MS]}_{\hat{p}j'p\sigma}(\Lambda_{c'}), \end{aligned} \tag{70}$$

where $\Lambda_c\Lambda_0 = R\Lambda_{c'}$ as in Eq. (64) and $D^{[S]}_{jj'}(\hat{p}, R) = D^{[S]}_{jj'}(R)$ since $(\hat{p}, R) = \Lambda_c(R\hat{p})R\Lambda_c^{-1}(\hat{p}) = R$. But Eq. (70) has the same form as Eq. (64) so that $D^{[M,S]}_{\hat{p}jp\sigma}(\Lambda_c)$ is a (non-square-integrable) coset function having the correct transformation properties.

IV. TENSOR PRODUCTS OF POSITIVE-MASS REPRESENTATIONS OF THE POINCARÉ GROUP; EXPLICIT FORM OF THE CLEBSCH-GORDAN COEFFICIENTS

The tensor-product space of positive-mass representations of the Poincaré group consists of square-integrable functions $f_{i_1 i_2}(\Lambda_{c_1}\Lambda_{c_2})$ with norm

$$\|f_{i_1 i_2}\|^2 = \sum_{i_1=-S_1}^{+S_1}\sum_{i_2=-S_2}^{+S_2}\int d\Lambda_{c_1}\, d\Lambda_{c_2}\, |f_{i_1 i_2}(\Lambda_{c_1}, \Lambda_{c_2})|^2 < \infty. \tag{71}$$

The choice of cosets comes about from the decomposition of the outer-product group (P_1, P_2):

$$\begin{aligned} &\left[\left\{\begin{pmatrix}\Lambda_1 & H(a_1)\Lambda_1^{-1+}\\ 0 & \Lambda_1^{-1+}\end{pmatrix}\right\}, \left\{\begin{pmatrix}\Lambda_2 & H(a_2)\Lambda_2^{-1+}\\ 0 & \Lambda_2^{-1+}\end{pmatrix}\right\}\right] \\ &\quad = \bigcup_{c_1 c_2}\left[\begin{pmatrix}SU(2)_1 & \{H(a_1)\}SU(2)_1\\ 0 & SU(2)_1\end{pmatrix}, \begin{pmatrix}SU(2)_2 & \{H(a_2)\}SU(2)_2\\ 0 & SU(2)_2\end{pmatrix}\right] \\ &\qquad \times\left[\begin{pmatrix}\Lambda_{c_1} & 0\\ 0 & \Lambda_{c_1}^{-1+}\end{pmatrix}, \begin{pmatrix}\Lambda_{c_2} & 0\\ 0 & \Lambda_{c_2}^{-1+}\end{pmatrix}\right], \end{aligned} \tag{72}$$

where Λ_{c_1} and Λ_{c_2} are both of the form given in Eq. (62). The representations of the inducing subgroup

$$\left[\begin{pmatrix}SU(2)_1 & \{H(a_1)\}SU(2)_1\\ 0 & SU(2)_1\end{pmatrix}, \begin{pmatrix}SU(2)_2 & \{H(a_2)\}SU(2)_2\\ 0 & SU(2)_2\end{pmatrix}\right]$$

are

$$e^{i\hat{p}_1\cdot a_1} D^{[S_1]}_{i_1 i_1'}(R_1) e^{i\hat{p}_2\cdot a_2} D^{[S_2]}_{i_2 i_2'}(R_2),$$

so that the induced representation on the tensor-product space is

$$
\begin{aligned}
&U(\Lambda_0, a_0) f_{i_1 i_2}(\Lambda_{c_1}, \Lambda_{c_2}) \\
&= f_{i_1 i_2}\left\{\left[\begin{pmatrix} \Lambda_{c_1} & 0 \\ 0 & \Lambda_{c_1}^{-1+} \end{pmatrix}, \begin{pmatrix} \Lambda_{c_2} & 0 \\ 0 & \Lambda_{c_2}^{-1+} \end{pmatrix}\right]\right. \\
&\quad \left.\times \left[\begin{pmatrix} \Lambda_0 & H(a_0)\Lambda_0^{-1+} \\ 0 & \Lambda_0^{-1+} \end{pmatrix}, \begin{pmatrix} \Lambda_0 & H(a_0)\Lambda_0^{-1+} \\ 0 & \Lambda_0^{-1+} \end{pmatrix}\right]\right\} \\
&= f_{i_1 i_2}\left\{\left[\begin{pmatrix} R_1 & \Lambda_{c_1} H(a_0)\Lambda_{c_1}^{+} R_1 \\ 0 & R_1 \end{pmatrix}, \begin{pmatrix} R_2 & \Lambda_{c_2} H(a_0)\Lambda_{c_2}^{+} R_2 \\ 0 & R_2 \end{pmatrix}\right]\right. \\
&\quad \left.\times \left[\begin{pmatrix} \Lambda_{c_1'} & 0 \\ 0 & \Lambda_{c_1'}^{-1+} \end{pmatrix}, \begin{pmatrix} \Lambda_{c_2'} & 0 \\ 0 & \Lambda_{c_2'}^{-1+} \end{pmatrix}\right]\right\} \\
&= e^{i\hat{p}_1 \cdot \Lambda_{c_1} a_0} e^{i\hat{p}_2 \cdot \Lambda_{c_2} a_0} \sum_{i_2'=-S_2}^{+S_2} \sum_{i_1'=-S_1}^{+S_1} D_{i_1 i_1'}^{[S_1]}(R_1) \\
&\quad \times D_{i_2 i_2'}^{[S_2]}(R_2) f_{i_1' i_2'}(\Lambda_{c_1'}, \Lambda_{c_2'}),
\end{aligned} \tag{73}
$$

where $\Lambda_{c_1}\Lambda_0 = R_1\Lambda_{c_1}$, and $\Lambda_{c_2}\Lambda_0 = R_2\Lambda_{c_2'}$ as before.

It is also possible to decompose (P_1, P_2) into double cosets with respect to the induced subgroups given in Eq. (72) and the diagonal subgroup

$$
\begin{aligned}
(P, P) = &\left\{\left[\begin{pmatrix} \Lambda & H(a)\Lambda^{-1+} \\ 0 & \Lambda^{-1+} \end{pmatrix}, \begin{pmatrix} \Lambda & H(a)\Lambda^{-1+} \\ 0 & \Lambda^{-1+} \end{pmatrix}\right]\right\}: \\
&\left[\begin{pmatrix} \Lambda_1 & H(a_1)\Lambda_1^{-1+} \\ 0 & \Lambda_1^{-1+} \end{pmatrix}, \begin{pmatrix} \Lambda_2 & H(a_2)\Lambda_2^{-1+} \\ 0 & \Lambda_2^{-1+} \end{pmatrix}\right] \\
&\quad = \left[\begin{pmatrix} R_1 & H(a_1)R_1 \\ 0 & R_1 \end{pmatrix}, \begin{pmatrix} R_2 & H(a_2)R_2 \\ 0 & R_2 \end{pmatrix}\right] \\
&\qquad \times \left[\begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}\begin{pmatrix} \Lambda_D & 0 \\ 0 & \Lambda_D^{-1+} \end{pmatrix}\right] \\
&\qquad \times \left[\begin{pmatrix} \Lambda & H(a)\Lambda^{-1+} \\ 0 & \Lambda^{-1+} \end{pmatrix}\begin{pmatrix} \Lambda & H(a)\Lambda^{-1+} \\ 0 & \Lambda^{-1+} \end{pmatrix}\right].
\end{aligned} \tag{74}
$$

Here Λ_D is a Lorentz transformation labeling the double coset; a convenient choice is

$$
\Lambda_D = \begin{pmatrix} D & 0 \\ 0 & 1/D \end{pmatrix}
$$

with D real. In the form given Λ_D consists of pure Lorentz transformations along the "z" axis.

Once the form of Λ_D has been chosen, it is necessary to find the group H_D of Eq. (14) which is the subgroup that induces representations in the subspace of the tensor-product space labeled by the double coset D:

$$
\begin{aligned}
H_D &= \begin{pmatrix} SU(2)_1 & \{H(a_1)\}SU(2)_1 \\ 0 & SU(2)_1 \end{pmatrix} \cap \begin{pmatrix} \Lambda_D & 0 \\ 0 & \Lambda_D^{-1+} \end{pmatrix}^{-1} \\
&\quad \times \begin{pmatrix} SU(2)_2 & \{H(a_2)\}SU(2)_2 \\ 0 & SU(2)_2 \end{pmatrix}\begin{pmatrix} \Lambda_D & 0 \\ 0 & \Lambda_D^{-1+} \end{pmatrix} \\
&= \begin{pmatrix} U(1) & \{H(a)\}U(1) \\ 0 & U(1) \end{pmatrix},
\end{aligned} \tag{75}
$$

since

$$
\begin{aligned}
&\Lambda_D^{-1} R_2 \Lambda_D \cap R_1 \\
&\quad = \begin{pmatrix} 1/D & 0 \\ 0 & D \end{pmatrix}\begin{pmatrix} u_2 & v_2 \\ -v_2^* & u_2^* \end{pmatrix}\begin{pmatrix} D & 0 \\ 0 & 1/D \end{pmatrix} \cap \begin{pmatrix} u_1 & v_1 \\ -v_1^* & u_1^* \end{pmatrix} \\
&\quad = \begin{pmatrix} u_1 & 0 \\ 0 & u_1^* \end{pmatrix} \in U(1), \quad |u_1|^2 = 1.
\end{aligned}
$$

H_D has representations inherited from H_1 and $\Lambda_D^{-1} H_2 \Lambda_D$, namely,

$$
e^{i\hat{p}_1 \cdot a} e^{i\hat{p}_2 \cdot a} D_{i_1 i_1'}^{[S_1]}(u) D_{i_2 i_2'}^{[S_2]}(u),
$$

with u an element of $U(1)$.

In the subspaces labeled by D the representations of P are induced by H_D of Eq. (75). The induced representations are defined on functions $F_{(i_1 i_2 D)}(\Lambda/U(1))$, where

$$
\frac{\Lambda}{U(1)} = \begin{pmatrix} \alpha & \beta \\ \gamma & |\delta| \end{pmatrix}
$$

is a right-coset decomposition of $SL(2, C)$ in which

$$
\Lambda = u \frac{\Lambda}{U(1)}.
$$

Then

$$
F_{(i_1 i_2 D)}\left(\frac{\Lambda}{U(1)}\right)
$$

is defined as

$$
\begin{aligned}
F_{(i_1 i_2 D)}\left(\frac{\Lambda}{U(1)}\right) &\equiv F_{(i_1 i_2)}\left(\frac{\Lambda}{U(1)}, \Lambda_D \frac{\Lambda}{U(1)}\right) \\
&= f_{i_1 i_2}(\Lambda_{c_1}, \Lambda_{c_2}).
\end{aligned} \tag{76}
$$

In order to see how $F_{(i_1 i_2 D)}(\Lambda/U(1))$ behaves under arbitrary Poincaré transformations—and also to see that D ought to be subscripted, indicating that it remains unchanged under a Poincaré transformation—it is necessary to see how $(\Lambda_{c_1}, \Lambda_{c_2})$ is related to $\Lambda/U(1)$ and Λ_D.

Equation (74) enables one to find in which double coset an arbitrary element of (P_1, P_2) is; to make (74) into an equation specifying all the elements on the right-hand side uniquely once the element of (P_1, P_2) is given, it is neccessary to divide out the diagonal subgroup elements by H_D of Eq. (75). Then

$$
\begin{aligned}
&\left[\begin{pmatrix} \Lambda_1 & H(a_1)\Lambda_1^{-1+} \\ 0 & \Lambda_1^{-1+} \end{pmatrix}\begin{pmatrix} \Lambda_2 & H(a_2)\Lambda_2^{-1+} \\ 0 & \Lambda_2^{-1+} \end{pmatrix}\right] \\
&\quad = \left[\begin{pmatrix} R_1 & H(a_1)R_1 \\ 0 & R_1 \end{pmatrix},\right. \\
&\qquad \left.\times \begin{pmatrix} R_2 & H(a_2)R_2 \\ 0 & R_2 \end{pmatrix}\right]\left[\begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}\begin{pmatrix} \Lambda_D & 0 \\ 0 & \Lambda_D^{-1+} \end{pmatrix}\right] \\
&\qquad \times \left[\begin{pmatrix} \Lambda/U(1) & 0 \\ 0 & [\Lambda/U(1)]^{-1+} \end{pmatrix}\right. \\
&\qquad \left.\times \begin{pmatrix} \Lambda/U(1) & 0 \\ 0 & [\Lambda/U(1)]^{-1+} \end{pmatrix}\right].
\end{aligned} \tag{77}
$$

In particular, the Lorentz transformations of (P_1, P_2) can be written as

$$\left[\begin{pmatrix}\alpha_1 & \beta_1\\ \gamma_1 & \delta_1\end{pmatrix},\begin{pmatrix}\alpha_2 & \beta_2\\ \gamma_2 & \delta_2\end{pmatrix}\right]=\left[\begin{pmatrix}u_1 & v_1\\ -v_1^* & u_1^*\end{pmatrix},\begin{pmatrix}u_2 & v_2\\ -v_2^* & u_2^*\end{pmatrix}\right]\times\left[\begin{pmatrix}\alpha & \beta\\ \gamma & |\delta|\end{pmatrix},\begin{pmatrix}D\alpha & D\beta\\ \gamma/D & |\delta|/D\end{pmatrix}\right] \quad (78)$$

so that for given (Λ_1, Λ_2), the values of u_1, v_1, u_2, v_2, $\alpha, \beta, |\delta|$, and D are all uniquely fixed. Thus, one can choose the element $(\Lambda_{c_1}, \Lambda_{c_2})$ and find how R_1, R_2, Λ_D, and $\Lambda/U(1)$ depend on Λ_{c_1} and Λ_{c_2} so that $F_{(i_1i_2D)}(\Lambda/U(1))$ in Eq. (76) is uniquely defined. $F_{(i_1i_2D)}(\Lambda/U(1))$ behaves, under an arbitrary Poincaré transformation, as

$$\begin{aligned}
&U(\Lambda_0, a_0)F_{(i_1i_2D)}(\Lambda/U(1))\\
&\quad=F_{(i_1i_2)}\Bigg\{\Bigg[\begin{pmatrix}\Lambda/U(1) & 0\\ 0 & [\Lambda/U(1)]^{-1+}\end{pmatrix},\\
&\qquad\qquad\begin{pmatrix}\Lambda_D(\Lambda/U(1)) & 0\\ 0 & [\Lambda_D(\Lambda/U(1))]^{-1+}\end{pmatrix}\Bigg]\\
&\qquad\times\left[\begin{pmatrix}\Lambda_0 & H(a_0)\Lambda_0^{-1+}\\ 0 & \Lambda_0^{-1+}\end{pmatrix}\begin{pmatrix}\Lambda_0 & H(a_0)\Lambda_0^{-1+}\\ 0 & \Lambda_0^{-1+}\end{pmatrix}\right]\Bigg\}\\
&\quad=F_{(i_1i_2)}\Bigg\{\Bigg[\begin{pmatrix}u & [\Lambda/U(1)]H(a_0)[\Lambda/U(1)]^{+}u\\ 0 & u\end{pmatrix}\\
&\qquad\times\begin{pmatrix}u & \Lambda_D(\Lambda/U(1))H(a_0)[\Lambda_D(\Lambda/U(1))]^{+}u\\ 0 & u\end{pmatrix}\Bigg]\\
&\qquad\times\Bigg[\begin{pmatrix}\Lambda'/U(1) & 0\\ 0 & [\Lambda'/U(1)]^{-1+}\end{pmatrix}\\
&\qquad\times\begin{pmatrix}\Lambda_D(\Lambda'/U(1)) & 0\\ 0 & [\Lambda_D(\Lambda'/U(1))]^{-1+}\end{pmatrix}\Bigg]\Bigg\}\\
&\quad=e^{i\hat{p}_1\cdot[\Lambda/U(1)]a_0}e^{i\hat{p}_2\cdot\Lambda_D[\Lambda/U(1)]a_0}\sum_{i_1'=-S_1}^{+S_1}\sum_{i_2'=-S_2}^{+S_2}D^{[S_1]}_{i_1i_1'}(u)\\
&\qquad\times D^{[S_2]}_{i_2i_2'}(u)F_{(i_1'i_2'D)}(\Lambda'/U(1)), \qquad (79)
\end{aligned}$$

where use has been made of the fact that $U(1)$ commutes with Λ_D, the representations of H_D are given in Eq. (75), and, as in the case of the right cosets, Λ_c, u, and $\Lambda'/U(1)$ are defined by $(\Lambda/U(1))\Lambda_0 = u(\Lambda'/U(1))$. Equation (79) can be simplified by defining $P_D = \hat{p}_1 + \Lambda_D^{-1}\hat{p}_2$ so that

$$H(P_D)=\begin{pmatrix}M_1 & 0\\ 0 & M_1\end{pmatrix}+\begin{pmatrix}M_2/D^2 & 0\\ 0 & M_2D^2\end{pmatrix}=\begin{pmatrix}M_1+M_2/D^2 & 0\\ 0 & M_1+M_2D^2\end{pmatrix} \quad (80)$$

and noting that $D^{[S]}_{ii'}(u) = [u]^i\delta_{ii'}$. Then

$$U(\Lambda_0, a_0)F_{(i_1i_2D)}(\Lambda/U(1))=e^{iP_D\cdot[\Lambda/U(1)]a_0}[u]^{i_1+i_2}F_{(i_1i_2D)}(\Lambda'/U(1)). \quad (81)$$

Equation (81) shows that the indices (i_1i_2) on

$$F_{(i_1i_2D)}(\Lambda/U(1))$$

are not indices which label the components of a vector in the $(2S_1 + 1) \times (2S_2 + 1)$-dimensional vector space as in the case of $f_{i_1i_2}(\Lambda_{c_1}, \Lambda_{c_2})$. Rather, they are degeneracy labels, distinguishing between the different ways in which $i_1 + i_2$ can give the same value. That is, $i_1 + i_2$ ranges between $-(2S + 1)(2S_2 + 1)$ and $+(2S_1 + 1)(2S_2 + 1)$; for a fixed but arbitrary value of k, $|k| \leq (2S_1 + 1)(2S_2 + 1)$, there will be various choices of i_1 and i_2 such that $i_1 + i_2 = k$ and the subscripts i_1, i_2 on $F_{(i_1i_2D)}(\Lambda/U(1))$ serve to distinguish these choices.

The functions $F_{(i_1i_2D)}(\Lambda/U(1))$ are elements of a Hilbert space having norm

$$\|F_{(i_1i_2D)}\|^2=\int d\frac{\Lambda}{U(1)}\left|F_{(i_1i_2D)}\left(\frac{\Lambda}{U(1)}\right)\right|^2<\infty. \quad (82)$$

The measure $d(\Lambda/U(1))$ is not specified since it is not needed.

We thus have the first stage of the tensor-product decomposition completed, namely the direct-integral (and sum) decomposition of

$$f_{i_1i_2}(\Lambda_{c_1}, \Lambda_{c_2}) \quad \text{into} \quad F_{(i_1i_2D)}(\Lambda/U(1)):$$

$$\|f_{i_1i_2}\|^2=\sum_{i_1i_2}\int dD\,\|F_{(i_1i_2D)}\|^2, \quad (83)$$

where dD is the direct-integral measure which is not specified, because it is not needed.

The second stage of the tensor-product decomposition consists in expanding $F_{(i_1i_2D)}(\Lambda/U(1))$ in a series of irreducible functions of P. The induced representation given in Eq. (81) is reducible since it is induced by the subgroup H_D, which is a "smaller" subgroup than H of Eq. (58) which induces irreducible representations of P. The difference is that H contains $SU(2)$ whereas H_D contains only $U(1)$. We will therefore expand $F_{(i_1i_2D)}(\Lambda/U(1))$ with respect to irreducible functions over $SU(2)/U(1)$, namely, $D^{[S]}_{ii}(SU(2)/U(1))$.

Before carrying out this expansion, however, it is necessary to look at the representation Eq. (81) somewhat more carefully. Notice that the standard or rest-frame vector $\hat{p}$ has the property that

$$SU(2)H(\hat{p})SU(2)^{+} = H(\hat{p});$$

we must also inquire as to which elements of $SL(2, C)$ leave $H(P_D)$ invariant. It is clear, first, that $U(1)$ leaves $H(P_D)$ invariant, for, according to Eq. (80), $H(P_D)$ has only diagonal entries.

Call $SU(2)_D$ that subgroup of Λ which leaves $H(P_D)$ invariant, i.e. $SU(2)_D H(P_D) SU(2)_D^+ = H(P_D)$.[17] Now, $\Lambda_c^{-1}(P_D)$ [see Eq. (68)] is the Lorentz transformation from the rest frame $\hat{P}$ to P_D. But

$$SU(2)H(\hat{P})SU(2)^+ = H(\hat{P})$$

and

$$H(P_D) = \Lambda_c^{-1}(P_D)H(\hat{P})\Lambda_c^{-1+}(P_D)$$

so that

$$SU(2)_D\Lambda_c^{-1}(P_D)H(\hat{P})\Lambda_c^{-1+}(P_D)SU(2)_D^+ = \Lambda_c^{-1}(P_D)H(\hat{P})\Lambda_c^{-1+}(P_D)$$

and, therefore,

$$\Lambda_c(P_D)SU(2)_D\Lambda_c^{-1}(P_D) = SU(2),$$

since any transformation leaving $H(\hat{P})$ invariant is a rotation. Clearly the representations of $SU(2)_D$ are the same as $SU(2)$, namely,

$$SU(2)_D \to D^{[S]}_{j'j}(SU(2)), \tag{84}$$

where the $SU(2)$ element is given by the equation $\Lambda_c(P_D)SU(2)_D\Lambda_c^{-1}(P_D)$.

We next decompose P into right cosets with respect to

$$\begin{pmatrix} SU(2)_D & \{H(a)\}SU(2)_D \\ 0 & SU(2)_D \end{pmatrix}$$

in preparation for writing an induced irreducible representation of P:

$$\begin{pmatrix} \Lambda & H(a)\Lambda^{-1+} \\ 0 & \Lambda^{-1+} \end{pmatrix} = \begin{pmatrix} R_D & H(a)R_D \\ 0 & R_D^{-1+} \end{pmatrix}\begin{pmatrix} \Lambda_c & 0 \\ 0 & \Lambda_c^{-1+} \end{pmatrix}, \tag{85}$$

where R_D is an element of $SU(2)_D$ and Λ_c is the same right coset as given in Eq. (60); this choice of right cosets is a most convenient one and it is not hard to see that $\{\Lambda_c\}$ and $SU(2)_D$ cover all elements of $SL(2, C)$ uniquely. The elements $\Lambda/U(1)$ can then be written as $[R_D/U(1)]\Lambda_c$, and it is possible to proceed with the expansion of $F_{(i_1i_2D)}(\Lambda/U(1))$:

$$F_{(i_1i_2D)}\left(\frac{\Lambda}{U(1)}\right) = F_{(i_1i_2D)}\left(\frac{R_D}{U(1)}\Lambda_c\right) = \sum_{S=0}^{\infty}\sum_{i=-S}^{+S}\left(\frac{2S+1}{4\pi}\right)^{\frac{1}{2}} D^{[S]}_{i_1+i_2,i}\left(\Lambda_c(P_D)\frac{R_D}{U(1)}\Lambda_c^{-1}(P_D)\right) f_{i(Si_1i_2D)}(\Lambda_c), \tag{86}$$

with

$$f_{i(Si_1i_2D)}(\Lambda_c) = \left(\frac{2S+1}{4\pi}\right)^{\frac{1}{2}}\int d\left(\frac{R_D}{U(1)}\right) D^{[S]*}_{i_1+i_2,i}\left(\Lambda_c(P_D)\frac{R_D}{U(1)}\Lambda_c^{-1}(P_D)\right) F_{(i_1i_2D)}\left(\frac{R_D}{U(1)}\Lambda_c\right). \tag{87}$$

Now we test how $f_{i(Si_1i_2D)}(\Lambda_c)$ transforms under an arbitrary Poincaré transformation:

$$\begin{aligned}
f_{i(Si_1i_2D)}(\Lambda_c) &\xrightarrow{(\Lambda_0,a_0)} \left(\frac{2S+1}{4\pi}\right)^{\frac{1}{2}}\int d\left(\frac{R_D}{U(1)}\right) D^{[S]*}_{i_1+i_2,i}\left(\Lambda_c(P_D)\frac{R_D}{U(1)}\Lambda_c^{-1}(P_D)\right) e^{iP_D\cdot\Lambda/U(1)a_0} F_{(i_1i_2D)}\left(\frac{\Lambda}{U(1)}\Lambda_0\right) \\
&= \left(\frac{2S+1}{4\pi}\right)^{\frac{1}{2}}\int d\left(\frac{R_D}{U(1)}\right) D^{[S]*}_{i_1+i_2,i}\left(\Lambda_c(P_D)\frac{R_D}{U(1)}\Lambda_c^{-1}(P_D)\right) e^{i[\Lambda^{-1}/U(1)]P_D\cdot a_0} F_{(i_1i_2D)}\left(\frac{R_D}{U(1)}\tilde{R}_D\Lambda_{c'}\right) \\
&= \left(\frac{2S+1}{4\pi}\right)^{\frac{1}{2}}\int d\left(\frac{R_D}{U(1)}\right) D^{[S]*}_{i_1+i_2,i}\left(\Lambda_c(P_D)\frac{R_D}{U(1)}\Lambda_c^{-1}(P_D)\right) e^{i\Lambda_c^{-1}P_D\cdot a_0} F_{(i_1i_2D)}\left(u\frac{R'_D}{U(1)}\Lambda_{c'}\right) \\
&= e^{i\Lambda_c^{-1}P_D\cdot a_0}\left(\frac{2S+1}{4\pi}\right)^{\frac{1}{2}}\int d\left(\frac{R_D}{U(1)}\right)[u]^{i_1+i_2} D^{[S]*}_{i_1+i_2,i}\left[\Lambda_c(P_D)(u)\frac{R'_D}{U(1)}\tilde{R}_D^{-1}\Lambda_c^{-1}(P_D)\right] F_{(i_1i_2D)}\left(\frac{R'_D}{U(1)}\Lambda_{c'}\right) \\
&= e^{i\Lambda_c^{-1}P_D\cdot a_0}\left(\frac{2S+1}{4\pi}\right)^{\frac{1}{2}}\int d\left(\frac{R_D}{U(1)}\right) D^{[S]*}_{i_1+i_2,i}\left[\Lambda_c(P_D)\frac{R'_D}{U(1)}\Lambda_c^{-1}(P_D)\Lambda_c(P_D)\tilde{R}_D^{-1}\Lambda_c^{-1}(P_D)\right] F_{(i_1i_2D)}\left(\frac{R'_D}{U(1)}\Lambda_{c'}\right) \\
&= e^{iP_D\cdot\Lambda_c a_0}\sum_{i'=-S}^{+S} D^{[S]*}_{i'i}(\Lambda_c(P_D)\tilde{R}_D^{-1}\Lambda_c^{-1}(P_D))\left(\frac{2S+1}{4\pi}\right)^{\frac{1}{2}}\int d\left(\frac{R'_D}{U(1)}\right) \\
&\quad\times D^{[S]*}_{i_1+i_2,i'}\left(\Lambda_c(P_D)\frac{R'_D}{U(1)}\Lambda_c^{-1}(P_D)\right) F_{(i_1i_2D)}\left(\frac{R'_D}{U(1)}\Lambda_{c'}\right) \\
&= e^{iP_D\cdot\Lambda_c a_0}\sum_{i'=-S}^{+S} D^{[S]}_{ii'}(\Lambda_c(P_D)\tilde{R}_D\Lambda_c^{-1}(P_D)) f_{i'(Si_1i_2D)}(\Lambda_{c'}).
\end{aligned} \tag{88}$$

Here, use has been made of the fact that $U(1)$ commutes with $\Lambda_c(P_D)$ and the measure $dR_D/U(1)$ is invariant under a rotation $\tilde{R}_D$; the elements $\tilde{R}_D$, $R'_D/U(1)$, $\Lambda_{c'}$, and u are defined by the following equations:

$$\frac{\Lambda}{U(1)}\Lambda_0 = \frac{R_D}{U(1)}\Lambda_c\Lambda_0 = \frac{R_D}{U(1)}\tilde{R}_D\Lambda_{c'} = u\frac{R'_D}{U(1)}\Lambda_{c'},$$

[17] The relation $SU(2)H(\hat{p})SU(2)^+ = H(\hat{p})$ can also be written in the more usual form

$$\begin{pmatrix} SO(3) & 1 \\ 0 & 0 \end{pmatrix}\hat{p} = \hat{p},$$

where $SO(3)$ is the three-dimensional rotation matrix whose elements are related to $SU(2)$ in a 2 to 1 fashion. There is obviously a similar correspondence between $SU(2)_D$ and $SO(3)_D$, where $SO(3)_D$ is a subgroup of the four-dimensional Lorentz group satisfying

$$SO(3)_D P_D = P_D.$$

so that

$$R_D/U(1) = uR'_D/U(1)\tilde{R}_D^{-1}.$$

But $\Lambda_c(P_D)\tilde{R}_D\Lambda_{c'}^{-1}(P_D) = R$ so that $f_{i(Si_1i_2D)}(\Lambda_c)$ is transformed into

$$e^{iP_D\cdot\Lambda_c a_0}\sum_{i'=-S}^{+S} D_{ii'}^{[S]}(R) f_{i'(Si_1i_2D)}(\Lambda_{c'}),$$

which agrees with Eq. (64) and, hence, shows that $f_{i(Si_1i_2D)}(\Lambda_c)$ is an element of an irreducible function space of P.

Thus the transformations which decompose $f_{i_1i_2}(\Lambda_{c_1}, \Lambda_{c_2})$ into irreducible functions $f_{i(Si_1i_2D)}(\Lambda_c)$ have been found. Symbolize these transformations by S so that

$$S: f_{i_1i_2}(\Lambda_{c_1}, \Lambda_{c_2}) \to f_{i(Si_1i_2D)}(\Lambda_c) = Sf_{i_1i_2}(\Lambda_{c_1}, \Lambda_{c_2}).$$

To get the Clebsch–Gordan coefficients, we consider the inner product

$$\begin{aligned}(f'_i(\Lambda_c), f_{i(Si_1i_2D)}(\Lambda_c)) &= \sum_{i=-S}^{+S}\int d\Lambda_c f'^{*}_i(\Lambda_c) f_{i(Si_1i_2D)}(\Lambda_c),\\ (f'_i(\Lambda_c), Sf_{i_1i_2}(\Lambda_{c_1}, \Lambda_{c_2})) &= \sum_{i=-S}^{+S}\int d\Lambda_c f'^{*}_i(\Lambda_c) Sf_{i_1i_2}(\Lambda_{c_1}, \Lambda_{c_2}),\end{aligned} \tag{89}$$

using for $f'_i(\Lambda_c)$, the function $D^{[M,S]}_{P_D iP\sigma}(\Lambda_c)$ and for $f_{i_1i_2}(\Lambda_{c_1}, \Lambda_{c_2})$ the functions $D^{[M_1S_1]}_{\hat{p}_1i_1p_1\sigma_1}(\Lambda_{c_1}) D^{[M_2S_2]}_{\hat{p}_2i_2p_2\sigma_2}(\Lambda_{c_2})$.

Making the substitutions of Eq. (78), whereby $(\Lambda_{c_1}, \Lambda_{c_2})$ is replaced by $\Lambda/U(1)$ and Λ_D, gives

$$\begin{aligned}D^{[M_1S_1]}_{\hat{p}_1i_1p_1\sigma_1}(\Lambda_{c_1}) D^{[M_2S_2]}_{\hat{p}_2i_2p_2\sigma_2}(\Lambda_{c_2}) &= D^{[M_1S_1]}_{\hat{p}_1i_1p_1\sigma_1}(\Lambda/U(1)) D^{[M_2S_2]}_{\hat{p}_2i_2p_2\sigma_2}[\Lambda_D(\Lambda/U(1))]\\ &= D^{[M_1S_1]}_{\hat{p}_1i_1p_1\sigma_1}(\Lambda/U(1)) D^{[M_2S_2]}_{(P_D-\hat{p}_1)i_2p_2\sigma_2}(\Lambda/U(1))\\ &= F_{(i_1i_2D)}(\Lambda/U(1)).\end{aligned} \tag{90}$$

Finally,

$$\begin{aligned}f_{i(Si_1i_2D)}(\Lambda_c) = \left(\frac{2S+1}{4\pi}\right)^{\frac{1}{2}}\int d(R_D/U(1)) D^{[S]^*}_{i_1+i_2,i}(\Lambda_c(P_D)[R_D/U(1)]\Lambda_c^{-1}(P_D))\\ \times D^{[M_1S_1]}_{\hat{p}_1i_1p_1\sigma_1}[R_D/U(1)]\Lambda_c D^{[M_2S_2]}_{(P_D-\hat{p}_1)i_2p_2\sigma_2}[R_D/U(1)]\Lambda_c,\end{aligned} \tag{91}$$

so that

$$\begin{aligned}&\langle [MS]\mathbf{P}\sigma; i_1i_2 \mid [M_1S_1]\mathbf{p}_1\sigma_1; [M_2S_2]\mathbf{p}_2\sigma_2\rangle\\ &\quad= N\sum_{i=-S}^{+S}\int d\Lambda_c D^{[MS]^*}_{P_D iP\sigma}(\Lambda_c)\left(\frac{2S+1}{4\pi}\right)^{\frac{1}{2}}\int d\left(\frac{R_D}{U(1)}\right)\\ &\qquad\times D^{[S]^*}_{i_1+i_2,i}\left[\Lambda_c(P_D)\frac{R_D}{U(1)}\Lambda_c^{-1}(P_D)\right] D^{[M_1S_1]}_{\hat{p}_1i_1p_1\sigma_1}\left(\frac{R_D}{U(1)}\Lambda_c\right)\\ &\qquad\times D^{[M_2S_2]}_{(P_D-\hat{p}_1)i_2p_2\sigma_2}\left(\frac{R_D}{U(1)}\Lambda_c\right)\\ &\quad= N\left(\frac{2S+1}{4\pi}\right)^{\frac{1}{2}}\sum_{i=-S}^{+S}\int d\Lambda_c\, d\left(\frac{R_D}{U(1)}\right)\delta^3(P_D - \Lambda_c P) D^{[S]^*}_{i\sigma}(P, \Lambda_c)\\ &\qquad\times D^{[S]^*}_{i_1+i_2,i}\left[\Lambda_c(P_D)\frac{R_D}{U(1)}\Lambda_c^{-1}(P_D)\right]\delta^3\left(\hat{p}_1 - \frac{R_D}{U(1)}\Lambda_c p_1\right) D^{[S_1]}_{i_1\sigma_1}\left(p_1, \frac{R_D}{U(1)}\Lambda_c\right)\\ &\qquad\times \delta^3\left(P_D - \hat{p}_1 - \frac{R_D}{U(1)}\Lambda_c p_2\right) D^{[S_2]}_{i_2\sigma_2}\left(p_2, \Lambda_D\frac{R_D}{U(1)}\Lambda_c\right)\\ &\quad= N\left(\frac{2S+1}{4\pi}\right)^{\frac{1}{2}}\delta^3(\mathbf{P} - \mathbf{p}_1 - \mathbf{p}_2)\int d\Lambda_c\, d\left(\frac{R_D}{U(1)}\right)\delta^3\left(\hat{p}_1 - \frac{R_D}{U(1)}\Lambda_c p_1\right)\\ &\qquad\times \delta^3\left(P_D - \hat{p}_1 - \frac{R_D}{U(1)}\Lambda_c p_2\right) D^{[S]^*}_{i_1+i_2,\sigma}\left[\Lambda_c(P_D)\frac{R_D}{U(1)}\Lambda_c\Lambda_c^{-1}(P)\right]\\ &\qquad\times D^{[S_1]}_{i_1\sigma_1}\left(p_1, \frac{R_D}{U(1)}\Lambda_c\right) D^{[S_2]}_{i_2\sigma_2}\left(p_2, \Lambda_D\frac{R_D}{U(1)}\Lambda_c\right),\end{aligned} \tag{92}$$

where delta-function-normalization factors have been absorbed in the factor N.

There are five integrations and six δ functions; the remaining δ function will give energy conservation or, what is equivalent, the value of the mass M. As with the group P_2 this is equivalent to fixing the double-coset label, for

$$M^2 = P_D^2 = |H(P_D)| = M_1^2 + M_2^2 + 2M_1M_2((D^2 + D^{-1})/2),$$

as can be seen from Eq. (80). Now the δ functions can be written as

$$\delta^3(\hat{p}_1 - \Lambda/U(1)p_1)\delta^3(\hat{p}_2 - \Lambda_D\Lambda/U(1)p_2),$$

so that

$$\hat{p}_1 = \Lambda/U(1)p_1 \quad \text{and} \quad \hat{p}_2 = \Lambda_D\Lambda/U(1)p_2.$$

Consider the quantity

$$p_1 \cdot p_2 = \Lambda^{-1}/U(1)\hat{p}_1 \cdot \Lambda^{-1}/U(1)\Lambda_D^{-1}\hat{p}_2 = \hat{p}_1 \cdot \Lambda_D^{-1}\hat{p}_2.$$

But

$$\hat{p}_1 \cdot \Lambda_D^{-1}\hat{p}_2 = \frac{|H(\hat{p}_1 + \Lambda_D^{-1}\hat{p}_2)| - M_1^2 - M_2^2}{2} = M_1M_2\frac{D^2 + D^{-2}}{2}$$

so that

$$p_1 \cdot p_2 = M_1M_2(D^2 + D^{-2})/2$$

and

$$M^2 = M_1^2 + M_2^2 + 2p_1 \cdot p_2 = (p_1 + p_2)^2.$$

To actually carry out the integrations explicitly is not difficult, since the integration is essentially

$$\int d(\Lambda/U(1))\delta^3(\hat{p}_1 - \Lambda/U(1)p_1)\delta^3(P_D - \hat{p}_1 - \Lambda/U(1)p_2)$$

and it is not difficult to obtain the measure $d(\Lambda/U(1))$.

It is, however, more useful to carefully specify the rotations involved in the Wigner functions, for these rotations are the quantities of most interest in the Clebsch–Gordan coefficients. To that end, consider Fig. 1 in which three frames are considered, one the arbitrary frame in which particles 1 and 2 have momenta $\mathbf{p}_1$ and $\mathbf{p}_2$ and two fixed frames, the rest frame of particle 1 and the center-of-mass frame (or, what is equivalent, the rest frame of the biparticle).

Notice that the Lorentz transformations $\Lambda_c^{-1}(p_1)$, which carries p_1 from its rest frame and $\Lambda_c^{-1}(P)$ which carries P from its rest frame (the center-of-mass frame) are both fixed as soon as $\mathbf{p}_1$ and $\mathbf{p}_2$ are chosen in some arbitrary coordinate system. Now

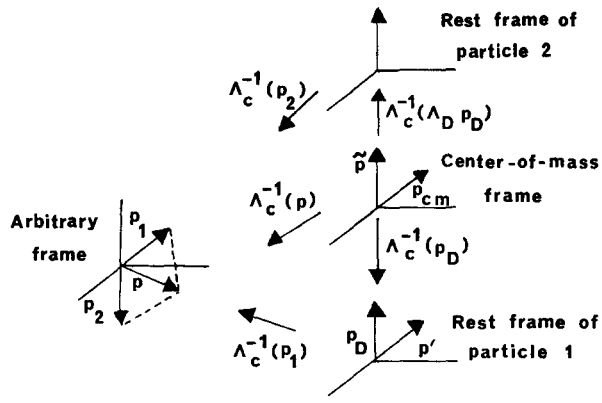

FIG 1. Coordinate frames relevant to the Clebsch–Gordon problem.

$$D^{[S_1]}_{i_1\sigma_1}(p_1, \Lambda/U(1))$$

deals with the rotation

$$(p_1, \Lambda/U(1)) = \Lambda_c[\Lambda/U(1)p_1][\Lambda/U(1)]\Lambda_c^{-1}(p_1) = [\Lambda/U(1)]\Lambda_c^{-1}(p_1)$$

since $\hat{p}_1 = \Lambda/U(1)p_1$. But $\Lambda_c(p_1)$ is fixed by the vector p_1 and $\Lambda/U(1)$ is fixed by $\hat{p}_1 = \Lambda/U(1)p_1$ and $\Lambda_D^{-1}\hat{p}_2 = \Lambda/U(1)p_2$, with D fixed by $\frac{1}{2}(D^2 + D^{-2}) = p_1 \cdot p_2/M_1M_2$. This defines the rotations, but to see which rotation this is in the figure, let $P' = \Lambda_c(p_1)P$. Now $P_D = \Lambda/U(1)P$ is aligned along the "z" axis in the rest frame of particle 1, so that $(p_1, \Lambda/U(1))$ is the rotation which carries the vector P' into the z axis. A similar sort of construction is easily carried out for

$$D^{[S_2]}_{i_2\sigma_2}(p_2, \Lambda_D\Lambda/U(1)).$$

Finally, there is the rotation in

$$D^{[S]*}_{i_1+i_2,\sigma}[\Lambda_c(P_D)R_D/U(1)\Lambda_c\Lambda_c^{-1}(P)].$$

It will be shown that this rotation carries p_{cm} into $\tilde{p}$ as shown in Fig. 1. To see this, define $\tilde{p} = \Lambda_c^{-1}(P_D)\hat{p}_1$. Then $\tilde{p}$ is aligned along the z axis in the center-of-mass frame since P_D has no x or y components. Also, $\hat{p}_1 = \Lambda/U(1)p_1 = R_D/U(1)\Lambda_c p_1$, where here Λ_c is the unique right-coset element carrying P to P_D so that

$$\tilde{p} = \Lambda_c(P_D)\hat{p}_1 = \Lambda_c(P_D)\frac{R_D}{U(1)}\Lambda_c p_1 = \Lambda_c(P_D)\frac{R_D}{U(1)}\Lambda_c\Lambda_c^{-1}(P)p_{cm}, \tag{93}$$

where $p_{cm} = \Lambda_c(P)p_1$; but $\Lambda_c(P_D)R_D/U(1)\Lambda_c\Lambda_c^{-1}(P)$ is the rotation contained in the $D^{[S]*}_{i_1+i_2,\sigma}$ function and this

is the rotation which carries p_{cm} to $\tilde{p}$. The Clebsch–Gordan coefficients are thus

$$\begin{aligned}\langle [MS]\mathbf{p}\sigma;\, i_1 i_2 \mid [M_1S_1]\mathbf{p}_1\sigma_1;\, [M_2S_2]\mathbf{p}_2\sigma_2\rangle \\ = N\left[\frac{(2S+1)}{4\pi}\right]^{\frac{1}{2}} \delta^3(\mathbf{p}-\mathbf{p}_1-\mathbf{p}_2) \\ \times \delta(M^2-(p_1+p_2)^2) \\ \times D^{[S]*}_{i_1+i_2,\sigma}(\mathbf{p}_{cm}\to\tilde{\mathbf{p}}) D^{[S]}_{i_1\sigma_1}(p_1, [\Lambda/U(1)]) \\ \times D^{[S_2]}_{i_2\sigma_2}(p_2, \Lambda_D[\Lambda/U(1)]). \qquad (94)\end{aligned}$$

The labels i_1 and i_2, as seen from Eq. (81), are degeneracy labels which serve to distinguish equivalent representations labeled by $[MS]$.

IV. OUTLINE OF CLEBSCH–GORDAN COEFFICIENTS OF THE LORENTZ GROUP $SL(2, C)$

Thus far, the direct-integral decomposition of tensor products has been carried out explicitly for two groups P_2 and P. In the case of $SL(2, C)$, the tensor-product decomposition has been carried out in a series of papers by Naimark.[10]

Consider, for the moment, the tensor-product decomposition for the principal series of representations of $SL(2, C)$.[9] For the principal series of representations, the inducing subgroup of $SL(2, C)$ is $\{\left(\begin{smallmatrix}\alpha & \beta\\ 0 & 1/\alpha\end{smallmatrix}\right)\}$, which has one-dimensional representations; the right cosets can be chosen to be $\left(\begin{smallmatrix}1 & 0\\ z & 1\end{smallmatrix}\right)$. Thus, induced representations are defined on functions $f(z)$, with norm

$$\|f\|^2 = \int dz |f(z)|^2 < \infty. \qquad (95)$$

The tensor-product space consists of functions $F(z_1, z_2)$ with norm

$$\|F\|^2 = \int dz_1\, dz_2\, |F(z_1 z_2)|^2 < \infty. \qquad (96)$$

According to Naimark, the operator S, which carries $F(z_1, z_2)$ into $f_{[m,\sigma]}(z)$, where $[m, \sigma]$ are the irreducible representation labels of $SL(2, C)$, can be written as a kernel transformation

$$f_{[m,\sigma]}(z) = \int dz_1\, dz_2 a(z_1, z_2, z, [m, \sigma]) F(z_1, z_2), \qquad (97)$$

where

$$\begin{aligned}a(z_1, z_2, z, [m, \sigma]) \\ = |z_2 - z_1|^{\frac{1}{2}(m+m_1+m_2)-i\frac{1}{2}(\sigma+\sigma_1+\sigma_2)-1}(z_2-z_1)^{\frac{1}{2}(-m+m_1+m_2)} \\ \times |z - z_1|^{\frac{1}{2}(-m-m_1+m_2)+i\frac{1}{2}(\sigma-\sigma_1+\sigma_2)-1}(z-z_1)^{\frac{1}{2}(m-m_1+m_2)} \\ \times |z_2 - z|^{\frac{1}{2}(-m+m_1-m_2)+i\frac{1}{2}(\sigma+\sigma_1-\sigma_2)-1}(z_2-z)^{\frac{1}{2}(m+m_1-m_2)}. \qquad (98)\end{aligned}$$

To get the Clebsch–Gordan coefficients it is necessary to obtain the functions $D^{[m,\sigma]}_{J'M'JM}(z)$ and check that they have the correct transformation properties. It seems, although this has not been shown, that the values of $J'M'$ will be fixed by demanding that $D^{[m,\sigma]}_{J'M'JM}(z)$ have the correct transformation properties, exactly as in Sec. II, where $D^{[M]}_{p'p}(\Lambda)$ has the correct transformation properties under $U(\Lambda_0, a_0)$ only if p' equals $\hat{p}$.

The Clebsch–Gordan coefficients are

$$\begin{aligned}\langle [m\sigma]JM \mid [m_1\sigma_1]J_1M_1;\, [m_2\sigma_2]J_2M_2\rangle \\ = N(D^{[m\sigma]}_{J'M'JM}(z), S\{D^{[m_1\sigma_1]}_{J_1'M_1'J_1M_1}(z_1) D^{[m_2\sigma_2]}_{J_2'M_2'J_2M_2}(z_2)\}) \\ = N\int dz\, dz_1\, dz_2 D^{[m\sigma]*}_{J'M'JM}(z) D^{[m_1\sigma_1]}_{J_1'M_1'J_1M_1}(z_1) \\ \times D^{[m_2\sigma_2]}_{J_2'M_2'J_2M_2}(z_2) a(z_1, z_2, z, [m\sigma]), \qquad (99)\end{aligned}$$

where, by the above remarks, all of the primed variables are fixed. The functions $D^{[m,\sigma]}_{J'M'JM}(z)$ have been obtained by several authors,[4] but, since they are rather complicated, will not be given here.

CONCLUSION

It has been shown that, for induced representation, functions over cosets are useful in evaluating Clebsch–Gordan coefficients *if* the tensor-product decomposition of two irreducible representations into a direct integral of irreducible representations is known. The method is sufficiently general to handle even those reductions which are not multiplicity free.

In actually computing the Clebsch–Gordan coefficients a key role is played by the functions

$$D^{[\chi]}_{x_0x}(g_c) = \langle [\chi]x_0 |\, U(g_c)\, |[\chi]x\rangle;$$

here x_0 is meant to denote a "standard state." The functions $D^{[\chi]}_{x_0x}(g_c)$ play a dual role in that they are, on the one hand, a concrete realization of $|[\chi]x\rangle$, while on the other hand all the manipulations as regards induced representation theory are done on the standard states x_0. It is because of this dual role that the $D^{[\chi]}_{x_0x}(g_c)$ functions serve as the bridge between the abstract spaces of induced representation theory and the concrete basis functions $|[\chi]x\rangle$.

ACKNOWLEDGMENTS

The author wishes to thank Dr. Edward McCliment of the University of Iowa and Dr. Ernest Thieleker of the Applied Mathematics Division of Argonne National Laboratories for many helpful discussions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Solution of the Faddeev Equation for Local Potentials by Approximate Product-Integration*

Y. E. KIM
Department of Physics, Purdue University, Lafayette, Indiana

(Received 24 March 1969)

A practical method of solving the Faddeev equation for short-range local potentials is proposed. The method consists of (a) transforming the Faddeev equation into a form in which the inhomogeneous term satisfies a Hölder condition and then (b) solving the resultant two-variable integral equations by approximate product-integration.

In recent years, the formal theory of the quantum-mechanical three-body problem has been greatly advanced through the work of Faddeev.[1] The Schrödinger equation for nonrelativistic three-particle systems can be reformulated into a set of integral equations of the Fredholm type, now known as the Faddeev equation. The application of the Faddeev equation to nonrelativistic three-body problems has been of considerable interest because the Faddeev equation treats on an equal basis all possible nonrelativistic processes such as bound states, elastic scattering, and break-up.

The Faddeev equation can be reduced to a set of coupled integral equations in two continuous variables with $3 \times (L+1) \times \min(2J+1, 2L+1)$ components, where L is the maximum orbital angular momentum in the two-body subsystems and J is the total angular momentum.[2,3] If the two-body interaction or t matrix is separable, then the Faddeev equation reduces to a set of coupled integral equations in one continuous variable. Considerable work has been done on the three-body problem with separable interactions,[4] because the numerical problems are manageable for solving one-variable integral equations. This separable approximation is expected to remain as a practical tool for investigating the three-body problem.

While simplicity is an attractive reason for using separable potentials, these potentials do not necessarily give accurate representations of a physical interaction. The local potential description of interactions is known to be applicable to atomic problems (Coulomb interaction) and to the two-nucleon system (at least for the one-pion exchange contribution to the potential). It is therefore desirable to develop a practical method of solving the three-body Faddeev equations with local interactions.

For simplicity, we consider the case of a $J = 0$ state of three identical spinless bosons interacting pairwise through an s-state local potential. Then the two-variable integral equation to be solved can be written as[2,5] (we use the same definitions of kinematic variables as in Refs. 2 and 5)

$$\psi_s(p, q) = \varphi_s(p, q; p_0, q_0) + \int_0^\infty dq_2^2 \int_{B(q,q_2)}^{A(q,q_2)} dp_2^2 K_s(p, q; p_2, q_2)\psi_s(p_2, q_2), \quad (1)$$

where the kernel is given by

$$K_s(p, q; p_2, q_2) = \frac{2}{(3)^{\frac{1}{2}}\pi q} t(p^2, p_1^2; S - q^2) \frac{1}{p_2^2 + q_2^2 - S},$$

and the integration limits A and B are given by

$$A(q, q_2) = \tfrac{1}{3}(2q + q_2)^2$$

and

$$B(q, q_2) = \tfrac{1}{3}(2q - q_2)^2.$$

The variable p is proportional to the magnitude of relative momentum of the pair and q is proportional to the magnitude of the momentum of the third particle in the three-body center-of-mass coordinate. The constants p_0 and q_0 are the initial momenta. $\psi_s(p, q)$ is three-body T matrix with the pair interacting with relative angular momentum $l = 0$ in the final state. The function $t(p^2, p_1^2; S - q^2)$ is the two-body t matrix for the interaction between the pair of particles and p_1 is defined by $p_1^2 = p_2^2 + q_2^2 - q^2$. It is normalized so that on the energy shell it is $t(p^2, p^2; p^2) = (e^{i\delta} \sin \delta)/p$, where δ is the s-wave phase shift. The total energy S of the three-particle system is given in

* Work supported by the U.S. Atomic Energy Commission.

[1] L. D. Faddeev, *Mathematical Aspects of the Three-Body Problem in the Quantum Scattering Theory* (Daniel Davey and Co., Inc., Hartford, Conn., 1965).

[2] A. Ahmezadeh and J. A. Tjon, Phys. Rev. **139**, B1085 (1965).

[3] T. A. Osborn and H. P. Noyes, Phys. Rev. Letters **17**, 215 (1966).

[4] I. Duck, in *Advances in Nuclear Physics*, M. Baranger and E. Vogt, Eds. (Plenum Press, Inc., New York, 1968), Vol. 1, pp. 343–409, and references cited therein.

[5] D. Y. Wong and G. Zambotti, Phys. Rev. **154**, 1050 (1967); J. S. Ball and D. Y. Wong, Phys. Rev. **169**, 1362 (1968).

the center-of-mass system. A small positive imaginary part $i\epsilon$ in S is suppressed here and in the following.

For the three-particle bound-state problem, we need not specify the inhomogeneous term $\varphi_s(p, q; p_0, q_0)$; furthermore, the initial states $|p_0 q_0\rangle$ need not be specified, because the kernel of Eq. (1) is independent of p_0 and q_0. However, most of physical experiments involve scattering of a particle by a bound pair. In this case, the initial states must be specified by the initial momenta p_0 and q_0, and the inhomogeneous term takes the form of

$$\varphi_s(p, q; p_0, q_0) = -(4/\pi q)t(p^2, p_0^2; S - q^2)\delta(q^2 - q_0^2). \tag{2}$$

The two-body t matrix contains poles corresponding $S - q^2 = S_0$, where S_0's are the two-particle bound-state (negative) energies. Therefore, it is convenient to write t as

$$t(p^2, p_0^2; S - q^2) = \tilde{t}(p^2, p_0^2; S - q^2)/D(S - q^2),$$

where $D(S - q^2)$ is the determinant $|1 + \mathbf{VG}|$ of a matrix equation which is obtained from the two-body Lippmann–Schwinger equation, $t = V + VGt$, after discretizing the momentum variables.[5] In the neighborhood of $S_0 = S - q^2$, the function $D(S - q^2)$ behaves as $[(S - q^2) - S_0]$.

As it stands now, the inhomogeneous term φ_s given by Eq. (2) does not satisfy a Hölder condition[6] due to the presence of singularities. For reasons which will become clear later, we transform Eq. (1) into a form in which the inhomogeneous term satisfies a Hölder condition:

$$\tilde{\psi}_s(p, q; p_0, q_0) = \tilde{\varphi}_s(p, q; p_0, q_0) + \int_0^\infty dq_2^2 \int_{B(q,q_2)}^{A(q,q_2)} dp_2^2 \tilde{K}_s(p, q; p_2, q_2)\tilde{\psi}_s(p_2, q_2; p_0, q_0), \tag{3}$$

where

$$\tilde{\varphi}_s(p, q; p_0, q_0) = qD(S - q^2)[\psi_s(p, q) - \varphi_s(p, q; p_0, q_0)]$$

and

$$\tilde{\varphi}_s(p, q; p_0, q_0) = -\frac{4}{\pi}\int_{B(q,q_0)}^{A(q,q_0)} dp_2^2 \tilde{K}_s(p, q; p_2, q_0)\tilde{t}(p_2^2, p_0^2; S - q_0^2),$$

with

$$\tilde{K}_s(p, q; p_2 q_2) = \frac{2}{(3)^{\frac{1}{2}}\pi}\frac{\tilde{t}(p^2, p_1^2; S - q^2)}{q_2 D(S - q_2^2)}\frac{1}{p_2^2 + q_2^2 - S}.$$

Equation (3) or (1) cannot be reduced directly into a matrix equation, because the integration limits for p_2 integration are not constants and are dependent on the q and q_2 variables.

We define a singular operator $\mathbf{K}$ by the following relation:

$$\mathbf{K}\tilde{\psi}_s(p, q) \equiv \tilde{\psi}_s(p, q) - \int_0^\infty dq_2^2 \int_{B(q,q_2)}^{A(q,q_2)} dp_2^2 \times \tilde{K}_s(p, q; p_2, q_2)\tilde{\psi}_s(p_2, q_2) = \tilde{\varphi}_s(p, q), \tag{4}$$

where we suppress the constants p_0 and q_0 in $\tilde{\psi}_s$ and $\tilde{\varphi}_s$. If we assume that $\tilde{\psi}_s(p, q)$ satisfies a Hölder condition $H(\mu, \nu)$, that is, for any given pairs of values (p'', q'') and (p', q'), there exists a relation

$$|\tilde{\psi}(p'', q'') - \tilde{\psi}_s(p', q')| \leq M|p'' - p'|^\mu + N|q'' - q'|^\nu,$$

with $0 < \mu \leq 1$ and $0 < \nu \leq 1$, where M and N are constants, then the singular operator $\mathbf{K}$ transforms $\tilde{\psi}_s(p, q)$ into a new function $\tilde{\varphi}(p, q)$, which also satisfies the Hölder condition. The above statement is valid if the singularities appearing in $\tilde{K}_s$ are of the Cauchy type.[7] Conversely, if $\tilde{\varphi}_s(p, q)$ satisfies the Hölder condition, then $\tilde{\psi}_s(p, q)$ also satisfies the Hölder condition under the transformation $\mathbf{K}$. The argument presented here can also serve as a proof of the existence of solution $\tilde{\psi}_s(p, q)$, where $\tilde{\psi}_s(p, q)$ is a function of two variables satisfying the Hölder condition. The same proof can be obtained by looking at the convergence of the Neumann series for $\tilde{\psi}_s(p, q)$, as is usually done for the solution of one-variable integral equation.

If $\tilde{\psi}_s(p, q)$ satisfies the condition $H(\mu, \nu)$, then $\tilde{\psi}_s(p, q)$ also satisfies the condition $H(\mu)$ for the variable p uniformly with respect to q and the condition $H(\nu)$ for q uniformly with respect to p.[6] Here $H(\mu)$ is the condition that

$$|\tilde{\psi}_s(p'', q'') - \tilde{\psi}_s(p', q'')| \leq M|p'' - p'|^\mu,$$

with $0 < \mu < 1$; similarly, for $H(\nu)$ it is

$$|\tilde{\psi}_s(p'', q'') - \tilde{\psi}_s(p'', q')| \leq N|q'' - q'|^\nu,$$

with $0 < \nu \leq 1$. Therefore, we can expand $\tilde{\psi}_s(p, q)$ in terms of a set of known linearly independent functions $F_{mn}(p, q)$ which satisfy the condition $H(\mu, \nu)$:

$$\tilde{\psi}_s(p, q) = \sum_{mn} a_{mn}(S)F_{mn}(p, q), \tag{5}$$

where $a_{mn}(S)$ are unknown complex coefficients. The p and q dependences of $\tilde{\psi}_s$ are included entirely in the function $F_{mn}(p, q)$. The above expansion is not restricted to a particular value of S, i.e., S can be either $S > 0$ or $S < 0$. The functions $F_{mn}(p, q)$ need not be separable in p and q variables.

[6] N. I. Muskhelishvili, *Singular Integral Equations* (P. Noordhoff Ltd., Groningen, The Netherlands, 1953), p. 12.

[7] See p. 50 of Ref. 6.

We select N_p points in the p variable and N_q points in the q variable, and write the integral part of Eq. (3) as

$$I \equiv \int_0^\infty dq_2^2 \int_{B(q_j,q_2)}^{A(q_j,q_2)} dp_2^2 \tilde{K}_s(p_i, q_j; p_2, q_2)\tilde{\psi}(p_2, q_2) \approx \sum_{kl} A_{kl}(p_i, q_j)\tilde{\psi}_s(p_{2k}, q_{2l}). \quad (6)$$

With this discretization of the continuous variables p and q, Eq. (3) becomes a matrix equation,

$$(\mathbf{1} - \mathbf{A})\mathbf{\Psi} = \mathbf{\Phi}, \quad (7)$$

where $\mathbf{A}$ is a $N \times N$ matrix with $N = N_p \times N_q$ and the matrix elements of $\mathbf{A}$ are given by $A_{kl}(p_i, q_j)$, which we determine below. Here, a pair of indices (i, j) or (k, l) is regarded as a single index running from 1 to N. Substituting (5) into (6) and requiring terms with the same coefficient a_{mn} to be equal, we obtain a set of simultaneous linear equations for $A_{kl}(p_i, q_j)$:

$$\int_0^\infty dp_2^2 \int_{B(p_j,q_2)}^{A(p_j,q_2)} dp_2^2 \tilde{K}_s(p_i, q_j; p_2, q_2)F_{mn}(p_2, q_2) = \sum_{kl} A_{kl}(p_i, q_j)F_{mn}(p_{2k}, q_{2l}). \quad (8)$$

Equation (8) represents a matrix equation of the form $\mathbf{FA} = \mathbf{M}$, where the matrix elements of $\mathbf{M}$ are given by moment integrals on the left-hand side of Eq. (8). The moment integrals are of the Cauchy type integral and hence they are bounded. Interestingly, the coefficients $a_{mn}(S)$ need not be calculated at all. Significantly, N_p points selected in the p variable are now over the entire range $(0, \infty)$ and need not lie within the integration limits (B, A).

The numerical method described above is a generalization of the method known as "approximate product-integration" for the one-variable to the two-variable case.[8] The approximate product-integration method for the one-variable case is found to be convenient for the bound-state problem, and is used in calculation of three-particle bound-state energies described below.

The bound-state energies are values of S at which the determinant $|\mathbf{1} - \mathbf{A}|$ vanishes. Therefore we look for zeros of the determinant as a function of S. We write the integral part I of Eq. (3) as

$$I = \sum_l^{N_q} W_l(p_i, q_j) \times \int_{B(q_j,q_{2l})}^{A(q_j,q_{2l})} dp_2^2 \tilde{K}_s(p_i, q_j; p_2, q_{2l})\tilde{\psi}_s(p_2, q_{2l}),$$

where W_l are the weights for Gaussian quadrature. This expansion is possible because for $S < S_0$ the kernel $\tilde{K}_s$ has no singularities. Otherwise, we must use the two-variable method described previously. For the p_2 integration we cannot use a standard quadrature method, because the integration limits are not constants. However, since $\tilde{\psi}_s(p_2, q_2)$ satisfies the condition $H(\mu)$ for the variable p_2 uniformly with respect to q_2, we can expand it as

$$\tilde{\psi}_s(p_2, q_{2l}) = \sum_m^{N_p} b_m(q_{2l})f_m(p_2),$$

and write the integral part I as

$$I = \sum_l^{N_q} W_l(p_i, q_j) \sum_k^{N_p} W_k(p_i, q_j; q_{2l})\tilde{\psi}_s(p_{2k}, q_{2l}),$$

where the W_k are to be determined by

$$\int_{B(q_j,q_{2l})}^{A(q_j,q_{2l})} dp_2^2 \tilde{K}_s(p_i, q_j; p_2, q_{2l})f_m(p_2) = \sum_k^{N_p} W_k(p_i, q_j; q_{2l})f_m(p_{2k}),$$

which is again a matrix equation. The matrix elements of $\mathbf{A}$ in Eq. (7) are now given by $W_l(p_i, q_j)W_k(p_i, q_j; q_{2l})$.

The three-particle bound-state energies are calculated as a function of the interaction strength λ for a local Yukawa potential $V(r) = \lambda e^{-r}/r$ (we set $\hbar = M = 1$, where M is the rest mass of the particle). Gaussian quadrature method is used to approximate the integral part of the two-body Lippmann–Schwinger equation, and the two-body t matrix is computed numerically by inverting the resultant matrix equation. The linearly independent functions $f_m(p)$ are chosen to be

$$f_m(p) = \frac{1}{p^2 + \alpha}\left(\frac{p^2}{p^2 + \alpha}\right)^m, \quad m = 0, 1, 2, \cdots, N_p - 1,$$

where α is a parameter taken to be 2. To test the accuracy of the solution, N_q is varied from 2 to 10, and N_p from 2 to 5. With $N_q = 10$, the bound-state energies calculated with $N_p = 4$ and 5 agree within 0.1%. Table I shows the convergence of the method

TABLE I. Successive approximations for the binding energies of the $J = 0$ state of three spinless bosons computed from the s-state local Yukawa potential with different values of the interaction strength λ. The number of points N_q in the q variable is set equal to 10.

λ	binding energy $N_p = 3$	$N_p = 4$	$N_p = 5$
−1.6	−0.1089	−0.1090	−0.1090
−1.8	−0.2876	−0.2881	−0.2881
−2.0	−0.5466	−0.5472	−0.5472
−2.2	−0.8880	−0.8868	−0.8869
−2.4	−1.3152	−1.3081	−1.3079

[8] A. Young, Proc. Roy. Soc. (London) **A224**, 552, 561 (1954).

used here for the bound-state energies at different values of λ. The bound-state energy of 0.2881 at $\lambda = -1.8$ can be compared with two other independent results of 0.295 by Wong and Ball and 0.293 by Osborn.[9]

The approximate product-integration method for the two-variable case described previously is also applicable to the bound-state problem, as well as to scattering problems with $S > 0$ or $0 > S > S_0$. The success of the method is critically dependent on a clever choice of linearly independent functions $F_{mn}(p, q)$. Some advantages of the method, either for the one-variable case or for the two-variable case, are that all rapidly varying parts, including the Cauchy-type singularities in the kernel, can be included in the moment integrals and that discrete points chosen in the continuous variable p_2 need not lie within the integration limits. Finally, we make an important remark that, to reduce the Faddeev equation to a matrix equation, we must transform it into a form in which the inhomogeneous term satisfies the Hölder condition.

[9] T. A. Osborn, "Faddeev's Equations for Local Potentials," SLAC Report No. 79, Stanford University, Stanford, California, 1967.

ACKNOWLEDGMENT

The author is indebted to A. Tubis for stimulating discussions on this subject.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Possible Instability of the Cut-Contributions to the Ring Equations

M. BAUS
Faculty of Science, University of Brussels, Brussels, Belgium

(Received 27 December 1968)

The specifically relativistic cut-contributions to the ring equations of a relativistic plasma as discussed in a previous paper are considered in more detail. The question of principle—whether or not these contributions can give rise to specifically relativistic instabilities—is investigated with the aid of a simple nonequilibrium velocity distribution which leads to unstable cut-contributions, but not necessarily to unstable Landau-pole contributions.

I. INTRODUCTION

In a previous paper[1] we derived the kinetic equations in the ring approximation for a relativistic plasma. A remarkable feature of these equations is the coupling which exists between the kinetic equation for the particles and the kinetic equation for the energy density of the normal modes of the electromagnetic field through which the particles interact. Another distinguishing feature of these equations is the appearance of cut-contributions arising from cuts in the complex plane on which the frequency-dependent collision operator is defined. As shown in I, the origin of these contributions can be traced back to the velocity cut-off introduced by the relativistic particle dynamics. These specifically relativistic contributions to the kinetic equations have been discussed qualitatively in Sec. 5 of I. It is the purpose of this paper to consider them in more detail. However, because of the complexity of these terms, we restrict ourselves to the question of principle, whether or not these cut-contributions can give rise to a specifically relativistic instability. For this purpose we consider the following particle-kinetic equation, taking dielectric screening into account:

$$\partial_t \rho_0(\mathbf{p}_j; t) = -\int_0^\infty d\tau \int d\mathbf{p}_i \Omega^{-2} \sum_i \sum_{\mathbf{k}_i, \mathbf{k}_j} \partial_j \cdot \mathcal{F}^{j|i}_{\mathbf{k}_i, \mathbf{k}_j}(t) \times (\mathcal{F}^{j|i}_{-\mathbf{k}_i, -\mathbf{k}_j}(t-\tau) \cdot \partial_j + \mathcal{F}^{i|j}_{-\mathbf{k}_j, -\mathbf{k}_i}(t-\tau) \cdot \partial_i) \times \rho_0(\mathbf{p}_i, \mathbf{p}_j; t). \quad (1)$$

Equation (1) was derived in Ref. 2 [see Eq. (II.35)], where the relevant notations have been defined. It was shown in II that (1) is a particular contribution to the general relativistic ring equations of I. We recall that $\rho_0(\mathbf{p}_j; t)$ is the momentum distribution of particle j, Ω the (large) volume, ∂_n the gradient with respect to $\mathbf{p}_n$, whereas $\mathcal{F}^{j|i}_{\mathbf{k}_i, \mathbf{k}_j}(t)$ is a Fourier transform with respect to the positions of particle i and j of the average force due to i and acting on j. We observe now that it is sufficient to investigate $\mathcal{F}(t)$ itself

[1] M. Baus Physica (to be published), referred to as I.

[2] M. Baus, Physica (to be published), referred to as II.

used here for the bound-state energies at different values of λ. The bound-state energy of 0.2881 at $\lambda = -1.8$ can be compared with two other independent results of 0.295 by Wong and Ball and 0.293 by Osborn.[9]

The approximate product-integration method for the two-variable case described previously is also applicable to the bound-state problem, as well as to scattering problems with $S > 0$ or $0 > S > S_0$. The success of the method is critically dependent on a clever choice of linearly independent functions $F_{mn}(p, q)$. Some advantages of the method, either for the one-variable case or for the two-variable case, are that all rapidly varying parts, including the Cauchy-type singularities in the kernel, can be included in the moment integrals and that discrete points chosen in the continuous variable p_2 need not lie within the integration limits. Finally, we make an important remark that, to reduce the Faddeev equation to a matrix equation, we must transform it into a form in which the inhomogeneous term satisfies the Hölder condition.

[9] T. A. Osborn, "Faddeev's Equations for Local Potentials," SLAC Report No. 79, Stanford University, Stanford, California, 1967.

ACKNOWLEDGMENT

The author is indebted to A. Tubis for stimulating discussions on this subject.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Possible Instability of the Cut-Contributions to the Ring Equations

M. BAUS
Faculty of Science, University of Brussels, Brussels, Belgium

(Received 27 December 1968)

The specifically relativistic cut-contributions to the ring equations of a relativistic plasma as discussed in a previous paper are considered in more detail. The question of principle—whether or not these contributions can give rise to specifically relativistic instabilities—is investigated with the aid of a simple nonequilibrium velocity distribution which leads to unstable cut-contributions, but not necessarily to unstable Landau-pole contributions.

I. INTRODUCTION

In a previous paper[1] we derived the kinetic equations in the ring approximation for a relativistic plasma. A remarkable feature of these equations is the coupling which exists between the kinetic equation for the particles and the kinetic equation for the energy density of the normal modes of the electromagnetic field through which the particles interact. Another distinguishing feature of these equations is the appearance of cut-contributions arising from cuts in the complex plane on which the frequency-dependent collision operator is defined. As shown in I, the origin of these contributions can be traced back to the velocity cut-off introduced by the relativistic particle dynamics. These specifically relativistic contributions to the kinetic equations have been discussed qualitatively in Sec. 5 of I. It is the purpose of this paper to consider them in more detail. However, because of the complexity of these terms, we restrict ourselves to the question of principle, whether or not these cut-contributions can give rise to a specifically relativistic instability. For this purpose we consider the following particle-kinetic equation, taking dielectric screening into account:

$$\partial_t \rho_0(\mathbf{p}_j; t) = -\int_0^\infty d\tau \int d\mathbf{p}_i \Omega^{-2} \sum_i \sum_{\mathbf{k}_i, \mathbf{k}_j} \partial_j \cdot \mathcal{F}^{j|i}_{\mathbf{k}_i, \mathbf{k}_j}(t) \times (\mathcal{F}^{j|i}_{-\mathbf{k}_i, -\mathbf{k}_j}(t-\tau) \cdot \partial_j + \mathcal{F}^{i|j}_{-\mathbf{k}_j, -\mathbf{k}_i}(t-\tau) \cdot \partial_i) \times \rho_0(\mathbf{p}_i, \mathbf{p}_j; t). \quad (1)$$

Equation (1) was derived in Ref. 2 [see Eq. (II.35)], where the relevant notations have been defined. It was shown in II that (1) is a particular contribution to the general relativistic ring equations of I. We recall that $\rho_0(\mathbf{p}_j; t)$ is the momentum distribution of particle j, Ω the (large) volume, ∂_n the gradient with respect to $\mathbf{p}_n$, whereas $\mathcal{F}^{j|i}_{\mathbf{k}_i, \mathbf{k}_j}(t)$ is a Fourier transform with respect to the positions of particle i and j of the average force due to i and acting on j. We observe now that it is sufficient to investigate $\mathcal{F}(t)$ itself

[1] M. Baus Physica (to be published), referred to as I.

[2] M. Baus, Physica (to be published), referred to as II.

rather than (1), because it follows from (1) that any instability in $\mathcal{F}(t)$ will induce an instability in the kinetic equation (1). We recall that the time behavior of the averaged force $\mathcal{F}(t)$ is given by (see II.27, 32)

$$\int_C d\omega \frac{e^{-i(\omega-\mathbf{k}\cdot\mathbf{v})t}}{\omega - \mathbf{k}\cdot\mathbf{v}'} \times \{(\mathbf{k}\cdot\mathbf{v} - \omega)\mathsf{D}_{\mathbf{k}}(\omega)\cdot\mathbf{v}' - \mathbf{v}\cdot\mathsf{D}_{\mathbf{k}}(\omega)\cdot\mathbf{v}'\,\mathbf{k}\}, \quad (2a)$$

$$\mathsf{D}_{\mathbf{k}}^{-1}(\omega) = \omega^2\boldsymbol{\epsilon}_{\mathbf{k}}(\omega) - (c\,|\mathbf{k}|)^2\mathsf{T}, \quad (2b)$$

where C is the Laplace transform inversion contour, $\mathbf{v}$, $\mathbf{v}'$ are particle velocities, $\boldsymbol{\epsilon}_{\mathbf{k}}(\omega)$ a dielectric tensor, and T is the unit tensor in transverse $\mathbf{k}$ space (see I). We can still simplify the investigation by considering a spatially isotropic dielectric tensor, in which case we can separate the longitudinal and transverse modes (see I.29): $\boldsymbol{\epsilon} = \epsilon_L\mathsf{L} + \epsilon_T\mathsf{T}$ ($\mathsf{T} = 1 - \mathsf{L}$, $\mathsf{L} = \mathbf{kk}/k^2$). For simplicity we consider only the longitudinal contribution. In this case the time behavior of (2) and thus of $\mathcal{F}(t)$ is governed by

$$-\frac{1}{k^2}\mathbf{k}\int_C d\omega \frac{e^{-i(\omega-\mathbf{k}\cdot\mathbf{v})t}}{\omega(\omega - \mathbf{k}\cdot\mathbf{v}')}\frac{\mathbf{k}\cdot\mathbf{v}'}{\epsilon_L(\mathbf{k},\omega)}. \quad (3)$$

Here we are interested only in the time behavior of (3) resulting from the cut-contribution discussed in I. Using the contour (I.65c), in which case the complex ω plane is cut from $c\,|\mathbf{k}|$ to $c\,|\mathbf{k}| - i\infty$ and from $-c\,|\mathbf{k}|$ to $-c\,|\mathbf{k}| - i\infty$ (c = velocity of light), we can write the cut-contributions to (3) as

$$I(t) = \sum_{\epsilon=\pm1}\int_0^\infty d\gamma \frac{e^{-i(\epsilon v-\mathbf{k}\cdot\mathbf{v})t-\gamma t}}{(\epsilon v - \mathbf{k}\cdot\mathbf{v}' - i\gamma)(\epsilon v - i\gamma)} \times \left(\frac{1}{\epsilon_L(\mathbf{k},\epsilon v + 0 - i\gamma)} - \frac{1}{\epsilon_L(\mathbf{k},\epsilon v - 0 - i\gamma)}\right), \quad (4)$$

where $\mathcal{F}^{\text{cut}}(t) \sim I(t)$ and $v = c\,|\mathbf{k}|$. In order now to study the time behavior of $I(t)$, we observe that at $\omega = \pm v \mp 0 - i\gamma$, $\epsilon_L(\mathbf{k},\omega)$ is defined by analytic continuation. This analytic continuation is easily obtained (as is explained, e.g., in Ref. 3) when $\epsilon_L(\mathbf{k},\omega)$ is written as a Cauchy integral. For this purpose it is necessary to perform the change of variables from $\mathbf{p}$ to the velocities $\mathbf{v}$. As a result of these operations one obtains the following expression for $\epsilon_L(\mathbf{k},\omega)$:

$$\epsilon_L(\mathbf{k},\omega) = 1 + \sum_n \frac{4\pi e_n^2 d_n}{k^2}\int_{v_n<c} d\mathbf{v}_n \frac{1}{\omega - \mathbf{k}\cdot\mathbf{v}_n} h(\mathbf{k},\mathbf{v}_n;t) + \theta(-\operatorname{Im}\omega)\theta(c\,|\mathbf{k}| - |\operatorname{Re}\omega|)g(\mathbf{k},\omega), \quad (5)$$

where $\operatorname{Im}\omega \neq 0$ and where $\theta(x)$ is the Heaviside step function [$\theta(x) = 1$ for $x > 0$, $\theta(x) = 0$ for $x \leq 0$]. The function h is defined as (m_0 being the rest mass)

$$h(\mathbf{k},\mathbf{v};t) = \frac{\partial(\mathbf{p})}{\partial(\mathbf{v})}\left(\mathbf{k}\cdot\frac{\partial}{\partial\mathbf{p}}\rho_0(\mathbf{p};t)\right)_{\mathbf{p}\to\mathbf{v}} = \frac{m_0^2}{(1 - v^2/c^2)^2}\left(\mathbf{k}\cdot\frac{\partial}{\partial\mathbf{v}} - \frac{1}{c^2}\mathbf{k}\cdot\mathbf{v}\,\mathbf{v}\cdot\frac{\partial}{\partial\mathbf{v}}\right)\rho_0(\mathbf{v};t), \quad (6)$$

whereas g is given by

$$g(\mathbf{k},\omega) = -2\pi i\sum_n \frac{4\pi l_n^2 d_n}{k^3} \times \int_{v_n<c} d\mathbf{v}_n\delta(\hat{\mathbf{k}}\cdot\mathbf{v}_n - u)h(\mathbf{k},\mathbf{v}_n;t)\Big|_{u=\omega/k}. \quad (7)$$

The time behavior of (4) is thus given by

$$I(t) = \sum_{\epsilon=\pm1}\int_0^\infty d\gamma \frac{e^{-i(\epsilon v-\mathbf{k}\cdot\mathbf{v})t-\gamma t}}{(\epsilon v - \mathbf{k}\cdot\mathbf{v}' - i\gamma)(\epsilon v - i\gamma)} \times \frac{\epsilon g(\mathbf{k},\epsilon v - i\gamma)}{\epsilon_L(\mathbf{k},\epsilon v + 0 - i\gamma)\epsilon_L(\mathbf{k},\epsilon v - 0 - i\gamma)}, \quad (8)$$

which we rewrite finally as

$$I(t) = \sum_{\epsilon=\pm1}\int_0^\infty d\gamma f_\epsilon(\mathbf{k};\mathbf{v},\mathbf{v}';\gamma;t)e^{-\gamma t}g(\mathbf{k},\epsilon v - i\gamma;t) \quad (9)$$

with

$$f_\epsilon(\mathbf{k};\mathbf{v},\mathbf{v}';\gamma;t) = \frac{\epsilon e^{-i(\epsilon v-\mathbf{k}\cdot\mathbf{v})t}}{(\epsilon v - \mathbf{k}\cdot\mathbf{v}' - i\gamma)(\epsilon v - i\gamma)} \times \frac{1}{\epsilon_L(\mathbf{k},\epsilon v + 0 - i\gamma)\epsilon_L(\mathbf{k},\epsilon v - 0 - i\gamma)}. \quad (10)$$

We observe now that the time behavior of (9) is governed, apart from the oscillating factor of (10), by the γ behavior of $e^{-\gamma t}g(\epsilon v - i\gamma)$. When the velocity distribution is of the exponential type, its analytic continuation contained in g will exhibit damped and growing oscillations in γ, so that the γ integral in (9) is not necessary well behaved. As a result one can not, even for time-independent velocity distributions, commute the asymptotic time limit $t \to t_R \gg \omega_p^{-1}$ and the γ integration and conclude that (9) is a transient (here t_R and ω_p^{-1} are, respectively, the relaxation and collision time). The behavior of (9), however, can be investigated asymptotically with the aid of the method of steepest descent,[4] but, as will soon become clear, no general statement can be made about (9). This situation is in contrast with the

[3] R. Balescu, *Statistical Mechanics of Charged Particles* (Interscience Publ. Co., New York, 1963).

[4] Ph. Dennery and A. Krzywicki, *Mathematics for Physicists* (Harper and Row, Inc., New York, 1967).

beautiful Landau-stability analysis (see, e.g., Ref. 3). We will now investigate (9) with the aid of a particular velocity distribution.

2. ASYMPTOTIC EVALUATION OF THE CUT-CONTRIBUTION FOR A PARTICULAR VELOCITY DISTRIBUTION

As far as the question of principle—whether or not (9) can become unstable—is concerned, it will be sufficient to evaluate (9) asymptotically ($t \to t_R$) with the method of steepest descent for a particular, but not too unrealistic, velocity distribution. Below, it will be seen that the instability condition as investigated along these lines requires that the analytic continuation of the velocity distribution contain at least three independent coefficients in γ space. This excludes Gaussian distributions and leads definitively to cubic nonequilibrium distributions. Moreover, in order to have simple computations, we will use spherical coordinates in velocity space in agreement with the symmetry of the integration domain ($v < c$), thereby avoiding the velocity-dependent integration limits encountered with other coordinate systems. However, in order to keep the Cauchy integral structure in (5), we delay the angle integrations. Let us take now a three-parameter spherically symmetric velocity distribution, e.g., a Gaussian with a cubic correction:

$$\rho_0(\mathbf{v}; t) \sim \exp(\alpha v^2 + \beta(v - \delta)^3). \tag{11}$$

For simplicity we rewrite (11) as

$$\rho_0(\mathbf{v}; t) \sim \exp(av + bv^2 + dv^3). \tag{12}$$

The cutoff of (12) for $v \geq c$ will be accounted for by taking a finite integration domain ($v < c$). Moreover, notice that, because of the finite integration domain ($v < c$), the normalization of (12) does not restrict the sign of the parameters a, b, d. Using now $\mathbf{k} \cdot \mathbf{v} = kv\mu$ ($\mu = \cos\theta$), we can rewrite (7) for (12) as

$$g(\mathbf{k}, \omega) = -2\pi i \sum_n \frac{4\pi e_n^2 d_n}{k^2} \int_{-1}^{1} d\mu_n \frac{2\pi m_{0n}^2}{1 - \omega^2/k^2\mu_n^2} \times k\mu_n \left(a + 2b\frac{\omega}{k\mu_n} + 3d\frac{\omega^2}{k^2\mu_n^2}\right) \rho_0\left(\frac{\omega}{k\mu_n}\right). \tag{13}$$

If we commute the μ and γ integration, we can rewrite (9) as

$$I(t) = \sum_\epsilon \sum_n \int_{-1}^{1} d\mu_n \mathcal{C}(t). \tag{14}$$

We recollect all exponential terms and write

$$\mathcal{C}(\tau^{\frac{1}{2}}) = \tau^{\frac{1}{2}} \int_0^\infty d\beta k(\beta) e^{\tau h(\beta)}, \tag{15}$$

where we have introduced the new variables $\tau = t^2$; $\beta = \gamma t^{-1}$. From (9), (12), and (13) we have

$$\begin{aligned} h(\beta) &= (-1 - ib_1)\beta - b_2\beta^2 + ib_3\beta^3, \\ b_1 &= (a/k\mu t) + (2b\epsilon v/k^2\mu^2 t) + (3v^2 d/k^3\mu^3 t), \\ b_2 &= (b/k^2\mu^2) + (3\epsilon v d/k^3\mu^3), \\ b_3 &= dt/k^3\mu^3. \end{aligned} \tag{16}$$

For large τ ($t \gg \omega_p^{-1}$) the main contribution to (15) will come from the saddle point

$$\frac{d}{d\beta} h(\beta) = 0, \quad \beta = \beta_\pm \equiv \beta'_\pm + i\beta''_\pm,$$

$$\beta'_\pm = \pm \frac{[(b_2^2 - 3b_1b_3)^2 + 9b_3^2]^{\frac{1}{4}}}{3b_3} \cdot \sin\frac{\theta}{2},$$

$$\beta''_\pm = -\frac{b_2}{3b_3} \mp \frac{[(b_2^2 - 3b_1b_3)^2 + 9b_3^2]^{\frac{1}{4}}}{3b_3} \cdot \cos\frac{\theta}{2}, \tag{17}$$

with

$$\tan\theta = 3b_3/(b_2^2 - 3b_1b_3). \tag{18}$$

The paths of steepest descent are given by

$$\mathrm{Im}\,[h(\beta) - h(\beta_\pm)] = 0 \tag{19}$$

and consist of two perpendicular straight lines through $\beta_\pm$ (degenerate hyperbola). The original contour of (15) can be deformed into the path of steepest descent if the real part of the saddle point is positive:

$$\beta'_\pm > 0. \tag{20}$$

On the other hand, as (12) does not vanish automatically when $v \geq c$, we can keep the original definition of the velocity distribution as long as we require the imaginary part of the saddle point to be negative:

$$\beta''_\pm < 0. \tag{21}$$

When (20) and (21) are fulfilled, (15) will behave like $\tau^{\frac{1}{2}} e^{\tau h(\beta_\pm)}$ and become unstable (growing oscillations) when

$$\mathrm{Re}\, h(\beta_\pm) \geq 0. \tag{22}$$

We now investigate the three conditions (20)–(22) as a function of the three parameters b_1, b_2, b_3 or a, b, d [see Eq. (16)]. From (16) we have, for (22),

$$\mathrm{Re}\, h(\beta_\pm) \equiv (-\beta'_\pm + b_1\beta''_\pm) + b_2(\beta''^2_\pm - \beta'^2_\pm) + b_3\beta''_\pm(\beta''^2_\pm - 3\beta'^2_\pm) \geq 0. \tag{23}$$

It is seen that $\mathrm{Re}\, h(\beta_+) > 0$ is trivially satisfied when

$$b_1 < 0, \tag{24}$$

$$b_2 > 0, \tag{25}$$

$$b_3 > 0, \tag{26}$$

$$1 < \frac{1}{b_1^2} \leq \frac{\beta''^2_+}{\beta'^2_+} < 3, \tag{27}$$

whereas (24)–(26) imply (20) and (21) on account of (18) for β_+. Clearly (24)–(27) is only a particular solution which guarantees that each term of (23) is positive. For any given $t \gg \omega_p^{-1}$ we can now look for the regions of a, b, d, k, and μ satisfying, for example, (24)–(27) and in these regions[5] $I(t)$ [(14)–(15)] will grow at least as t. [The exponential could eventually tend to a constant, but cannot decay because of (23).] This shows that, in principle, the cut-contribution (14) can exhibit instabilities. However, it should be noticed that this does not necessarily imply an instability for the kinetic equation (1), because the k, μ regions where the conditions [e.g., (24)–(27)] are satisfied can be of no importance with respect to the k, μ, and v integrations which we still have to perform.

3. LANDAU STABILITY AND CUT-CONTRIBUTION

Let us now look under what conditions the cut-instability will be specifically relativistic, i.e., will not be accompanied by unstable Landau-pole contributions. Using (12) in (5) for $\operatorname{Im}\omega \equiv \omega'' > 0$, we can write

$$\epsilon_L(\mathbf{k}, \omega' + i\omega'') = 1 + \sum_n \frac{4\pi e_n^2 d_n}{k^2} \int_{v_n<c} d\mathbf{v}_n \frac{m_{0n}^2}{1 - v_n^2/c^2} \mathbf{k}\cdot\hat{\mathbf{v}}_n \times \frac{(\omega' - \mathbf{k}\cdot\mathbf{v}_n) - i\omega''}{(\omega' - \mathbf{k}\cdot\mathbf{v}_n)^2 + \omega''^2}(a + 2bv_n + 3dv_n^2)\rho_0(v_n). \tag{28}$$

Using the fact that for a zero of (28) $\operatorname{Im}\epsilon_L = 0$, following the usual procedure[6] we write

$$\operatorname{Re}\epsilon_L(\mathbf{k}, \omega' + i\omega'') = 1 + \sum_n \frac{4\pi e_n^2 d_n}{k^2} \int_{v_n<c} d\mathbf{v}_n \frac{m_{0n}^2}{1 - v_n^2/c^2} \times \frac{-\mathbf{k}\cdot\hat{\mathbf{v}}\,\mathbf{k}\cdot\mathbf{v}_n}{(\omega' - \mathbf{k}\cdot\mathbf{v}_n)^2 + \omega''^2}(a + 2bv_n + 3dv_n^2)\rho_0(v_n). \tag{29}$$

Therefore the instability condition $\operatorname{Re}\epsilon_L = 0$, $\omega'' > 0$ cannot be satisfied when

$$a + 2bv_n + 3dv_n^2 < 0 \quad \text{for} \quad 0 < v_n < c, \tag{30}$$

i.e., when the roots of

$$3dv_n^2 + 2bv_n + a = 0 \tag{31}$$

are complex or lie outside the domain $(0, c)$.

The regions of a, b, d satisfying (30) together with, e.g., (24)–(27) will thus lead to a specifically relativistic cut-instability for which no traditional pole instability exists.

4. CONCLUSIONS

The cut-contributions to the kinetic equations for a relativistic plasma introduced in I have been investigated in more detail. It has been shown that, in principle, these contributions can become significant for not too unrealistic nonequilibrium situations and yield a specifically relativistic instability which is not necessarily accompanied by a traditional Landau-pole instability. However, a more detailed study of this type of instability is still required.

ACKNOWLEDGMENTS

We would like to thank Professor I. Prigogine and Professor R. Balescu for the interest they took in this work.

[5] As an example one can check that (27) can be satisfied for any positive b_3 when $b_2 \to 0$ and with $-1 < b_1 < -\frac{1}{3}$ or by $-\sqrt{3}/3 < b_1 < 0$ when $b_3 \to +\infty$ for any positive b_2.

[6] J. D. Jackson, J. Nucl. Energy 1C, 171 (1960).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Bounds Associated with Integral Equations of Kirkwood–Riseman Type

N. ANDERSON, A. M. ARTHURS, AND P. D. ROBINSON
Department of Mathematics, University of York, England

(Received 25 November 1968)

Variational principles are used to find approximate solutions of integral equations arising in the Kirkwood–Riseman theory of the properties of flexible macromolecules. Upper and lower bounds are obtained for the translational diffusion constant.

I. INTRODUCTION

Integral equations of the form

$$\varphi(x) = f(x) + \lambda \int_{-1}^{1} |x - t|^{-\alpha} \varphi(t)\, dt, \qquad (1)$$

where $0 < \alpha < 1$ and λ is negative, arise in the Kirkwood–Riseman theory[1] of intrinsic viscosities and diffusion constants of flexible macromolecules. Numerical solutions of this kind of equation have been obtained by Ullman[2,3] and by Schlitt.[4] In this paper complementary variational principles for integral equations[5] are used to find approximate solutions for equations of the form (1). When $f(x) = 1$, these principles lead to upper and lower bounds for the translational diffusion constant.

II. COMPLEMENTARY BOUNDS

Consider the Fredholm integral equation

$$\varphi(x) = f(x) + \lambda K\varphi(x), \qquad (2)$$

where

$$K\varphi(x) = \int \mathcal{K}(x, t)\varphi(t)\, dt, \qquad (3)$$

in which the kernel $\mathcal{K}(x, t)$ is symmetric and positive definite. If λ is negative, it follows from the theory of Sec. 3 of Ref. 5 that the complementary bounds

$$G(\Phi) \leq I(\varphi) \leq J(\Psi) \quad (\lambda < 0) \qquad (4)$$

hold, where

$$I(\varphi) = \int f(x)\varphi(x)\, dx, \qquad (5)$$

$$G(\Phi) = \int \{2f\Phi - \Phi^2 + \Phi\lambda K\Phi\}\, dx, \qquad (6)$$

$$J(\Psi) = \int \{(f + \lambda K\Psi)^2 - \Psi\lambda K\Psi\}\, dx. \qquad (7)$$

Here φ is the exact solution of the integral equation (2), and Φ and Ψ are any trial functions. The nearer Φ and Ψ are to φ, the closer the bounds G and J are to I and to each other.

Since

$$\mathcal{K}(x, t) = |x - t|^{-\alpha}, \quad -1 \leq x, t \leq 1, \qquad (8)$$

is symmetric and positive definite (this latter property following from the analysis of Auer and Gardner[6]), these results can be applied to the integral equation (1).

III. CALCULATIONS

With the kernel (8), calculations have been performed for two cases: (i) $f(x) = 1$, and (ii) $f(x) = x^2$. In each case the trial functions used were

$$\Phi = A_1x^2 + B_1, \quad \Psi = A_2x^2 + B_2, \qquad (9)$$

the parameters A_1, B_1, A_2, B_2 being found from the stationary conditions

$$\frac{\partial G}{\partial A_1} = 0, \quad \frac{\partial G}{\partial B_1} = 0, \quad \frac{\partial J}{\partial A_2} = 0, \quad \frac{\partial J}{\partial B_2} = 0, \qquad (10)$$

which optimize G and J. Since the same functional form is employed for both Φ and Ψ, the closeness of the optimum parameter values measures the accuracy of approximate solutions of this nature, as does the closeness of G and J also. In Tables I and II we give the results for a range of λ values and for $\alpha = 0.2, 0.5$, and 0.8.

When $f(x) = 1$, we see from Eq. (5) that

$$I(\varphi) = \int_{-1}^{1} \varphi(x)\, dx, \qquad (11)$$

and this by the Kirkwood–Riseman theory[1] is inversely proportional to the translational diffusion constant D; that is,

$$\int_{-1}^{1} \varphi(x)\, dx = c/D, \qquad (12)$$

where $c = 2M_0kT/M\zeta$, in which kT is the Boltzmann energy, ζ is a friction constant characteristic of the

[1] J. G. Kirkwood and J. Riseman, J. Chem. Phys. **16**, 565 (1948).
[2] R. Ullman, J. Chem. Phys. **40**, 2193 (1964).
[3] N. Ullman and R. Ullman, J. Math. Phys. **7**, 1743 (1966).
[4] D. W. Schlitt, J. Math. Phys. **9**, 436 (1968).
[5] P. D. Robinson and A. M. Arthurs, J. Math. Phys. **9**, 1364 (1968).
[6] P. L. Auer and C. S. Gardner, J. Chem. Phys. **23**, 1545 (1955).

TABLE I. Variational parameters and bounds: $f = 1$.

$-\lambda$	A_1	B_1	Lower bound G	A_2	B_2	Upper bound J	$(J-G)/J$
			(a) $\alpha = 0.2$[a]				
2	0.06322	0.15079	0.34372	0.06070	0.15160	0.34378	0.00017
4	0.05227	0.07675	0.18834	0.04981	0.07754	0.18843	0.00046
8	0.03685	0.03722	0.99004(−1)	0.03485	0.03786	0.99132(−1)	0.00129
16	0.02281	0.01781	0.50834(−1)	0.02145	0.01825	0.51002(−1)	0.00329
32	0.01289	0.00859	0.25772(−1)	0.01207	0.00885	0.25966(−1)	0.00747
64	0.00689	0.00419	0.12977(−1)	0.00643	0.00434	0.13189(−1)	0.01610
128	0.00356	0.00207	0.65118(−2)	0.00332	0.00215	0.67342(−2)	0.03303
256	0.00181	0.00103	0.32618(−2)	0.00169	0.00107	0.34894(−2)	0.06523
512	0.00091	0.00051	0.16324(−2)	0.00085	0.00053	0.18628(−2)	0.12370
1024	0.00046	0.00026	0.81656(−3)	0.00043	0.00027	0.10484(−2)	0.22112
			(b) $\alpha = 0.5$				
2	0.05529	0.09956	0.23598	0.05227	0.10047	0.23660	0.00262
4	0.03441	0.05131	0.12557	0.03241	0.05192	0.12636	0.00627
8	0.01943	0.02596	0.64880(−1)	0.01826	0.02632	0.65794(−1)	0.01389
16	0.01036	0.01304	0.32994(−1)	0.00972	0.01324	0.33978(−1)	0.02896
32	0.00535	0.00653	0.16639(−1)	0.00502	0.00664	0.17663(−1)	0.05798
64	0.00272	0.00327	0.83556(−2)	0.00255	0.00332	0.94004(−2)	0.11114
128	0.00137	0.00164	0.41868(−2)	0.00129	0.00166	0.52424(−2)	0.20136
256	0.00069	0.00082	0.20956(−2)	0.00065	0.00083	0.31566(−2)	0.33612
512	0.00035	0.00041	0.10484(−2)	0.00032	0.00042	0.21120(−2)	0.50379
1024	0.00017	0.00020	0.52436(−3)	0.00016	0.00021	0.15893(−2)	0.96701
			(c) $\alpha = 0.8$				
2	0.01171	0.04591	0.99618(−1)	0.01148	0.04598	0.99698(−1)	0.00080
4	0.00623	0.02347	0.51088(−1)	0.00611	0.02350	0.51178(−1)	0.00176
8	0.00321	0.01187	0.25876(−1)	0.00315	0.01189	0.25972(−1)	0.00367
16	0.00163	0.00597	0.13023(−1)	0.00160	0.00598	0.13122(−1)	0.00757
32	0.00082	0.00299	0.65328(−2)	0.00081	0.00300	0.66338(−2)	0.01523
64	0.00041	0.00150	0.32718(−2)	0.00041	0.00150	0.33736(−2)	0.03018
128	0.00021	0.00075	0.16372(−2)	0.00020	0.00075	0.17394(−2)	0.05877
256	0.00010	0.00037	0.81894(−3)	0.00010	0.00038	0.92138(−3)	0.11118
512	0.00005	0.00019	0.40956(−3)	0.00005	0.00019	0.51208(−3)	0.20020
1024	0.00003	0.00009	0.20480(−3)	0.00003	0.00009	0.30738(−3)	0.33372

[a] Here $N(-m)$ means $N \times 10^{-m}$.

TABLE II. Variational parameters and bounds: $f = x^2$.

$-\lambda$	A_1	B_1	Lower bound G	A_2	B_2	Upper bound J	$(J-G)/J$
			(a) $\alpha = 0.2$				
0.1	0.98363	−0.05753	0.35510	0.98246	−0.05728	0.35512	0.00006
0.5	0.91128	−0.14787	0.26594	0.90668	−0.14659	0.26618	0.00090
1.0	0.82827	−0.17270	0.21618	0.82101	−0.17053	0.21676	0.00268
2.0	0.69756	−0.16961	0.16595	0.68763	−0.16654	0.16721	0.00752
4.0	0.52877	−0.14022	0.11803	0.51767	−0.13672	0.12047	0.02025
8.0	0.35590	−0.09886	0.76454(−1)	0.34603	−0.09572	0.80492(−1)	0.05017
			(b) $\alpha = 0.5$				
0.1	0.90868	−0.05610	0.32608	0.90548	−0.05530	0.32618	0.00031
0.5	0.65125	−0.09346	0.19819	0.64379	−0.09141	0.19934	0.00574
1.0	0.47755	−0.08213	0.13627	0.46983	−0.07995	0.13850	0.01608
2.0	0.31069	−0.05932	0.84730(−1)	0.30432	−0.05749	0.88284(−1)	0.04026
4.0	0.18270	−0.03691	0.48472(−1)	0.17838	−0.03566	0.53206(−1)	0.08897
8.0	0.10014	−0.02084	0.26164(−1)	0.09757	−0.02009	0.31736(−1)	0.17557
			(c) $\alpha = 0.8$				
0.1	0.62552	−0.03356	0.22784	0.62443	−0.03343	0.22808	0.00105
0.5	0.24862	−0.02215	0.84682(−1)	0.24781	−0.02202	0.85544(−1)	0.01008
1.0	0.14169	−0.01376	0.47504(−1)	0.14118	−0.01368	0.48588(−1)	0.02231
2.0	0.07616	−0.00774	0.25302(−1)	0.07587	−0.00770	0.26532(−1)	0.04636
4.0	0.03956	−0.00412	0.13079(−1)	0.03940	−0.00409	0.14393(−1)	0.09130
8.0	0.02017	−0.00213	0.66520(−2)	0.02009	−0.00211	0.80116(−2)	0.16970

fluid, M is the molecular weight of the polymer unit, and M_0 is the molecular weight of the monomer unit. Since G and J provide complementary bounds for $I(\varphi)$, it follows that c/G and c/J obtained from Table I provide upper and lower bounds for D, the translational diffusion constant. Thus

$$c/J \leq D \leq c/G \quad (f = 1). \tag{13}$$

When $f(x) = x^2$, the functional $I(\varphi)$ in (5) has no direct significance, but other writers[3,4] have also considered this case.

IV. DISCUSSION

Previous numerical solutions of Eq. (1) have been concerned with the case $\alpha = 0.5$, which corresponds to a Gaussian model for the statistics of the polymer chain. It is to be expected that statistics other than Gaussian will lead to values of α lying somewhere between zero and unity. Using the variational method of this paper, we have obtained results which enable us to compare the solutions for Gaussian and non-Gaussian models.

Judging by the closeness of the bounds G and J, and by the closeness of the parameters A_1, A_2 and B_1, B_2, we see from Tables I and II that the solutions are quite accurate for the smaller λ values shown. With these values in case (i) corresponding to $f(x) = 1$, we are able to place close upper and lower limits on values for the diffusion constant D. For fixed values of λ in the range -2 to -256, we see from Table I and Eq. (13) that the diffusion constant D increases as α increases. For the larger λ values, some variation of accuracy occurs as α changes, the least accuracy being recorded for $\alpha = 0.5$.

TABLE III. Comparison between results of Ullman and this paper for $f(x) = x^2$, $\alpha = 0.5$, $\lambda = -0.5$.

x	ϕ (Ref. 3)	Φ	Ψ
0.019511	−0.081022	−0.093212	−0.09116
0.116084	−0.073681	−0.08468	−0.08273
0.227786	−0.052520	−0.05967	−0.05801
0.413779	0.015277	0.01804	0.01882
0.502804	0.062386	0.07118	0.07135
0.636054	0.15394	0.17001	0.16904
0.778306	0.28370	0.30104	0.29857
0.912234	0.45436	0.44849	0.44433
0.999554	0.65425	0.55721	0.55181

In case (ii), corresponding to $f(x) = x^2$, some comparison with previous work is possible. Table III contains the numerical solution of Ullman and Ullman[3] for $\alpha = 0.5$, $\lambda = -0.5$, and our variational solutions Φ and Ψ, which for this case are given by

$$\Phi = (0.65125)x^2 - 0.09346, \tag{14}$$

$$\Psi = (0.64379)x^2 - 0.09141. \tag{15}$$

While these variational solutions are accurate to 0.57% in terms of the bounds G and J, we see from Table III that the agreement with the numerical solution is only moderately good. This, however, must be attributed to the extreme simplicity of the trial functions used here. More elaborate trial functions, for example $\Phi = Ax^4 + Bx^2 + C$, will undoubtedly lead to more accurate variational solutions and closer bounds. Even in a situation like this when the bounds are not of direct physical interest, the advantage of an approximate analytical solution valid for all x is worth stressing.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Schwarzschild Singularity

LUIS BEL
Laboratoire de Physique Théorique associé au C.N.R.S., Institut H. Poincaré, Paris

(Received 21 March 1968)

A new point of view is presented for which the Schwarzschild singularity becomes a real point singularity on which the sources of Schwarzschild's exterior solution are localized.

INTRODUCTION

The exterior Schwarzschild metric may be written in polar harmonic coordinates as[1]

$$ds^2 = -\frac{r-\alpha}{r+\alpha}c^2\,dt^2 + \frac{r+\alpha}{r-\alpha}dr^2 + (r+\alpha)^2\,d\omega^2,$$
$$d\omega^2 = d\theta^2 + \sin^2\theta\,d\phi^2, \qquad (1)$$
$$\alpha = GM/c^2.$$

By polar harmonic coordinates is meant that t, $x^1 = r\sin\theta\cos\phi$, $x^2 = r\sin\theta\sin\phi$, and $x^3 = r\cos\theta$ are harmonic functions. These coordinates $\{t, x^i\}$ are admissible only for $r > \alpha$.

Most people feel that the singularity of (1) as $r \to \alpha$—the Schwarzschild singularity—is not an intrinsic one, since the only curvature invariant K, of which every other invariant is a regular function, is $K = m/(r+\alpha)^3$ and does not display any special behavior at $r = \alpha$. One may then ask whether or not the metric (1) can be "extended" regularly. Actually, several extensions have been proposed in the literature, the most commonly quoted being those of Finkelstein[2] and Kruskal.[3] Both extensions lead to space–time models which are not globally static and are consequently inadequate for representing the exterior solution of a source in static equilibrium.

In this paper we aim to show that, besides the extension point of view, there is another one for which the Schwarzschild singularity becomes instead a real point singularity on which are localized the sources of the exterior static solution. For anyone who accepts it, this point of view eliminates any speculation about the physics of objects supposed to have collapsed beyond the Schwarzschild singularity.

1. THE SPACE MANIFOLD $\bar{V}_3$

(a) The space metric which corresponds to (1) could be, of course,

$$d\bar{s}^2 = \frac{r+\alpha}{r-\alpha}dr^2 + (r+\alpha)^2\,d\omega^2, \qquad (2)$$

but it could also be

$$d\bar{s}^2 = dr^2 + (r^2-\alpha^2)\,d\omega^2. \qquad (3)$$

($d\bar{s}^2 = \xi^2\,ds^2$, ξ^2 are, up to a sign, the square of the generator of the group of timelike motions.) Corresponding to the coordinates which we are using, $\xi^2 = (r-\alpha)/(r+\alpha)$. Several local properties of (3) as compared to the corresponding ones of (2) have been discussed by Fock,[1] Ehlers and Kund,[4] and Bel and Escard.[5]

We choose (3) as the metric of the space. This choice is essential in the following construction of the space manifold $\bar{V}_3$.

(b) Let $\bar{V}_3^*$ be the open submanifold of R_3 defined by $x \in \bar{V}_3^*$ if $r > \alpha$. $d\bar{s}^2$ defines a distance d^* onto $\bar{V}_3^*$ by

$$d^*(x_1, x_2) = \inf\int_{l(x_1,x_2)} d\bar{s}, \quad x_1, x_2 \in \bar{V}_3^*, \qquad (4)$$

$l(x_1, x_2)$ being the set of all piecewise-differentiable paths joining x_1 and x_2.

The metric space $\bar{V}_3^*(d^*)$ is not complete. But one can prove very easily the following result:

The completion $\bar{V}_3(d)$ of $\bar{V}_3^*(d^*)$ contains only one additional point. Let x_0 be this point. Then $d(x_0, x) = r - \alpha \mid (x \in \bar{V}_3^*)$.

The steps of the proof are the following. Every Cauchy sequence x_n of $\bar{V}_3^*(d^*)$ is either of type (1) or of type (2). x_n is of type (1) if there exists k such that for $n > k$, $r_n > a > \alpha$. Every Cauchy sequence of this type converges to a point $x \in \bar{V}_3^*$. x_n is of type (2) if for every $\epsilon > 0$ there exist n such that $r_n - \alpha < \epsilon$. No Cauchy sequence of this type converges.

If x_m and x_n' are any two points of $\bar{V}_3^*$, one has that

$$d^*(x_m, x_n') \leq |r_m - r_n'| + \pi(r_n'^2 - \alpha^2)^{\frac{1}{2}}. \qquad (5)$$

From this inequality it follows that, if x_m and x_n' are two Cauchy sequences of type (2),

$$\lim_{n=m\to\infty} d^*(x_m, x_n') = 0.$$

[1] V. Fock, *The Theory of Space-Time and Gravitation* (Pergamon Press, Inc., New York, 1964).
[2] D. Finkelstein, Phys. Rev. **110**, 965 (1958).
[3] M. Kruskal, Phys. Rev. **119**, 1743 (1960).
[4] J. Ehlers and W. Kund, *Contribution to Gravitation*, L. Witten, Ed. (John Wiley & Sons, Inc., New York, 1962).
[5] L. Bel and J. C. Escard, Rend. Atti. Accad. Nazl. Lincei. (VIII) **41**, 476 (1966).

This means that any two such sequences are equivalent. This proves the first part of our result.

From (5) with $x_m = x$ and from

$$d^*(x_1, x'_n) \geq r - r_n - \pi(r'^2_n - \alpha^2)^{\frac{1}{2}}, \qquad (6)$$

valid for x'_n of type (2) and n large enough, it follows

$$\lim_{n\to\infty} d^*(x_1, x'_n) = r - \alpha.$$

This proves the second part.

(c) $\bar{V}_3(d)$ is by construction a complete metric space to which we may add a differentiable structure compatible with that of V_3^*. We need only define a local chart (U, ϕ), U being, for example, an open ball $d(x_0, x) < b$ of center x_0 and ϕ a homeomorphism of U into a Euclidean ball of center 0 $\{y^1 = y^2 = y^3 = 0\}$. We define ϕ by

$$\phi: x_0 \to 0, \quad x \to y \equiv \{y^1 = u \sin\theta\cos\phi, \quad y^2 = u \sin\theta \sin\phi, y^3 = u\cos\theta\},$$
$$u = (3\alpha)^{\frac{1}{3}}(r - \alpha)^{\frac{2}{3}}. \qquad (7)$$

Let $\bar{g}_{\{y\}}$ be the determinant of $d\bar{s}^2$ corresponding to this local system of coordinates. Then,[6]

$$\bar{g}_{\{y\}} = (r + \alpha)/2\alpha.$$

The space manifold $\bar{V}_3$ can now be built from the union of two local charts (U, ϕ) and the original one as defined by the coordinates $\{x^i\}$. One can prove, then, the existence of a global coordinate system $\{z^i\}$ compatible with the asymptotic Euclidean behavior of $d\bar{s}^2$ and such that $\bar{g}_{\{z^i\}} = 1$ everywhere.

2. METRIC AND FIELD SINGULARITIES

We shall use in the neighborhood of the origin (point x_0) the coordinates $\{u, \theta, \phi\}$. The space metric is

$$d\bar{s}^2 = 3(4\alpha)^{-1}u\,du^2 + S^2(u)\,d\omega^2,$$
$$S^2 = (3\alpha)^{-1}u^{\frac{3}{2}}[u^{\frac{3}{2}} + (12\alpha^3)^{\frac{1}{2}}]. \qquad (8)$$

Consider the orthonormal cobasis: $\bar{\theta}^1 = (3u/4\alpha)^{\frac{1}{2}}$, $\bar{\theta}^2 = S\,d\theta$, $\bar{\theta}^3 = S\sin\theta\,d\phi$.

The strict components of the Rieman tensor which are not identically zero in this cobasis are

$$\bar{R}_{12,12} = \bar{R}_{31,31} = -\bar{R}_{23,23} = \alpha^2 S^{-4}. \qquad (9)$$

If $u \to 0$, these components tend to infinity. Consequently the point x_0 is singular with respect to the Riemannian structure of $\bar{V}_3$. It is also a singular point of the gravitational field[7,8] $\mathbf{E} = -c^2$ grad ln ξ, where

$$\xi = u^{\frac{3}{2}}/[u^{\frac{3}{2}} + (12\alpha^3)^{\frac{1}{2}}]^{\frac{1}{2}}. \qquad (10)$$

In fact, in the same cobasis one has

$$E_1 = -c^2\alpha S^{-2}, \quad E_2 = E_3 = 0. \qquad (11)$$

3. THE SOURCE OF THE EXTERIOR SOLUTION

(a) Let us consider a spherically symmetrical space–time which is globally static with metric (1) for $r > R > \alpha$ and

$${}_1d\bar{s}^2 = -{}_1\xi^2(\zeta^k)\,dt^2 + {}_1d\hat{s}^2, \quad {}_1d\hat{s}^2 = {}_1\hat{g}_{ij}(\zeta^k)\,ds^i\,ds^j,$$
$$\zeta^1 = r, \quad \zeta^2 = \theta, \quad \zeta^3 = \phi, \qquad (12)$$

for $0 \leq r \leq \alpha$. We assume that ${}_1d\bar{s}^2$ is of class C^3 and a solution of Einstein's interior field equations:

$$S_{\alpha\beta} = \chi T_{\alpha\beta}, \quad \chi = 8\pi G/c^2. \qquad (13)$$

We assume also that the global metric is C^1, piecewise C^3 (across the hypersurface $r = R$).

It is well known[9] that one of the field equations (14) may be written as

$${}_1\bar{\Delta}\,{}_1U = 4\pi G\mu, \quad {}_1U = c^2\log\xi, \quad \mu = {}_1\xi^{-2}(T^0_0 - T^i_i), \qquad (14)$$

where ${}_1\bar{\Delta}$ is the Laplacian operator corresponding to ${}_1d\bar{s}^2 = {}_1\xi^2\,{}_1d\hat{s}^2$. From (14) one can prove that

$$M = \int_{\bar{V}_3} \mu({}_1\bar{g})^{\frac{1}{2}}\,dr\,d\theta\,d\phi, \quad {}_1\bar{g} = \det({}_1\bar{g}_{ij}). \qquad (15)$$

(b) We now consider the exterior solution only. For any open set which does not contain x_0 we have, of course, $\bar{\Delta}U = 0$ $(U = c^2 \ln\xi)$. We wish to define $\bar{\Delta}U$ as a distribution. The function U is locally integrable using as volume element the volume element of $d\bar{s}^2$:

$$(\bar{g})^{\frac{1}{2}} = \frac{u^2}{2\alpha^{\frac{3}{2}}3^{\frac{1}{2}}}[u^{\frac{3}{2}} + (12\alpha^3)^{\frac{1}{2}}]\sin\theta.$$

It then defines a distribution

$$\langle U, f\rangle = \int_{\bar{V}_3} Uf(\bar{g})^{\frac{1}{2}}\,du\,d\theta\,d\phi, \quad f \in D^0(\bar{V}_3). \qquad (16)$$

$\bar{\Delta}U$ is now defined as a distribution by

$$\langle \Delta U, f\rangle = \lim_{\epsilon\to 0}\int_{\bar{V}_3 - B(\epsilon)} U\bar{\Delta}(\bar{g})^{\frac{1}{2}}\,du\,d\theta\,d\phi, \qquad (17)$$

where $B(\epsilon)$ is the open ball of center x_0 and radius ϵ. To calculate (18) we may use the generalization of Green's formula to Riemannian spaces and write

$$\langle \bar{\Delta}U, f\rangle = -\lim_{\epsilon\to 0}\int_{S_2(\epsilon)}\left[U\frac{\partial f}{\partial u} - f\frac{dU}{du}\right]\frac{4\alpha}{u}(\bar{g})^{\frac{1}{2}}\,d\theta\,d\phi,$$
$$S_2(\epsilon) = \partial B(\epsilon). \qquad (18)$$

[6] $\lim_{u\to 0}\bar{g}_{\{y\}} = 1$. If instead we had defined ϕ by $u = r - \alpha$, as seems most natural, we would have had $\lim_{u\to 0}\bar{g}_{\{y\}} = 0$.

[7] Justifications of this terminology can be found in Refs. 1, 4, 5, and 8.

[8] C. Cattaneo, Nuovo Cimento **10**, 318 (1958).

[9] See, for instance, Ref. 1.

Using the fact that

$$\lim_{\epsilon\to 0} \epsilon \log \epsilon = 0,$$

the calculation can be finished as in the Newtonian case and we get

$$\langle \bar{\Delta} U, f\rangle = 4\pi G M f(x_0)$$

or

$$\bar{\Delta} U = 4\pi G \delta_{x_0}. \tag{19}$$

We may say then that the source of the exterior solution is $M\delta_{x_0}$ and write symbolically, but consistent with (14) and (15),

$$M = \int_{V_3} M\delta_{x_0}(\bar{g})^{\frac{1}{2}}\, du\, d\theta\, d\phi. \tag{20}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Initial Conditions in General Relativity: Lapse and Shift Formulation

ELLIOT P. BELASCO* AND HANS C. OHANIAN†
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey

(Received 15 November 1967; Revised Manuscript Received 8 April 1969)

We examine the system of coupled differential equations to which the constraints on the Cauchy data reduce if expressed in terms of the "shift" vector N_k and "lapse" N_0. If ${}^{(3)}g_{ij}$ and $\partial\, {}^{(3)}g_{ij}/\partial t$ are given and Dirichlet boundary conditions are imposed, the solution N_k is found to be unique if $2 \times$ (energy density) $-$ (three-curvature) > 0, but need not be unique when this inequality is not satisfied. No general existence theorem is known, but we list some conditions which make solutions impossible.

I. INTRODUCTION

In the general theory of relativity the causal development of the metric field is completely determined, at least for some finite time, once the initial data are given on a spacelike hypersurface. The correct data are of the Cauchy type and, if the spacelike hypersurface is designated by $x^0 = 0$, this means we are to specify ${}^{(4)}g_{\mu\nu}$, ${}^{(4)}\dot{g}_{\mu\nu} \equiv \partial\, {}^{(4)}g_{\mu\nu}/\partial x^0$, and the energy–momentum tensor $T_{\mu\nu}$ at $x^0 = 0$. However, it is seen that, of the ten Einstein equations, the four equations

$$G^0_\mu(g_{\alpha\beta}, \dot{g}_{\alpha\beta}) = T^0_\mu \tag{1}$$

do not contain second time derivatives at all and therefore must be regarded as four constraints on the initial data.

In order to find a consistent set of initial data, one can try to proceed in this fashion: prescribe arbitrarily some of the $g_{\mu\nu}$ and $\dot{g}_{\mu\nu}$, then find those remaining by solving Eqs. (1) for them. Foures-Bruhat[1] has reviewed several ways of formulating this problem of initial conditions. We will be concerned with the approach discussed by Wheeler.[2–4] In this formulation one defines the "shift" vector N_k and the "lapse" N_0 by

$$ds^2 = g_{ik}\, dx^i\, dx^k + 2N_i\, dx^i\, dx^0 + ({}^{(3)}g^{ik}N_iN_k - N_0^2)(dx^0)^2. \tag{2}$$

Here ${}^{(3)}g_{ik}$ is the metric induced on the hypersurface $x^0 = 0$ (we will hereafter abbreviate $g_{ik} \equiv {}^{(3)}g_{ik}$). It can be shown that, within the hypersurface, N_k transforms like a 3-vector and N_0 like a scalar. These shift and lapse functions have a very simple geometrical interpretation: they determine how hypersurfaces near $x^0 = 0$ are to be constructed.

We now take g_{ik} and $\dot{g}_{ik}$ as prescribed and attempt to consider Eqs. (1) as four equations for the unknowns N_k and N_0. These equations can be written as

$$[(\gamma_{mn} - g_{mn}\gamma^s_s)/N_0]^{|m} = S_n, \tag{3}$$

$$N_0 = [\gamma^*/(2\epsilon - R)]^{\frac{1}{2}}, \tag{4}$$

where

$$\gamma_{rs} = \tfrac{1}{2}(N_{r|s} + N_{s|r} - \dot{g}_{rs}), \tag{5}$$

$$\gamma^* = (\gamma^s_s)^2 - \gamma_{rs}\gamma^{rs}. \tag{6}$$

---

* Present address: Massachusetts Institute of Technology, Cambridge, Massachusetts.

† Present address: Rensselaer Polytechnic Institute, Troy, New York.

[1] Y. Foures-Bruhat, in *Gravitation: An Introduction to Current Research*, L. Witten, Ed. (John Wiley & Sons Inc., New York, 1962), p. 130.

[2] R. F. Baierlein, D. H. Sharp, and J. A. Wheeler, Phys. Rev. **126**, 1864 (1962).

[3] J. A. Wheeler, in *Relativity, Groups and Topology*, C. DeWitt and B. DeWitt, Eds. (Gordon and Breach, Science Publishers, New York, 1964), p. 317.

[4] J. A. Wheeler, in *Gravitation and Relativity*, H. Y. Chiu and W. F. Hoffman, Eds. (W. A. Benjamin, Inc., New York, 1964), p. 303.

Using the fact that

$$\lim_{\epsilon\to 0} \epsilon \log \epsilon = 0,$$

the calculation can be finished as in the Newtonian case and we get

$$\langle \bar{\Delta} U, f\rangle = 4\pi G M f(x_0)$$

or

$$\bar{\Delta} U = 4\pi G \delta_{x_0}. \tag{19}$$

We may say then that the source of the exterior solution is $M\delta_{x_0}$ and write symbolically, but consistent with (14) and (15),

$$M = \int_{V_3} M\delta_{x_0}(\bar{g})^{\frac{1}{2}}\, du\, d\theta\, d\phi. \tag{20}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Initial Conditions in General Relativity: Lapse and Shift Formulation

ELLIOT P. BELASCO* AND HANS C. OHANIAN†
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey

(Received 15 November 1967; Revised Manuscript Received 8 April 1969)

We examine the system of coupled differential equations to which the constraints on the Cauchy data reduce if expressed in terms of the "shift" vector N_k and "lapse" N_0. If ${}^{(3)}g_{ij}$ and $\partial\, {}^{(3)}g_{ij}/\partial t$ are given and Dirichlet boundary conditions are imposed, the solution N_k is found to be unique if $2 \times$ (energy density) $-$ (three-curvature) > 0, but need not be unique when this inequality is not satisfied. No general existence theorem is known, but we list some conditions which make solutions impossible.

I. INTRODUCTION

In the general theory of relativity the causal development of the metric field is completely determined, at least for some finite time, once the initial data are given on a spacelike hypersurface. The correct data are of the Cauchy type and, if the spacelike hypersurface is designated by $x^0 = 0$, this means we are to specify ${}^{(4)}g_{\mu\nu}$, ${}^{(4)}\dot{g}_{\mu\nu} \equiv \partial\, {}^{(4)}g_{\mu\nu}/\partial x^0$, and the energy–momentum tensor $T_{\mu\nu}$ at $x^0 = 0$. However, it is seen that, of the ten Einstein equations, the four equations

$$G^0_\mu(g_{\alpha\beta}, \dot{g}_{\alpha\beta}) = T^0_\mu \tag{1}$$

do not contain second time derivatives at all and therefore must be regarded as four constraints on the initial data.

In order to find a consistent set of initial data, one can try to proceed in this fashion: prescribe arbitrarily some of the $g_{\mu\nu}$ and $\dot{g}_{\mu\nu}$, then find those remaining by solving Eqs. (1) for them. Foures-Bruhat[1] has reviewed several ways of formulating this problem of initial conditions. We will be concerned with the approach discussed by Wheeler.[2–4] In this formulation one defines the "shift" vector N_k and the "lapse" N_0 by

$$ds^2 = g_{ik}\, dx^i\, dx^k + 2N_i\, dx^i\, dx^0 + ({}^{(3)}g^{ik}N_iN_k - N_0^2)(dx^0)^2. \tag{2}$$

Here ${}^{(3)}g_{ik}$ is the metric induced on the hypersurface $x^0 = 0$ (we will hereafter abbreviate $g_{ik} \equiv {}^{(3)}g_{ik}$). It can be shown that, within the hypersurface, N_k transforms like a 3-vector and N_0 like a scalar. These shift and lapse functions have a very simple geometrical interpretation: they determine how hypersurfaces near $x^0 = 0$ are to be constructed.

We now take g_{ik} and $\dot{g}_{ik}$ as prescribed and attempt to consider Eqs. (1) as four equations for the unknowns N_k and N_0. These equations can be written as

$$[(\gamma_{mn} - g_{mn}\gamma^s_s)/N_0]^{|m} = S_n, \tag{3}$$

$$N_0 = [\gamma^*/(2\epsilon - R)]^{\frac{1}{2}}, \tag{4}$$

where

$$\gamma_{rs} = \tfrac{1}{2}(N_{r|s} + N_{s|r} - \dot{g}_{rs}), \tag{5}$$

$$\gamma^* = (\gamma^s_s)^2 - \gamma_{rs}\gamma^{rs}. \tag{6}$$

---

* Present address: Massachusetts Institute of Technology, Cambridge, Massachusetts.

† Present address: Rensselaer Polytechnic Institute, Troy, New York.

[1] Y. Foures-Bruhat, in *Gravitation: An Introduction to Current Research*, L. Witten, Ed. (John Wiley & Sons Inc., New York, 1962), p. 130.

[2] R. F. Baierlein, D. H. Sharp, and J. A. Wheeler, Phys. Rev. **126**, 1864 (1962).

[3] J. A. Wheeler, in *Relativity, Groups and Topology*, C. DeWitt and B. DeWitt, Eds. (Gordon and Breach, Science Publishers, New York, 1964), p. 317.

[4] J. A. Wheeler, in *Gravitation and Relativity*, H. Y. Chiu and W. F. Hoffman, Eds. (W. A. Benjamin, Inc., New York, 1964), p. 303.

R is the curvature scalar formed out of the 3-metric g_{ik} and ϵ and S_n are, respectively, the energy density and the density of energy flow (the precise definitions are given in Sec. IV). From Eqs. (3) and (4) we obtain the following condensed initial-value equations which only involve the vector N_k:

$$\{[(2\epsilon - R)/\gamma^*]^{\frac{1}{2}}(\gamma_{mn} - g_{mn}\gamma^s_s)\}^{|m} = S_n. \quad (7)$$

These last equations can be obtained from the action integral

$$J = \int \{-[\gamma^*(2\epsilon - R)]^{\frac{1}{2}} + S^k N_k\} g^{\frac{1}{2}}\, d^3x \quad (8)$$

by seeking the extremum with respect to variations of N_k.

The problem of finding consistent Cauchy data reduces to that of solving Eqs. (7) for the shift N_k, where ϵ, S_n, g_{mn}, and $\dot{g}_{mn}$ are given functions. The solution must satisfy the restriction

$$\gamma^*/(2\epsilon - R) > 0 \quad (9)$$

everywhere. It then becomes necessary to ask, Do the Eqs. (7) have one and only one solution whenever some appropriate boundary conditions are given? Sections II and III discuss this question. Section IV contains a few remarks about the source terms.

II. UNIQUENESS

The uniqueness problem for Eqs. (7) is of particular interest because of its relation to the Mach principle as formulated by Wheeler.[4] It is also relevant to the question of the absence of a gravitational analog to electric charge.[5] As domain for the solution of Eqs. (7) we take a region C of finite proper volume contained in the hypersurface $x^0 = 0$. We will suppose that C is the union of a finite number of oriented curvilinear tetrahedra, i.e., C is a 3-chain. This will make it possible to apply Gauss's theorem. We do not place any further restrictions on the topology of C; in particular, "wormholes" are not excluded.

We will consider boundary conditions of the Dirichlet type: the function N_k is to be prescribed on the boundary $d(C)$ of C. We then find that the solution is unique, provided that everywhere

$$2\epsilon - R > 0. \quad (10)$$

However, it may happen that the boundary $d(C)$ vanishes: for example, if the hypersurface $x^0 = 0$ is a 3-sphere and C is all of this 3-sphere. We will call such a space "closed." In this case no boundary conditions are imposed on N_k, but nevertheless the solution of Eq. (7) remains unique *up to Killing vectors*. Since such Killing vectors added to N_k are physically irrelevant, the shift is essentially unique. If Eq. (10) is not satisfied, we can show by means of an example that Dirichlet boundary conditions are insufficient to guarantee uniqueness.

The uniqueness theorem can be stated as follows:

Theorem: Suppose $2\epsilon - R > 0$. Whenever the equations

$$\{[(2\epsilon - R)/\gamma^*]^{\frac{1}{2}}(\gamma_{mn} - g_{mn}\gamma^s_s)\}^{|m} = S_n, \quad (11)$$

with

$$\gamma_{rs} = \tfrac{1}{2}(N_{r|s} + N_{s|r} - \dot{g}_{rs}), \quad (12)$$

$$\gamma^* = (\gamma^s_s)^2 - \gamma_{rs}\gamma^{rs} > 0, \quad (13)$$

have two solutions $N_k = V_k(x)$ and $N_k = W_k(x)$, taking the same boundary values N_s on $d(C)$, then the two solutions can differ at most by a Killing vector. If $d(C) \neq 0$, the solutions must be identical throughout C.

The proof proceeds in two steps: first we establish

$$\alpha_{mn}(x) = \lambda(x)\beta_{mn}(x), \quad (14)$$

where $\lambda(x) > 0$ is a differentiable scalar function and

$$\alpha_{mn} = \tfrac{1}{2}(V_{m|n} + V_{n|m} - \dot{g}_{mn}), \quad (15)$$

$$\beta_{mn} = \tfrac{1}{2}(W_{m|n} + W_{n|m} - \dot{g}_{mn}). \quad (16)$$

Then we show that $\lambda(x) = 1$.

Define

$$M_k = uV_k + (1 - u)W_k, \quad (17)$$

$$\Gamma_{mn}(u) = u\alpha_{mn} + (1 - u)\beta_{mn} = \tfrac{1}{2}(M_{m|n} + M_{n|m} - \dot{g}_{mn}), \quad (18)$$

$$\Gamma^*(u) = (\Gamma^s_s)^2 - \Gamma_{rs}\Gamma^{rs}, \quad (19)$$

where u is a real parameter, $0 \leq u \leq 1$. $\Gamma^*(u)$, regarded as function of u (for fixed x), is a quadratic polynomial in u. It can be shown that $\Gamma^*(u)$ has no roots in the interval $0 \leq u \leq 1$.[6] Since $\Gamma^*(0) = \beta^*$ and $\Gamma^*(1) = \alpha^*$ are positive, the function $[\Gamma^*(u)]^{\frac{1}{2}}$ is real and all its derivatives with respect to u exist and are real in the interval $0 \leq u \leq 1$.

Next define

$$J(u) = \int_C \{-[(2\epsilon - R)\Gamma^*(u)]^{\frac{1}{2}} + S^k M_k\} g^{\frac{1}{2}}\, d^3x. \quad (20)$$

We will show that the derivatives of $J(u)$ with respect to u have the following properties:

(i) $J'(0) = J'(1) = 0$;

(ii) $J''(u) \geq 0$ in the interval $0 \leq u \leq 1$;

(iii) $J''(u) > 0$ in the interval $0 \leq u \leq 1$ unless Eq. (14) is true.

[5] J. A. Wheeler, in *Relativistic Fluid Mechanics and Magnetohydrodynamics*, R. Wasserman and C. P. Wells, Eds. (Academic Press Inc., New York, 1963), p. 1.

[6] The proof is given in Appendix A.

These contradictory properties of the function $J'(u)$ establish that Eq. (14) must hold.

The demonstration of (i) is trivial since Eq. (11) is the Euler–Lagrange equation for the integral J and, therefore, $J(u)$ at $u = 0$ and $u = 1$ is stationary for arbitrary variations of the solutions W_k and V_k, respectively. Items (ii) and (iii) are obtained by straightforward calculation of $J''(u)$.[7]

To show that $\lambda(x) = 1$ we begin by observing that this is certainly true on $d(C)$ because of the Dirichlet boundary condition. Suppose now that there exists some region C' such that $\lambda \neq 1$ in C' and $\lambda = 1$ on $d(C')$. It follows from Eq. (14) that, wherever $\lambda \neq 1$,

$$W_{k|m} + W_{m|k} - \dot{g}_{mk} = (W_{k|m} + W_{m|k} - V_{k|m} - V_{m|k})/(1 - \lambda), \quad (21)$$

so that $W_k - V_k$ must be a solution of the "homogeneous" initial-value equation [i.e., Eq. (11) with $\dot{g}_{mn} = 0$]. Designate by $C - C'$ the region where $\lambda(x) = 1$; then, in $C - C'$, $W_k - V_k$ must be a Killing vector. This makes the existence of any region C' impossible because the homogeneous equation cannot have any solution L_k which reduces to a Killing vector ξ_k on $d(C')$ and also satisfies the restriction (9). Since Killing vectors have zero divergence and since ξ_k vanishes on $d(C)$,

$$\begin{aligned} 0 &= \int_{C-C'} \xi^{|s}_s \, d^3x = \int_{d(C)} \xi_s n^s \, d^2x - \int_{d(C')} \xi_s n^s \, d^2x \\ &= -\int_{d(C')} \xi_s n^s \, d^2x \\ &= -\int_{d(C')} L_s n^s \, d^2x = -\int_{C'} L^{|s}_s \, d^3x. \end{aligned} \quad (22)$$

Therefore, $L^{|s}_s$ must vanish somewhere in C' and Eq. (9) is not satisfied.

We conclude finally that $\lambda(x) = 1$ everywhere so that W_k and V_k differ at most by a Killing vector. If $d(C) \neq 0$, this Killing vector would have to vanish on the boundary and must therefore be zero everywhere.

In absence of the condition $2\epsilon - R > 0$, it is possible to construct counterexamples to uniqueness. Take the closed space consisting of the cube $|x^1| \leq a$, $|x^2| \leq a$, $|x^3| \leq a$ with periodic boundary conditions imposed on all functions in this space (3-torus). Suppose $g_{mn} = \delta^n_m$, but $\dot{g}_{mn} \neq 0$. Suppose further that we have found some solution N_k of Eq. (3) such that γ_{mn} is constant and $\gamma_{22} + \gamma_{33} = 0$. These conditions are compatible. It is easy to check that $N_k + \delta^1_k f(x^1)$ is also a solution of Eq. (3) whenever $f(x^1)$ is a twice differentiable, periodic function of x^1. Our example is unphysical because ϵ must be negative, but similar constructions are possible in spaces of positive 3-curvature with positive ϵ. This lack of uniqueness is not surprising because examination of the characteristic determinant of the system of equations (3) reveals that the determinant has no definite sign if $2\epsilon - R < 0$. The equations are then not elliptic and Dirichlet boundary conditions are not natural. Unfortunately this seems to be the only type of boundary condition which is physically meaningful.

III. EXISTENCE[8]

No general existence theorem for Eqs. (7) is known. Solutions do not exist under all possible assignments of the functions g_{ij}, $\dot{g}_{ij}$, ϵ, and S_i. Here are three sets of conditions on these functions, each of which guarantees that no solution exists.

(i) If the geometry is closed and admits a Killing vector ξ_i and

$$\int S_i \xi^i g^{\frac{1}{2}} \, d^3x \neq 0, \quad (23)$$

no solution to any of the associated problems of initial conditions (g_{ij} and ϵ arbitrary) can be found. If a solution did exist, we would have $S^i = \pi^{ij}_{|j}$ for some tensor π^{ij}. Thus

$$\begin{aligned} \int S^i \xi_i g^{\frac{1}{2}} \, d^3x &= \int \pi^{ij}_{|j} \xi_i g^{\frac{1}{2}} \, d^3x \\ &= -\int \pi^{ij} \xi_{(i|j)} g^{\frac{1}{2}} \, d^3x = 0, \end{aligned} \quad (24)$$

which would contradict Eq. (23).

(ii) If the geometry is closed and $\dot{g}_{ij} = L_{i|j} + L_{j|i}$ for some vector L_k, and if $2\epsilon - R > 0$, no solution to any associated initial-value problem (S_i arbitrary) can exist. This is proved as follows: Consider a solution γ_{rs}; it must satisfy

$$(\gamma^s_s)^2 = 2\epsilon - R + \gamma^{rs}\gamma_{rs} > 0.$$

This is impossible, because

$$\gamma^s_s = \tfrac{1}{2}(N^s - L^s)_{|s}$$

and

$$\int (N^s - L^s)_{|s} g^{\frac{1}{2}} \, d^3x = 0.$$

(iii) For a closed geometry, if $g^{rs}\dot{g}_{rs} = \dot{g}/g = 0$ and $2\epsilon - R > 0$, no solution to any associated initial-value problem (S_i arbitrary) exists. The proof of this runs analogously to that of (ii), once it is noted that

$$\gamma^s_s = \tfrac{1}{2}(N^{|s}_s + N^s_{|s}).$$

IV. COMMENTS ON THE SOURCE TERMS[8]

A serious question arises when one considers conditions like $2\epsilon - R > 0$, which have appeared

[7] See Appendix B.

[8] E. P. Belasco, Senior thesis, Princeton University, 1967.

again and again. Although ϵ and S_i have, in the "lapse and shift" formulation of the problem of initial conditions, been assumed to be given functions, this representation is not perfect.

Given an Einstein 4-geometry, let us generate a problem of initial conditions. Consider the surface S given by $x^0 = 0$. We seek a three-dimensional metric, its derivative with respect to time, and energy terms. For coordinates on S we choose (in a representative patch) the coordinates x^i. The metric induced in S is thus g_{ij}. Examination of the Einstein equations reveals that $\epsilon = -k\bar{T}^0_0$ and $S_i = -k\bar{T}^0_i$, where $\bar{T}^\nu_\mu$ is the stress-energy tensor T^ν_μ expressed in a coordinate system that agrees with x^i on S but is *Gaussian normal* to S.

The unreasonableness of this situation is apparent. The metric g_{ij} was chosen as part of the freely specifiable data because of its measurability. Why should the measurement process produce data in two *different* coordinate systems? To be consistent, perhaps the terms S_k and ϵ should be written as follows:

$$\epsilon = -\bar{T}^0_0 = (T_{00} - 2T_{0i}N^i + T_{ij}N^iN^j)/N_0^2, \quad (25)$$

$$S_i = -\bar{T}^0_i = (T_{0i} - T_{si}N^s)/N_0. \quad (26)$$

The above forms, although true in all cases, are not always natural. Additional coupled equations may have to be introduced and other changes made in order to describe specific matter sources. Some particular cases of this are discussed in Ref. 1.

V. CONCLUDING REMARKS

We have seen that Eq. (7) may have no solutions, and even if it does the solution need not be unique. Hence, arbitrary specification of g_{ij} and $\dot{g}_{ij}$ may be inconsistent and even if it is consistent it need not be sufficient to determine the time development of the geometry uniquely. The quite plausible separation of the Cauchy data into two parts, one of which $(g_{ij}, \dot{g}_{ij})$ is freely specifiable and another (N_k, N_0) which is completely determined by the constraint equations, is therefore not always workable. The separation is only possible for a restricted class of functions g_{ij}, $\dot{g}_{ij}$, ϵ, and S_k for which Eq. (7) has one and only one solution N_k. The sufficient condition for uniqueness and the several necessary conditions for existence of Secs. II and III suggest that the requirements for membership in this privileged class are likely to be rather complex and exclusive.

ACKNOWLEDGMENTS

We express our appreciation to Professor J. A. Wheeler, who suggested this problem to us and provided constant advice and encouragement. We also thank Dr. R. P. Geroch for many discussions.

APPENDIX A

The function $\Gamma^*(u)$ defined by Eq. (19) can be written as a quadratic polynomial in u:

$$\Gamma^*(u) = u^2[\alpha^* + \beta^* - 2(\alpha\beta)^*] + 2u[(\alpha\beta)^* - \beta^*] + \beta^*, \quad (A1)$$

where $(\alpha\beta)^* = \alpha^*_r\beta^s_s - \alpha_{rs}\beta^{rs}$. One can show that roots of $\Gamma^*(u)$ occur in the interval $0 < u < 1$ if and only if

$$(\alpha\beta)^* < 0, \quad (A2)$$

$$[(\alpha\beta)^*]^2 - \alpha^*\beta^* \geq 0. \quad (A3)$$

[The case of equality in Eq. (A3) corresponds to double roots.] Suppose these conditions are satisfied and $\Gamma^*(u)$ has roots. If we replace β_{mn} by $-\beta_{mn}$, condition (A2) will not be satisfied any more since this operation changes the sign of $(\alpha\beta)^*$. Write the function defined by Eq. (18) as $\Gamma_{mn}(u; \alpha_{rs}, \beta_{rs})$ so as to display the dependence on α_{rs} and β_{rs}. The preceding argument shows that if $\Gamma^*(u; \alpha_{rs}, \beta_{rs})$ has roots in the interval $0 < u < 1$, then $\Gamma^*(u; \alpha_{rs}, -\beta_{rs})$ does not.

Consider $\alpha^n_m(x)$ and designate by $A_1(x)$, $A_2(x)$, $A_3(x)$ its eigenvalues in the local coordinate frame in which $g_{ks} = \delta^s_k$. We can introduce a three-dimensional space of eigenvalues in which the above eigenvalues are represented by an "eigenpoint" with Cartesian coordinates (A_1, A_2, A_3). The condition $\alpha^* > 0$ demands that the eigenpoints lie inside a double cone whose apex is at the origin, axis in the direction $(1, 1, 1)$, and which has the three coordinate axes lying on its surface.[4] The two halves of the double cone are distinguished by the value of $\alpha^k_k(x)$: it is positive in one cone and negative in the other. Now we observe that $\alpha^k_k(x)$ [and also $\beta^k_k(x)$] cannot change sign anywhere in C since that would violate $\alpha^* > 0$ (or $\beta^* > 0$). Further, by Eqs. (15) and (16) and the boundary conditions,

$$\int_C (\alpha^k_k - \beta^k_k)\, d^3x = \int_C (V^{|k}_k - W^{|k}_k)\, d^3x = \int_{a(C)} (V_k - W_k)n^k\, d^2x = 0.$$

Therefore, $\alpha^k_k = \beta^k_k$ somewhere within C and then α^k_k and β^k_k must have the same sign everywhere.

Next examine the behavior of $\Gamma^*(u; \alpha_{rs}, \beta_{rs})$ in the eigenvalue space (keeping x fixed). $\Gamma_{mn}(0; \alpha_{rs}, \beta_{rs})$ has the eigenpoint $\mathbf{B} = (B_1, B_2, B_3)$ and $\Gamma_{mn}(1; \alpha_{rs}, \beta_{rs})$ the point $\mathbf{A} = (A_1, A_2, A_3)$. As u varies from 0 to

1, the eigenpoint of $\Gamma_{mn}(u; \alpha_{rs}, \beta_{rs})$ traces out a path from **B** to **A**. Roots of $\Gamma^*(u; \alpha_{rs}, \beta_{rs})$ correspond to intersections of this path with the cone. The replacement of β_{mn} by $-\beta_{mn}$ changes **B** to $-$**B** and, since **A** and **B** lie in the same cone, **A** and $-$**B** will lie in opposite cones. But then the path from **A** to $-$**B** must necessarily intersect the cone, i.e., there must exist roots of $\Gamma^*(u; \alpha_{rs}, -\beta_{rs})$. However, we know that if the latter polynomial has roots in the interval $0 < u < 1$, then $\Gamma^*(u; \alpha_{rs}, \beta_{rs})$ does not.

This establishes the absence of roots of $\Gamma^*(u)$ needed for the proof of the uniqueness theorem.

APPENDIX B

Appendix A shows that $\Gamma^*(u; \alpha_{rs}, \beta_{rs})$ has no roots in the interval $0 < u < 1$, but that $\Gamma^*(u; \alpha_{rs}, -\beta_{rs})$ does. The existence of roots of $\Gamma^*(u; \alpha_{rs}, -\beta_{rs})$ implies that

$$(\alpha\beta)^* > 0, \tag{B1}$$

$$[(\alpha\beta)^*]^2 - \alpha^*\beta^* \geq 0. \tag{B2}$$

By Eq. (20),

$$J''(u) = -\int_C (2\epsilon - R)^{\frac{1}{2}} \left[\frac{d^2}{du^2}\Gamma^{*\frac{1}{2}}\right] g^{\frac{1}{2}}\, d^3x. \tag{B3}$$

If we use the expression (A1) for Γ^*, we obtain

$$\frac{d^2}{du^2}\Gamma^{*\frac{1}{2}} = -\Gamma^{*-\frac{3}{2}}\{[(\alpha\beta)^*]^2 - \alpha^*\beta^*\}. \tag{B4}$$

From (B4), (B2), and (B3) we see that $J''(u) \geq 0$.

We can further show that $J''(u) = 0$ if and only if

$$\alpha_{mn} = \lambda\beta_{mn}. \tag{B5}$$

It is obvious that (B5) is sufficient to make

$$[(\alpha\beta)^*]^2 - \alpha^*\beta^*,$$

and hence also $J''(u)$, vanish. To show that (B5) is necessary, we write

$$\alpha_{mn} = (\alpha_s^s/\beta_r^r)\beta_{mn} + q_{mn}. \tag{B6}$$

The tensor q_{mn} defined by this equation is traceless. Substituting the expression for α_{mn} given by (B6) and making use of $q_s^s = 0$, we find

$$\begin{aligned}[(\alpha\beta)^*]^2 - \alpha^*\beta^* &\equiv (\alpha_s^s\beta_r^r - \alpha_{rs}\beta^{rs})^2 \\ &\quad - [(\alpha_s^s)^2 - \alpha_{rs}\alpha^{rs}][(\beta_m^m)^2 - \beta_{mn}\beta^{mn}] \\ &= (\beta_{rs}q^{rs})^2 + \beta^* q_{rs}q^{rs}. \end{aligned} \tag{B7}$$

If this last quantity is to vanish, we must have $q_{rs} = 0$. Hence $J''(u) > 0$, unless Eq. (B5) holds.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Single-Particle Condensate and Pair-Correlation Theory of Inhomogeneous Boson Systems

DONALD H. KOBE*
Fysisk Laboratorium I, H. C. Ørsted Institutet, Copenhagen Ø, Denmark

(Received 10 December 1968)

A theory of inhomogeneous boson systems is developed which takes into account both the single-particle condensate and pair correlations. A set of three coupled nonlinear integrodifferential equations is derived for the condensate wavefunction and the two Bogoliubov quasiparticle wavefunctions. Half-integral h/m circulation can exist only when the single-particle condensate is completely depleted. General considerations of off-diagonal long-range order and circulation quantization also lead to this conclusion.

1. INTRODUCTION

One of the most remarkable developments in the microscopic theory of superfluidity was the prediction by Onsager[1] and Feynman[2] that vortices should be quantized in He II. The argument was based on the single-valuedness[3] of the condensate wavefunction, and the quanta of circulation was shown to be h/m. This prediction was first verified experimentally by Vinen,[4] using a vibrating wire. In a beautiful experiment on the motion of ions trapped in vortex rings, Rayfield and Reif[5] established conclusively that vortices are quantized with a circulation of h/m.

The presence of vortices with multiple quanta of circulation have also been observed. Whitmore and Zimmermann,[6] using the Vinen method, found

* Present address: Department of Physics, North Texas State University, Denton, Texas 76203.

[1] L. Onsager, Nuovo Cimento **6**, Suppl. 2, 249 (1949).

[2] R. P. Feynman, in *Progress in Low Temperature Physics*, C. J. Gorter, Ed. (North-Holland Publ. Co., Amsterdam, 1955), Vol. 1, Chap. II.

[3] For a discussion of the single-valuedness condition see, e.g., E. Merzbacher, Am. J. Phys. **30**, 237 (1962).

[4] W. F. Vinen, Proc. Roy. Soc. (London) **A260**, 218 (1961).

[5] G. W. Rayfield and F. Reif, Phys. Rev. **136**, A1194 (1964); Phys. Rev. Letters **11**, 305 (1963).

[6] S. C. Whitmore and W. Zimmermann, Jr., Phys. Rev. Letters **15**, 389 (1965); Phys. Rev. **166**, 181 (1968).

1, the eigenpoint of $\Gamma_{mn}(u; \alpha_{rs}, \beta_{rs})$ traces out a path from **B** to **A**. Roots of $\Gamma^*(u; \alpha_{rs}, \beta_{rs})$ correspond to intersections of this path with the cone. The replacement of β_{mn} by $-\beta_{mn}$ changes **B** to $-$**B** and, since **A** and **B** lie in the same cone, **A** and $-$**B** will lie in opposite cones. But then the path from **A** to $-$**B** must necessarily intersect the cone, i.e., there must exist roots of $\Gamma^*(u; \alpha_{rs}, -\beta_{rs})$. However, we know that if the latter polynomial has roots in the interval $0 < u < 1$, then $\Gamma^*(u; \alpha_{rs}, \beta_{rs})$ does not.

This establishes the absence of roots of $\Gamma^*(u)$ needed for the proof of the uniqueness theorem.

APPENDIX B

Appendix A shows that $\Gamma^*(u; \alpha_{rs}, \beta_{rs})$ has no roots in the interval $0 < u < 1$, but that $\Gamma^*(u; \alpha_{rs}, -\beta_{rs})$ does. The existence of roots of $\Gamma^*(u; \alpha_{rs}, -\beta_{rs})$ implies that

$$(\alpha\beta)^* > 0, \tag{B1}$$

$$[(\alpha\beta)^*]^2 - \alpha^*\beta^* \geq 0. \tag{B2}$$

By Eq. (20),

$$J''(u) = -\int_C (2\epsilon - R)^{\frac{1}{2}} \left[\frac{d^2}{du^2}\Gamma^{*\frac{1}{2}}\right] g^{\frac{1}{2}}\, d^3x. \tag{B3}$$

If we use the expression (A1) for Γ^*, we obtain

$$\frac{d^2}{du^2}\Gamma^{*\frac{1}{2}} = -\Gamma^{*-\frac{3}{2}}\{[(\alpha\beta)^*]^2 - \alpha^*\beta^*\}. \tag{B4}$$

From (B4), (B2), and (B3) we see that $J''(u) \geq 0$.

We can further show that $J''(u) = 0$ if and only if

$$\alpha_{mn} = \lambda\beta_{mn}. \tag{B5}$$

It is obvious that (B5) is sufficient to make

$$[(\alpha\beta)^*]^2 - \alpha^*\beta^*,$$

and hence also $J''(u)$, vanish. To show that (B5) is necessary, we write

$$\alpha_{mn} = (\alpha_s^s/\beta_r^r)\beta_{mn} + q_{mn}. \tag{B6}$$

The tensor q_{mn} defined by this equation is traceless. Substituting the expression for α_{mn} given by (B6) and making use of $q_s^s = 0$, we find

$$\begin{aligned}[(\alpha\beta)^*]^2 - \alpha^*\beta^* &\equiv (\alpha_s^s\beta_r^r - \alpha_{rs}\beta^{rs})^2 \\ &\quad - [(\alpha_s^s)^2 - \alpha_{rs}\alpha^{rs}][(\beta_m^m)^2 - \beta_{mn}\beta^{mn}] \\ &= (\beta_{rs}q^{rs})^2 + \beta^* q_{rs}q^{rs}. \end{aligned} \tag{B7}$$

If this last quantity is to vanish, we must have $q_{rs} = 0$. Hence $J''(u) > 0$, unless Eq. (B5) holds.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Single-Particle Condensate and Pair-Correlation Theory of Inhomogeneous Boson Systems

DONALD H. KOBE*
Fysisk Laboratorium I, H. C. Ørsted Institutet, Copenhagen Ø, Denmark

(Received 10 December 1968)

A theory of inhomogeneous boson systems is developed which takes into account both the single-particle condensate and pair correlations. A set of three coupled nonlinear integrodifferential equations is derived for the condensate wavefunction and the two Bogoliubov quasiparticle wavefunctions. Half-integral h/m circulation can exist only when the single-particle condensate is completely depleted. General considerations of off-diagonal long-range order and circulation quantization also lead to this conclusion.

1. INTRODUCTION

One of the most remarkable developments in the microscopic theory of superfluidity was the prediction by Onsager[1] and Feynman[2] that vortices should be quantized in He II. The argument was based on the single-valuedness[3] of the condensate wavefunction, and the quanta of circulation was shown to be h/m. This prediction was first verified experimentally by Vinen,[4] using a vibrating wire. In a beautiful experiment on the motion of ions trapped in vortex rings, Rayfield and Reif[5] established conclusively that vortices are quantized with a circulation of h/m.

The presence of vortices with multiple quanta of circulation have also been observed. Whitmore and Zimmermann,[6] using the Vinen method, found

* Present address: Department of Physics, North Texas State University, Denton, Texas 76203.

[1] L. Onsager, Nuovo Cimento **6**, Suppl. 2, 249 (1949).

[2] R. P. Feynman, in *Progress in Low Temperature Physics*, C. J. Gorter, Ed. (North-Holland Publ. Co., Amsterdam, 1955), Vol. 1, Chap. II.

[3] For a discussion of the single-valuedness condition see, e.g., E. Merzbacher, Am. J. Phys. **30**, 237 (1962).

[4] W. F. Vinen, Proc. Roy. Soc. (London) **A260**, 218 (1961).

[5] G. W. Rayfield and F. Reif, Phys. Rev. **136**, A1194 (1964); Phys. Rev. Letters **11**, 305 (1963).

[6] S. C. Whitmore and W. Zimmermann, Jr., Phys. Rev. Letters **15**, 389 (1965); Phys. Rev. **166**, 181 (1968).

vortices with one, two, and three quanta of circulation. However, in the last hours of some of their runs, they found rather stable circulation values at $h/2m$ and $\sim 3h/2m$, but do not attach any significance to them.[7] Steyert, Taylor, and Kitchens[8] observed the trajectory of frozen HD particles in He II, and found, in addition to integral multiples of h/m, significant peaks at $h/2m$ and $3h/2m$. Di Castro[9] has suggested that a possible explanation for these peaks is that a "pair condensate" coexists with the single-particle condensate. The integral multiples of h/m cannot be explained as being due to even multiples of $h/2m$ because much more energy is required to produce one vortex line with multiple quanta of circulation than several lines with single quanta having the same total circulation.[2] If vortices with circulation $h/2m$ exist, they are rare compared to those with circulation h/m, so the single-particle condensate must predominate. Further experiments should be performed to establish conclusively the existence or nonexistence of vortices with half-integral quanta of circulation.

The idea of a "pair condensate" is not new. A number of theories have taken correlations between particles of equal and opposite momentum into account. In the special case that the pairing correlations augment the single-particle condensate in the zero-momentum state, it can be said that a "pair condensate" is present. However, depending on the interaction, the effect of the pairing correlations may reduce the effective condensate. In inhomogeneous systems, the single-particle condensate is no longer only in the zero-momentum state, and correlations must be taken into account between pairs of particles in time-reversed states.

After Valatin[10] developed a theory of superconductivity independently of Bogoliubov,[11] he and Butler[12] reapplied the canonical transformation method to boson systems taking pair correlations into account. Their treatment of the zero-momentum state as consisting of pairs is not the same, however, as the original boson theory of Bogoliubov.[13] Since pairs of bosons with equal and opposite momentum satisfy approximate commutation relations in the same way that Cooper pairs in the theory of superconductivity do, it is perhaps to be expected that they would also condense if the interaction between the particles in the pair is sufficiently attractive.

Although the pair-correlation theory has been developed by a number of authors,[14] it has not been shown to lead to agreement with the observed energy spectrum of He II. Due to mathematical difficulties, the complicated integral equations have not yet been solved for a realistic potential. A calculation by the author[15] using a delta-function potential displaced from the origin was shown to lead to agreement with the experimental spectrum, but the depletion was over 100%. However, the pair-correlation theory goes beyond Bogoliubov's original theory[13] in a systematic way and thus merits further attention.

In this paper, a theory of inhomogeneous boson systems is developed which takes into account both the single-particle condensate and pair correlations. In this sense it is in the same spirit as an extended theory of Valatin,[16] but attention is not focused on the ground state. It is also similar to a recent theory of Cummings and Johnston,[17] but is more tractable and does not assume plane-wave expansions. The first to suggest such a theory combining both single-particle condensate and pair correlations was Gross.[18] Pitaevski[19] generalized the original canonical transformation of Bogoliubov[13] to the spatially inhomogeneous case. However, in his equations and those of subsequent authors[20] the contribution of pair correlations was overlooked, since the interaction Hamiltonian for the Bogoliubov quasiparticles was not put in normal order. When the normal ordering is performed, no contributions to the unperturbed ground-state energy are neglected. The normal ordering of the operators gives a pair potential for the quasiparticles which, in addition to a contribution from the single-particle condensate, has a term describing the pairing of particles in time-reversed states. This term is called

[7] W. Zimmermann, Jr., private communication, 1967.

[8] W. A. Steyert, R. D. Taylor, and T. A. Kitchens, Phys. Rev. Letters **15**, 546 (1965).

[9] C. Di Castro, Phys. Letters **24A**, 191 (1967). The term "pair condensate" is used to describe the effects of pair correlations, even though the pair correlations need not necessarily augment the single-particle condensate.

[10] J. G. Valatin, Nuovo Cimento **7**, 843 (1958).

[11] N. N. Bogoliubov, Zh. Eksp. Teor. Fiz. **34**, 58 (1958) [Sov. Phys.—JETP **7**, 41 (1958)]; Nuovo Cimento **7**, 794 (1958).

[12] J. G. Valatin and D. Butler, Nuovo Cimento **10**, 37 (1958). This paper has recently been criticized for excluding a true single-particle condensate. See Ref. 15.

[13] N. N. Bogoliubov, J. Phys. (USSR) **11**, 23 (1947).

[14] M. Girardeau and R. Arnowitt, Phys. Rev. **113**, 755 (1959); G. Wentzel, Phys. Rev. **120**, 1579 (1960); M. Luban, Phys. Rev. **128**, 965 (1962); D. N. Zubarev and Iu. Tserkovnikov, Dokl. Akad. Nauk SSSR **120**, 991 (1958) [Sov. Phys.—Dokl. **3**, 603 (1958)]; V. V. Tolmachev, Dokl. Akad. Nauk SSSR **134**, 1324 (1960) [Sov. Phys.—Dokl. **5**, 984 (1960)].

[15] D. H. Kobe, Ann. Phys. (N.Y.) **47**, 15 (1968); Nuovo Cimento **59B**, 187 (1969).

[16] J. G. Valatin, in *Lectures in Theoretical Physics*, W. E. Brittin and W. R. Chappell, Eds. (University of Colorado Press, Boulder, Colo., 1964), Vol. VI, pp. 245–372. On p. 362, when the equations become somewhat similar to ours, he arbitrarily sets $\phi(x) = 0$ for simplicity, which is not justified.

[17] F. W. Cummings and J. R. Johnston, Phys. Rev. **151**, 105 (1966).

[18] E. P. Gross, Ann. Phys. (N.Y.) **9**, 292 (1960).

[19] L. P. Pitaevski, Zh. Eksp. Teor. Fiz. **40**, 646 (1961) [Sov. Phys.—JETP **13**, 451 (1961)].

[20] See, e.g., D. Pines and P. Nozières, *Theory of Quantum Liquids* (W. A. Benjamin, Inc., New York, to be published), Vol. II, Chap. 6.

the pair-condensate wavefunction, although it need not necessarily augment the condensate. If both a pair condensate and single-particle condensate coexist, the pair condensate is locked in phase with the single-particle condensate, which gives circulation quantized only in units of h/m. However, if the single-particle condensate is completely depleted leaving only the pair condensate, circulation in units of $h/2m$ is possible. The theory here reduces to previous theories[20] if the pair condensate is neglected and a delta-function potential is used.

In the next section, a space- and time-dependent canonical transformation is made to separate the effect of the single-particle condensate. Then another space- and time-dependent canonical transformation is made to Bogoliubov quasiparticles or bogolons. The Hamiltonian is put in normal order in terms of the bogolon operators. In Sec. 3 the equations of motion for the coefficients in the transformation are obtained. The single-particle condensate wavefunction is coupled to the bogolon wavefunctions, and vice versa. The concept of off-diagonal long-range order (ODLRO) is discussed in Sec. 4. Intrinsic ODLRO in the two-particle density matrix when ODLRO is present in the one-particle density matrix is defined and shown to be present in this theory. On the basis of general arguments in Sec. 5, it is shown that the pair condensate is locked in phase with the single-particle condensate. The only way to have half-integral h/m circulation is for the single-particle condensate to be completely depleted. In Sec. 6, these general considerations are shown to be valid on the basis of the wave equations for a vortex line in the single-particle condensate in the presence of the pair condensate. The wave equations are shown in Sec. 7 to imply the possibility of half-integral h/m circulation when the single-particle condensate is completely depleted.

2. THE BOGOLON HAMILTONIAN

The Hamiltonian for a system of bosons interacting with a two-body potential V is[21]

$$H = \int d\mathbf{x}\psi^\dagger(x)T(\mathbf{x})\psi(x) + \tfrac{1}{2}\iint d\mathbf{x}\, d\mathbf{y}\psi^\dagger(x)\psi^\dagger(y)V(\mathbf{x},\mathbf{y})\psi(y)\psi(x). \quad (2.1)$$

The variable x represents the spatial coordinates $\mathbf{x}$ and the time t and similarly for y. The operator T is the sum of the kinetic energy and the external potential V_{ext}, if it is present, minus the chemical potential:

$$T(\mathbf{x}) = -\frac{\hbar^2}{2m}\nabla^2 + V_{\text{ext}}(\mathbf{x}) - \mu. \quad (2.2)$$

The field operators in the Heisenberg picture are

$$\psi(x) = e^{iHt}\psi(\mathbf{x})e^{-iHt}. \quad (2.3)$$

The time-independent field creation operator $\psi^\dagger(\mathbf{x})$ and annihilation operator $\psi(\mathbf{x})$ satisfy the usual boson commutation relations

$$\begin{aligned}[\psi(\mathbf{x}), \psi^\dagger(\mathbf{y})] &= \delta(\mathbf{x}-\mathbf{y}),\\ [\psi(\mathbf{x}), \psi(\mathbf{y})] &= 0,\\ [\psi^\dagger(\mathbf{x}), \psi^\dagger(\mathbf{y})] &= 0,\end{aligned} \quad (2.4)$$

which is equivalent to saying that the wavefunction for bosons must be symmetric.

The density of particles at a point in a boson system can be high, since the wavefunction does not vanish if two particles occupy the same point, as is the case for fermions. The expectation value of the density operator in the true ground state,

$$\rho(x) = \langle\psi^\dagger(x)\psi(x)\rangle, \quad (2.5)$$

can therefore be large at some points. It is now possible to make a canonical transformation to a type of quasiparticle which has, on the average, a very low density at each point. The field operator for the new quasiparticles is defined as

$$\chi(x) = \psi(x) - \phi(x), \quad (2.6)$$

and was first introduced by Pitaevski[19] generalizing the idea of Bogoliubov[13] to the spatially inhomogeneous case. The function $\phi(x)$, called the condensate wavefunction, describes the condensation of single particles. It is determined in the next section from the equation of motion. Since the field operators $\chi^\dagger(x)$ and $\chi(x)$ satisfy the commutation relations of Eq. (2.4), the quasiparticles are bosons and the transformation is canonical. The quasiparticle operator in Eq. (2.6) describes the deviation of the operator from the function $\phi(x)$ [usually taken to be its average value $\langle\psi(x)\rangle$[22]] and can hence be called a *devion* operator.

The density in Eq. (2.5) can then be expressed in terms of the devion operators as

$$\rho(x) = |\phi(x)|^2 + 2\,\text{Re}\,\langle\chi^\dagger(x)\rangle\phi(x) + \langle\chi^\dagger(x)\chi(x)\rangle. \quad (2.7)$$

[21] For a discussion of second quantization see, e.g., D. H. Kobe, Am. J. Phys. 34, 1150 (1960).

[22] P. Nozières, in *Quantum Fluids*, D. F. Brewer, Ed. (North-Holland Publ. Co., Amsterdam, 1966), p. 9. The expectation value can be defined either between states of fixed particle number N and $N-1$, or in a ground state in which the number of particles is no longer a good quantum number because a source term has been added to the Hamiltonian.

The function $|\phi(x)|^2$ can be chosen to be a good approximation to $\rho(x)$, so the other terms will be small.

The devions themselves can be expressed in terms of Bogoliubov quasiparticles, or bogolons, by introducing the canonical transformation[19,23]

$$\chi(x) = \sum_n [u_n(x)\gamma_n + v_n^*(x)\gamma_n^\dagger]. \tag{2.8}$$

The bogolons are bosons and so the creation and annihilation operators satisfy the usual commutation relations:

$$\begin{aligned} &[\gamma_n, \gamma_m^\dagger] = \delta_{nm}, \\ &[\gamma_m, \gamma_n] = 0, \\ &[\gamma_n^\dagger, \gamma_m^\dagger] = 0. \end{aligned} \tag{2.9}$$

Since the devions are also bosons and satisfy the commutation relations in Eq. (2.4), the completeness relations

$$\sum_n [u_n(x)u_n^*(y) - v_n^*(x)v_n(y)] = \delta(\mathbf{x} - \mathbf{y}) \tag{2.10}$$

and

$$\sum_n [u_n(x)v_n^*(y) - v_n^*(x)u_n(y)] = 0 \tag{2.11}$$

must be satisfied.

The bogolon Hamiltonian can be obtained by first using the transformation to devions on the Hamiltonian in Eq. (2.1), and then using the transformation to bogolons in Eq. (2.8). The creation and annihilation operators for the bogolons must be put in normal order, so that as much of the interaction as possible will be thrown into the ground-state energy and the bogolon kinetic energy.[24] In this way the expectation value of the bogolon interaction terms with respect to the bogolon vacuum state can be made zero. The bogolon interaction terms do not contribute as much in a subsequent perturbation expansion for the ground-state energy or the bogolon self-energy. In this paper the full Hamiltonian is used, whereas previous treatments[20] have neglected the devion interaction terms before making the transformation to bogolons.

The Hamiltonian in terms of the bogolon operators can be written as

$$H = H_{00} + H_{01} + H_{10} + H_{11} + H_{20} + H_{02} + H_{\text{int}}. \tag{2.12}$$

The bogolon interaction term H_{int} involves at least three or four bogolon operators

$$H_{\text{int}} = \sum_{j,k} H_{jk}, \tag{2.13}$$

where $j + k = 3$ or 4. In Eqs. (2.12) and (2.13) the first subscript in H_{jk} denotes the number of creation operators and the second denotes the number of annihilation operators. The term can thus be written as

$$H_{jk} = \sum_{1,2,\cdots,j+k} h_{jk}(1, 2, \cdots, j+k) \times \gamma_1^\dagger \cdots \gamma_j^\dagger \gamma_{j+1} \cdots \gamma_{j+k}, \tag{2.14}$$

for $j, k = 0, 1, 2, 3, 4$ and the sum is over all states. Since the Hamiltonian is hermitian,

$$H_{jk} = H_{kj}^\dagger \tag{2.15}$$

for all allowed j and k. The coefficients in Eq. (2.14) thus have the property

$$h_{jk}(1, 2, \cdots, j+k) = h_{kj}^*(j+k, \cdots, 2, 1), \tag{2.16}$$

so that not all of them need be given explicitly.

The unperturbed bogolon ground-state energy H_{00} is obtained if the expectation value of Eq. (2.12) is taken with respect to the bogolon vacuum. It is given in the Appendix along with a discussion of the variational principle and the compensation of the lowest-order dangerous diagrams.[25] In the Hamiltonian of Eq. (2.12) there are now two types of "dangerous diagrams." The usual ones in superconductivity theory describe the creation or annihilation of two bogolons from the vacuum, and are given by H_{20} and H_{02}. However, it is now possible to have "dangerous diagrams" that lead from the vacuum to a single bogolon state. The usual method of "compensating the dangerous diagrams" leads to the time-independent equations.

The coefficient in the dangerous term H_{01} is given by

$$h_{01}(n) = \int d\mathbf{x}[W^*(x)u_n(x) + W(x)v_n(x)], \tag{2.17}$$

where the function W is defined as

$$\begin{aligned} W(x) = T(\mathbf{x})\phi(x) &+ \int d\mathbf{y} V(\mathbf{x}, \mathbf{y})\phi^*(y)\phi(y)\phi(x) \\ &+ \int d\mathbf{y} V(\mathbf{x}, \mathbf{y})[\langle\chi^\dagger(y)\chi(x)\rangle\phi(x) \\ &+ \langle\chi^\dagger(y)\chi(x)\rangle\phi(y) + \langle\chi(y)\chi(x)\rangle\phi^*(y)]. \end{aligned} \tag{2.18}$$

The devion density matrix in the bogolon vacuum state can be obtained from Eq. (2.8) and is

$$\langle\chi^\dagger(x)\chi(y)\rangle = \sum_n v_n(x)v_n^*(y). \tag{2.19}$$

The devion pair amplitude, which plays the role of an order parameter for the "pair condensate," is

$$\langle\chi(x)\chi(y)\rangle = \sum_n u_n(x)v_n^*(y). \tag{2.20}$$

[23] W. F. Vinen, in *Quantum Fluids*, D. F. Brewer, Ed. (North-Holland Publ. Co., Amsterdam, 1966), p. 95. However, he expands ψ in terms of bogolons, instead of χ, and neglects the terms coming from the normal ordering of the interaction terms.

[24] This procedure is absolutely essential in the theory of superconductivity. See Refs. 10 and 11.

[25] For a discussion of the principle of compensation of dangerous diagrams see D. H. Kobe, Phys. Rev. **140**, A825 (1965); Ann. Phys. (N.Y.) **40**, 395 (1966); J. Math. Phys. **8**, 1200 (1967); J. Math. Phys. **9**, 1779, 1795 (1968).

It is analogous to the anomalous propagator in the Green's function theory of superconductivity[26] and superfluidity.[27] The terms involving the devion density and pair amplitude in Eq. (2.18) would not occur if the devion interaction terms had not been included when the bogolon transformation was made.

The bogolon kinetic energy term H_{11} in Eq. (2.12) has the coefficient

$$h_{11}(n, m) = \iint d\mathbf{x}\, d\mathbf{y}\{u_n^*(x)U(x, y)u_m(y) + v_m(x)U(x, y)v_n^*(y) + \tfrac{1}{2}\Delta(x, y)[v_n^*(x)u_m(y) + u_m(x)v_n^*(y)] + \tfrac{1}{2}\Delta^*(x, y)[u_n^*(x)v_m(y) + v_m(x)u_n^*(y)]\}, \quad (2.21)$$

which is not in diagonal form. The coefficient of the "dangerous" term H_{02} in Eq. (2.12) is given by

$$h_{02}(n, m) = \tfrac{1}{2}\iint d\mathbf{x}\, d\mathbf{y}\{v_n(x)U(x, y)u_m(y) + v_m(x)U(x, y)u_n(y) + \Delta(x, y)u_n(x)u_m(y) + \Delta^*(x, y)v_n(x)v_m(y)\}. \quad (2.22)$$

In both Eqs. (2.21) and (2.22) the operator U and the function Δ appear. The operator U is essentially the Hartree–Fock Hamiltonian in the sense that it is the sum of the kinetic energy and the self-consistent potential due to all the other particles with exchange

$$U(x, y) = \delta(\mathbf{x} - \mathbf{y})\Big\{T(\mathbf{x}) + \int d\mathbf{z}V(\mathbf{x}, \mathbf{z})[\phi^*(z)\phi(z) + \langle\chi^\dagger(z)\chi(z)\rangle]\Big\} + V(\mathbf{x}, \mathbf{y})[\phi^*(y)\phi(x) + \langle\chi^\dagger(y)\chi(x)\rangle]. \quad (2.23)$$

The function Δ is the pair potential

$$\Delta(x, y) = V(\mathbf{x}, \mathbf{y})[\phi^*(x)\phi^*(y) + \langle\chi^\dagger(x)\chi^\dagger(y)\rangle] \quad (2.24)$$

and is related to the probability of a pair of particles condensing into either the single-particle condensate or the "pair condensate." It is the analog of the pair potential in the theory of superconductivity.[28]

In the next section the Hamiltonian in Eq. (2.12) is used to obtain the equation of motion for the functions in the canonical transformations.

3. EQUATIONS OF MOTION

The Heisenberg equation of motion for the field operator $\psi(x)$ is determined by differentiating Eq. (2.3) with respect to time, which gives

$$i\dot{\psi}(x) = [\psi(x), H]. \quad (3.1)$$

The Hamiltonian of Eq. (2.12) is substituted directly into Eq. (3.1) in order to obtain the equations of motion for the functions in the canonical transformations of Eqs. (2.6) and (2.8). The bogolon interaction terms involving three or more operators are neglected. The canonical transformations to devions and bogolons in Eqs. (2.6) and (2.8) can be combined to give

$$\psi(x) = \phi(x) + \sum_n [u_n(x)\gamma_n + v_n^*(x)\gamma_n^\dagger]. \quad (3.2)$$

Then Eq. (3.2) can be substituted into Eq. (3.1) and the coefficients of the same operator on the left and the right sides can be equated.

The coefficients of the unit operator give the equation for the condensate wavefunction:

$$i\dot{\phi}(x) = \sum_n [h_{10}(n)u_n(x) - h_{01}(n)v_n^*(x)]. \quad (3.3)$$

The coefficients of the bogolon annihilation operator γ_n give the equation for the amplitude $u_n(x)$:

$$i\dot{u}_n(x) = \sum_m [h_{11}(m, n)u_m(x) - 2h_{02}(m, n)v_m^*(x)]. \quad (3.4)$$

Likewise, the coefficients of the bogolon creation operator $\gamma_n^\dagger$ give the equation

$$i\dot{v}_n(x) = \sum_m [h_{11}(m, n)v_m(x) - 2h_{02}(m, n)u_m^*(x)]. \quad (3.5)$$

If the coefficients h_{01}, h_{11}, and h_{20} in Eqs. (2.17), (2.21), and (2.22) are substituted into the equations of motion in Eqs. (3.3)–(3.5), and use is made of the completeness relations in Eqs. (2.10) and (2.11), the equations for ϕ, u_n, and v_n are obtained. The equation for the condensate wavefunction ϕ is

$$i\dot{\phi}(x) = T(\mathbf{x})\phi(x) + \int d\mathbf{y}V(\mathbf{x}, \mathbf{y})\phi^*(y)\phi(y)\phi(x) + \int d\mathbf{y}V(\mathbf{x}, \mathbf{y})[\langle\chi^\dagger(y)\chi(y)\rangle\phi(x) + \langle\chi^\dagger(y)\chi(x)\rangle\phi(y) + \langle\chi(x)\chi(y)\rangle\phi^*(y)]. \quad (3.6)$$

It is a generalization of the usual Gross[29]–Pitaevski[19] equation for the condensate wavefunction, since it involves coupling to the bogolon functions u_n and v_n through the devion density matrix $\langle\chi^\dagger(x)\chi(y)\rangle$ and devion pair amplitude $\langle\chi(x)\chi(y)\rangle$ defined in Eqs. (2.19) and (2.20), respectively. The equations of motion for the bogolon u_n and v_n are

$$i\dot{u}_n(x) = \int d\mathbf{y}U(x, y)u_n(y) + \int d\mathbf{y}\Delta^*(x, y)v_n(y) \quad (3.7)$$

[26] L. P. Gorkov, Zh. Eksp. Teor. Fiz. **34**, 735 (1958) [Sov. Phys.—JETP **7**, 289 (1958)].

[27] S. T. Beliaev, Zh. Eksp. Teor. Fiz. **34**, 417 (1958) [Sov. Phys.—JETP **7**, 289 (1958)].

[28] See, e.g., P. G. de Gennes, *Superconductivity of Metals and Alloys* (W. A. Benjamin, Inc., New York, 1966), pp. 137–145.

[29] E. P. Gross, Ann. Phys. (N.Y.) **4**, 57 (1958).

and

$$-i\dot{v}_n(x) = \int dy U^*(x, y)v_n(y) + \int dy \Delta(x, y)u_n(y). \tag{3.8}$$

The Hartree–Fock Hamiltonian U is defined in Eq. (2.23) and the pair potential is defined in Eq. (2.24). If the devion density matrix and pair amplitude were neglected in Eqs. (2.23) and (2.24) and a delta function potential were used, Eqs. (3.7) and (3.8) would reduce to previously given equations for u_n and v_n.[20] However, these terms are not obviously small, and should thus be included.

The equations derived for the condensate wavefunction and the bogolon wavefunctions u_n and v_n are analyzed in more detail in Secs. 6 and 7. The next two sections examine some general concepts in order to obtain a better understanding of the pair condensate.

4. OFF-DIAGONAL LONG-RANGE ORDER

The concept of off-diagonal long-range order (ODLRO) in the reduced density matrices has been developed by Yang[30] on the basis of ideas introduced by Penrose[31] and Penrose and Onsager.[32] This concept provides the most general criterion for the existence of superfluidity or superconductivity, since it avoids all mention of Hamiltonians, interactions, or state vectors. The existence of superfluidity in boson systems is due to ODLRO in the one-particle density matrix ρ_1, whereas superconductivity is due to ODLRO in the two-particle density matrix ρ_2.

The explanation given by Di Castro[9] for the occurrence of half-integral h/m circulation is that there could be an intrinsic ODLRO in the two-particle density matrix, as well as ODLRO in the one-particle density matrix. He suggested that this intrinsic ODLRO was due to a "pair condensate" of a type similar to that in superconductivity, and that it automatically implied the existence of half-integral h/m circulation. It is shown in this section that there is indeed intrinsic ODLRO in the two-particle density matrix, but subsequent sections show that half-integral h/m circulation is not automatically implied.

Yang[30] has defined ODLRO for ρ_1 and for ρ_2 in the absence of ODLRO in ρ_1. However, it is possible to extend the concept to include ODLRO in ρ_2 in the presence of ODLRO in ρ_1. The following four cases can be distinguished:

(a) No ODLRO in ρ_1 and no ODLRO in ρ_2;
(b) No ODLRO in ρ_1 and ODLRO in ρ_2;
(c) ODLRO in ρ_1 and no ODLRO in ρ_2;
(d) ODLRO in ρ_1 and ODLRO in ρ_2.

[30] C. N. Yang, Rev. Mod. Phys. **34**, 694 (1962).
[31] O. Penrose, Phil. Mag. **42**, 1373 (1951).
[32] O. Penrose and L. Onsager, Phys. Rev. **104**, 576 (1956).

Yang considered the first two cases and lumped the last two together under ODLRO in ρ_1. However, it is useful to distinguish between the ODLRO in ρ_2 as a result of ODLRO in ρ_1 and the *intrinsic* ODLRO in ρ_2. In the following we will say that ρ_2 has ODLRO only when it has intrinsic ODLRO.

The one-particle density matrix ρ_1 is defined as

$$\rho_1(\mathbf{x}, \mathbf{y}) = \langle \psi^\dagger(\mathbf{x})\psi(\mathbf{y})\rangle, \tag{4.1}$$

where the expectation value is taken in the true ground state. If ρ_1 does not have ODLRO, then

$$\rho_1(\mathbf{x}, \mathbf{y}) \to 0, \quad \text{as} \quad |\mathbf{x} - \mathbf{y}| \to \infty; \tag{4.2}$$

but if ρ_1 has ODLRO, then

$$\rho_1(\mathbf{x}, \mathbf{y}) \nrightarrow 0, \quad \text{as} \quad |\mathbf{x} - \mathbf{y}| \to \infty, \tag{4.3}$$

which is equivalent to saying that ρ_1 has an eigenvalue of order N.

The two-particle density matrix is defined as

$$\rho_2(\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}) = \langle \psi^\dagger(\mathbf{w})\psi^\dagger(\mathbf{x})\psi(\mathbf{y})\psi(\mathbf{z})\rangle. \tag{4.4}$$

In order to separate the effects of ρ_1 on ρ_2, an *intrinsic two-particle density matrix* is defined as

$$\Delta\rho_2(\mathbf{x}, \mathbf{y}) = \rho_2(\mathbf{x}, \mathbf{x}, \mathbf{y}, \mathbf{y}) - \rho_1(\mathbf{x}, \mathbf{y})^2. \tag{4.5}$$

The absence of intrinsic ODLRO (or just ODLRO) in ρ_2 is defined to be the absence of ODLRO in the intrinsic two-particle density matrix

$$\Delta\rho_2(\mathbf{x}, \mathbf{y}) \to 0, \quad \text{as} \quad |\mathbf{x} - \mathbf{y}| \to \infty, \tag{4.6}$$

regardless if ρ_1 had ODLRO or not. In the case where ρ_1 does not have ODLRO, Eq. (4.6) reduces to

$$\rho_2(\mathbf{x}, \mathbf{x}, \mathbf{y}, \mathbf{y}) \to 0, \quad \text{as} \quad |\mathbf{x} - \mathbf{y}| \to \infty, \tag{4.7}$$

which is just Yang's definition of the absence of ODLRO in ρ_2. In the case where ρ_1 has ODLRO, Eq. (4.6) reduces to

$$\rho_2(\mathbf{x}, \mathbf{x}, \mathbf{y}, \mathbf{y}) \to \lim \rho_1(\mathbf{x}, \mathbf{y})^2, \quad \text{as} \quad |\mathbf{x} - \mathbf{y}| \to \infty, \tag{4.8}$$

which is the nonintrinsic part of the ODLRO in ρ_2.

The presence of intrinsic ODLRO in ρ_2 is defined to be the same as ODLRO in the intrinsic two-particle density matrix

$$\Delta\rho_2(\mathbf{x}, \mathbf{y}) \nrightarrow 0, \quad \text{as} \quad |\mathbf{x} - \mathbf{y}| \to \infty. \tag{4.9}$$

If ρ_1 does not have ODLRO, then this definition reduces to

$$\rho_2(\mathbf{x}, \mathbf{x}, \mathbf{y}, \mathbf{y}) \nrightarrow 0, \quad \text{as} \quad |\mathbf{x} - \mathbf{y}| \to \infty, \tag{4.10}$$

which is just Yang's definition of ODLRO in ρ_2. In the case where ρ_1 has ODLRO, Eq. (4.9) can be

TABLE I. Presence or absence of off-diagonal long-range order (ODLRO) in the one[a]- and two[b]-particle density matrices.

ρ_1 \ ρ_2	No ODLRO	ODLRO
No ODLRO	$\rho_2 \to 0$ $\Delta\rho_2 \to 0$	$\rho_1 \to 0$ $\Delta\rho_2 \nrightarrow 0$
ODLRO	$\rho_1 \nrightarrow 0$ $\Delta\rho_2 \to 0$	$\rho_1 \nrightarrow 0$ $\Delta\rho_2 \nrightarrow 0$

[a] $\rho_1 = \langle\psi^\dagger(x)\psi(y)\rangle$.
[b] $\rho_2 = \langle\psi^\dagger(x)\psi^\dagger(x)\psi(y)\psi(y)\rangle$, $\Delta\rho_2 = \rho_2 - \rho_1^2$, all limits being as $|x - y| \to \infty$.

rewritten as

$$\rho_2(\mathbf{x}, \mathbf{x}, \mathbf{y}, \mathbf{y}) \nrightarrow \lim \rho_1(\mathbf{x}, \mathbf{y})^2, \quad \text{as} \quad |\mathbf{x} - \mathbf{y}| \to \infty, \tag{4.11}$$

which shows that there is some contribution to the ODLRO in ρ_2 other than that due to ρ_1. The four cases discussed here are summarized in Table I.

In order to illustrate these ideas, we can apply them to the density matrices calculated from the canonical transformation introduced in Sec. 2. If the devion operator is neglected in Eq. (2.6), the field operator becomes just the condensate wavefunction ϕ. Then the one- and two-particle density matrices defined in Eqs. (4.1) and (4.4) are

$$\rho_1(\mathbf{x}, \mathbf{y}) = \phi^*(x)\phi(y) \tag{4.12}$$

and

$$\rho_2(\mathbf{x}, \mathbf{x}, \mathbf{y}, \mathbf{y}) = \phi^*(x)\phi^*(x)\phi(y)\phi(y). \tag{4.13}$$

This approximation is equivalent to the Hartree approximation[33] for the ground-state wavefunction. In this approximation, ρ_1 has ODLRO according to Eq. (4.3). The two-particle density matrix has ODLRO according to Yang's definition in Eq. (4.10), but since the intrinsic two-particle density matrix from Eq. (4.5) is

$$\Delta\rho_2 = 0, \tag{4.14}$$

ρ_2 does not have any intrinsic ODLRO according to Eq. (4.6).

The situation is different if the full canonical transformation of Eq. (3.2) is used. The one-particle density matrix defined in Eq. (4.1) then becomes

$$\rho_1(\mathbf{x}, \mathbf{y}) = \phi^*(x)\phi(y) + \langle\chi^\dagger(x)\chi(y)\rangle, \tag{4.15}$$

where the devion density matrix is defined in Eq. (2.19). Equation (4.15) shows that ρ_1 has ODLRO according to Eq. (4.3). The intrinsic two-particle density matrix in Eq. (4.5) becomes

$$\begin{aligned}\Delta\rho_2(\mathbf{x}, \mathbf{y}) = {} & \phi^*(x)\phi^*(x)\langle\chi(y)\chi(y)\rangle \\ & + \langle\chi^\dagger(x)\chi^\dagger(x)\rangle\phi(y)\phi(y) \\ & + 2\phi^*(x)\phi(y)\langle\chi^\dagger(x)\chi(y)\rangle \\ & + \langle\chi^\dagger(x)\chi^\dagger(x)\rangle\langle\chi(y)\chi(y)\rangle \\ & + \langle\chi^\dagger(x)\chi(y)\rangle^2,\end{aligned} \tag{4.16}$$

which has ODLRO according to Eq. (4.9) if the devion pair amplitude defined in Eq. (2.20) is not zero. In the case where the condensate wavefunction ϕ is zero, there is no ODLRO in ρ_1, but there is still ODLRO in the two-particle density matrix due to the pair amplitude. In this case, which is analogous to superconductivity, there might still be a form of superfluidity due to the presence of the pair condensate described by the pair amplitude. In the next section the idea of circulation quantization is considered including both the single-particle and "pair condensate."

5. CIRCULATION QUANTIZATION

The quantization of circulation in He II was first suggested by Onsager[1] in a footnote, and later developed independently by Feynman.[2] The condition used was the single-valuedness[3] of the condensate wavefunction ϕ and the quantum of circulation is h/m. If there is also intrinsic ODLRO in the two-particle density matrix, there would be a pair condensate of the type important in superconductivity coexisting with the single-particle condensate. The pair condensate could also undergo motion and the circulation in it would be expected to be quantized.

The order parameter for the single-particle condensate is ϕ, which should be single-valued. The order parameter for the pair condensate is taken as

$$\langle\chi(x)\chi(x)\rangle = \Phi(x),$$

the devion pair amplitude, which should also be single-valued. The total effective condensate is taken to be the two-particle amplitude

$$\Psi(x) = \langle\psi(x)\psi(x)\rangle = \phi(x)\phi(x) + \Phi(x), \tag{5.1}$$

which is related to the pair potential in Eq. (2.24). The function Ψ describes both the single-particle and the pair contribution to the condensate. It can be treated *ad hoc* as a wavefunction for a pair of particles and a velocity of the total condensate is defined as the current divided by the density or

$$\mathbf{v} = \frac{\hbar}{2m} \frac{\operatorname{Im} \Psi^*(x)\nabla\Psi(x)}{\Psi^*(x)\Psi(x)}. \tag{5.2}$$

The factor $2m$ appears because the two-particle

[33] E. P. Gross, J. Math. Phys. 4, 195 (1963).

amplitude is considered. The single-particle condensate can be written as

$$\phi(x) = |\phi(x)|\, e^{iS(x)} \tag{5.3}$$

and the pair condensate as

$$\Phi(x) = |\Phi(x)|\, e^{iQ(x)}, \tag{5.4}$$

where S and Q are real functions. If Eqs. (5.3) and (5.4) are substituted into Eq. (5.1), the result is

$$\Psi(x) = e^{i2S(x)}[|\phi(x)|^2 + |\Phi(x)|\, e^{iQ(x)-2S(x)}]. \tag{5.5}$$

When Eq. (5.5) is substituted into Eq. (5.2), the result is, in general, complicated and the velocity is not just the gradient of a function.

In the special cases where the pair condensate is zero

$$|\Phi(x)| = 0 \tag{5.6}$$

or the phases are related by

$$Q(x) = 2S(x), \tag{5.7}$$

Eq. (5.2) becomes

$$\mathbf{v} = (\hbar/2m)2\nabla S. \tag{5.8}$$

The circulation κ can be calculated, assuming that the wavefunction is single-valued:

$$\kappa = \oint \mathbf{v} \cdot d\mathbf{l} = nh/m, \tag{5.9}$$

where n is an integer. This equation is just the usual circulation-quantization condition. Thus, in order to obtain circulation quantization, the "pair condensate" must be zero or it must be locked in phase with the single-particle condensate. Therefore, when a single-particle condensate exists, it is not possible to have half-integral h/m circulation.

However, in the special case where the single-particle condensate vanishes, leaving only the pair condensate, the velocity becomes

$$\mathbf{v} = (\hbar/2m)\nabla Q. \tag{5.10}$$

Then Eq. (5.9) gives circulation quantized in units of $h/2m$. This situation is analogous to superconductivity, since only the pair condensate contributes. Since the single-particle condensate wavefunction ϕ is usually considered to be the order parameter for the superfluid phase, its vanishing would mean that the system was normal. However, a new phase with a nonvanishing order parameter $\langle\chi\chi\rangle$ could conceivably exist if the interparticle potential were sufficiently attractive. In fact, some model calculations do show such a phase.[34]

[34] A. Coniglio, F. Mancini, and M. Maturi, "On the Coexistence of Single- and Two-Particle Condensation in an Interacting Boson Gas," preprint, Istituto di Fisica Teorica dell' Università, Napoli, Italy, 1968.

The experimental situation regarding the half-integral h/m circulations is not clear. The apparent half-integral h/m circulations observed by the Vinen method[6] could easily be explained by assuming a fractional vortex line attached to the wire. If the complete data of Whitmore and Zimmermann[35] is examined, there is little evidence for half-integral h/m. In the last hour of their runs they found stable circulations, some of which appear to be at half-integral h/m and which Di Castro[9] has considered significant. However, these stable circulations occurred when the surface of the helium bath was within one centimeter of the wire, and could be shifted to a different value, still not an integral multiple of h/m, by heating the wire. At integral values of h/m this could not be done, so the significance of these peaks is questionable. The experiment of Steyert *et al.*[8] definitely shows peaks at $h/2m$ and $3h/2m$, which the authors themselves considered significant. The experiment has been repeated to obtain better statistics, and the half-integral h/m peaks continue to persist.[36] The reason that the macroscopic particles observed should move with the superfluid velocity is not clear, however. There is the possibility of detecting vortices in a pair condensate by using the method of Rayfield and Reif,[5] but this method requires that the single-particle condensate be completely depleted.

6. VORTEX LINE IN THE SINGLE-PARTICLE CONDENSATE

The previous general arguments can be verified by a consideration of forms of solutions admitted by Eqs. (3.6)–(3.8). The case considered in this section is that of a combined single-particle and pair condensate in which there is a single vortex line.[37] It is shown that the pair condensate becomes locked into the single-particle condensate and rotates with it. This behavior is perhaps to be expected, since it would otherwise be difficult to have virtual transitions from the single-particle condensate to the pair condensate, and vice versa. This behavior follows from the fact that Eqs. (3.6)–(3.8) are coupled.

If there is a single vortex line present along the axis of a cylinder, the single-particle condensate wavefunction has the time-independent form[19,38]

$$\phi(\mathbf{x}) = \phi(r, \theta, z) = N^{\frac{1}{2}} f(r) e^{i\theta}, \tag{6.1}$$

[35] S. C. Whitmore, Ph.D. thesis, University of Minnesota, 1966.

[36] W. A. Steyert, private communication, 1967. He and R. D. Taylor repeated the experiment in Footnote 8 with frozen particles of Ne, N, air, and D_2, using a laser for improved resolution in time and space and a heat current to produce the vortices. The new results also show half-integral h/m circulations (unpublished).

[37] A. L. Fetter, Phys. Rev. **138A**, 709 (1965). His derivation is based on Green's functions.

[38] A. L. Fetter, Phys. Rev. **140A**, 452 (1965).

in cylindrical coordinates. The number of particles is N and f is a real function.

In order for Eqs. (3.6)–(3.8) to be separable in the cylindrical coordinates, the bogolon wavefunctions must be of the form

$$u(x) = u_{nkl}(r)e^{ikz}e^{i(l+1)\theta}e^{-iEt} \tag{6.2}$$

and

$$v(x) = v_{nkl}(r)e^{ikz}e^{i(l-1)\theta}e^{-iEt}. \tag{6.3}$$

The wavenumber in the z direction is k, and l is an angular-momentum quantum number. If Eqs. (6.2) and (6.3) are used in Eq. (2.19) for the devion density matrix, the result is

$$\langle\chi^\dagger(x)\chi(x)\rangle = Nh(r) = \sum_{nkl}|v_{nkl}(r)|^2, \tag{6.4}$$

which shows that the density in the pair condensate is a radial function only. The devion pair amplitude which describes the pair condensate can be obtained by substituting Eqs. (6.2) and (6.3) into Eq. (2.20), which gives

$$\langle\chi(x)\chi(x)\rangle = Ng(r)e^{i2\theta} = \sum_{nkl}u_{nkl}(r)v^*_{nkl}(r)e^{i2\theta}. \tag{6.5}$$

Thus the pair condensate has the same angular behavior as the contribution of the single-particle condensate to the pair potential in Eq. (2.24), so the pair potential can be factored into a radial part and an angular part.

If Eqs. (6.1), (6.4), and (6.5) are substituted into Eq. (3.6) for the condensate wavefunction, the angular dependence cancels out and the equation for the radial part is

$$f'' + \xi^{-1}f' - [\xi^{-2} - \alpha + f^2 + 2h + g]f = 0, \tag{6.6}$$

when the delta function potential

$$V(\mathbf{x}, \mathbf{y}) = V\delta(\mathbf{x} - \mathbf{y}) \tag{6.7}$$

is used. Equation (6.6) is written in terms of the dimensionless distance $\xi = r/a$, where a is the de Broglie wavelength defined as

$$a = (2mNV)^{-\frac{1}{2}}, \tag{6.8}$$

and the reduced chemical potential is

$$\alpha = \mu/NV. \tag{6.9}$$

Equation (6.6) reduces to the usual Gross–Pitaevski equation in cylindrical coordinates when the terms g and h are neglected.[39]

The equations for the bogolon wavefunctions in Eqs. (3.7) and (3.8) under the same assumption of delta-function potential can be obtained by substituting Eqs. (6.2) and (6.3) into them, which gives

$$-\epsilon u = u'' + \xi^{-1}u' + [-\xi^{-2}(l+1)^2 - \kappa^2 + \alpha - 2f^2 - 2h]u - [f^2 + g]v \tag{6.10}$$

and

$$\epsilon v = v'' + \xi^{-1}v' + [-\xi^{-2}(l-1)^2 - \kappa^2 + \alpha - 2f^2 - 2h]v - [f^2 + g]u. \tag{6.11}$$

The dimensionless energy parameter ϵ is

$$\epsilon = E/NV, \tag{6.12}$$

where E is defined in Eqs. (6.2) and (6.3), and the dimensionless wave vector in the z direction is

$$\kappa = ka. \tag{6.13}$$

The three equations for f, u, and v form a set of coupled nonlinear differential equations which is very difficult to solve exactly. These equations have been investigated in the approximation that the nonlinear terms g and h are neglected by Pitaevski[19] and Fetter.[37] Both bound-state and scattering-state solutions exist, the lowest lying bound states occurring for $l = \pm 1$. The scattering of quasiparticles has also been investigated by Fetter,[38] who determined the phase shifts and cross sections in the long wavelength limit. The nonlinear terms are not obviously negligible, especially if the depletion is large, and their effect should be investigated.

7. COMPLETELY DEPLETED SINGLE-PARTICLE CONDENSATE

In order to investigate the possible existence of vortex lines quantized in units of $h/2m$, the case of complete depletion of the single-particle condensate is now considered. It is only in this case that the pair condensate is not coupled to the single-particle condensate and rotates with it. The condition that the pair-condensate wavefunction be single-valued then leads to the half-integral h/m circulation quantization condition.

Equation (3.6) has the trivial solution $\phi = 0$ corresponding to complete depletion of the single-particle condensate. Equations (3.7) and (3.8) for the bogolon wavefunction admit the solutions

$$u(x) = u(r)e^{ikz}e^{i(l+\frac{1}{2})\theta}e^{-iEt} \tag{7.1}$$

and

$$v(x) = v(r)e^{ikz}e^{i(l-\frac{1}{2})\theta}e^{-iEt}. \tag{7.2}$$

[39] E. P. Gross, Nuovo Cimento **20**, 454 (1961).

The equations are then separable and the radial parts are

$$-\epsilon u = u'' + \xi^{-1}u' + [-\xi^{-2}(l+\tfrac{1}{2})^2 - \kappa^2 + \alpha - 2h]u - gv \quad (7.3)$$

and

$$\epsilon v = v'' + \xi^{-1}v' + [-\xi^{-2}(l-\tfrac{1}{2})^2 - \kappa^2 + \alpha - 2h]v - gu. \quad (7.4)$$

If these two coupled nonlinear equations have nontrivial solutions, a new phase analogous to superconductivity would exist. In the absence of a vortex, $l \pm \frac{1}{2}$ is replaced by l, and the equations reduce to those for a homogeneous boson system with completely depleted zero-momentum state.[15] For the delta function potential of Eq. (6.7) the only solutions are the trivial ones. For other potentials, however, nontrivial solutions may exist.[34] Thus it is worthwhile to investigate Eqs. (3.7) and (3.8) for a more realistic potential.

8. CONCLUSION

The theory developed in this paper is an extension to inhomogeneous systems of the theories developed for homogeneous systems which take pair correlations into account.[15] The previous theories for inhomogeneous systems[20,37] have just been generalizations of Bogoliubov's original theory.[13] The equations obtained here should have applicability in systems with nonnegligible depletion and should serve as a more realistic model of superfluid helium than the Bogoliubov approximation. The effect of the nonlinear terms in the equations is under investigation. The main conclusion of the paper is that half-integral quanta of circulation cannot exist unless there is complete depletion of the single-particle condensate, leaving only a "pair condensate." [40]

ACKNOWLEDGMENTS

This work has benefited greatly from stimulating discussions with Dr. J. R. Schrieffer, Dr. S. Teitler, and Dr. R. Mattuck, and with Mag. scients. H. Smith and C. Fogedby, and to them I would like to express my sincere gratitude. I would like to thank Professor Højgaard Jensen for his hospitality at the H. C. Ørsted Institute and for his interest and encouragement in this work. I would also like to thank Dr. A. L. Fetter for some very helpful comments.

[40] An early version of this paper was prepared in March, 1967, and subsequently revised.

APPENDIX: THE BOGOLON GROUND-STATE ENERGY

The unperturbed ground-state energy H_{00} in Eq. (2.12) is given by

$$\begin{aligned} H_{00} = &\int d\mathbf{x}\phi^*(x)T(\mathbf{x})\phi(x) \\ &+ \tfrac{1}{2}\iint d\mathbf{x}\, d\mathbf{y}\phi^*(x)\phi^*(y)V(\mathbf{x},\mathbf{y})\phi(y)\phi(x) \\ &+ \int d\mathbf{x}\Big\langle \chi^\dagger(x)\Big[T(\mathbf{x}) \\ &+ \int d\mathbf{y}\phi^*(y)V(\mathbf{x},\mathbf{y})\phi(y)\Big]\chi(x)\Big\rangle \\ &+ \iint d\mathbf{x}\, d\mathbf{y}\phi^*(y)\phi(x)V(\mathbf{x},\mathbf{y})\langle\chi^\dagger(x)\chi(y)\rangle \\ &+ \tfrac{1}{2}\iint d\mathbf{x}\, d\mathbf{y}V(\mathbf{x},\mathbf{y})[\phi(x)\phi(y)\langle\chi^\dagger(x)\chi^\dagger(y)\rangle \\ &+ \phi^*(x)\phi^*(y)\langle\chi(y)\chi(x)\rangle \\ &+ \langle\chi^\dagger(x)\chi(y)\rangle\langle\chi^\dagger(y)\chi(x)\rangle \\ &+ \langle\chi^\dagger(x)\chi(x)\rangle\langle\chi^\dagger(y)\chi(y)\rangle \\ &+ \langle\chi^\dagger(x)\chi^\dagger(y)\rangle\langle\chi(y)\chi(x)\rangle], \end{aligned} \quad (A1)$$

where the devion density matrix and two-devion amplitude are given in Eqs. (2.19) and (2.20).

In the case of time-independent canonical transformations, the ground-state energy can be minimized with respect to the functions in it. Minimizing H_{00} with respect to $\phi^*(\mathbf{x})$ gives the equation

$$\frac{\delta H_{00}}{\delta\phi^*(\mathbf{x})} = W(\mathbf{x}) = 0, \quad (A2)$$

where W is defined in Eq. (2.18). Equation (A2) is a generalized time-independent Gross–Pitaevski equation, which has coupling to the bogolon functions. This procedure is also equivalent to setting the "dangerous diagram" in Eq. (2.17) equal to zero.

Since there are constraints on the bogolon wavefunctions given by Eqs. (2.10) and (2.11), it is necessary to use Lagrangian multipliers and vary

$$\Lambda = H_{00} + A + B, \quad (A3)$$

where

$$A = \iint d\mathbf{x}\, d\mathbf{y}A(\mathbf{x},\mathbf{y})\Big\{\sum_n [u_n(\mathbf{x})u_n^*(\mathbf{y}) - v_n^*(\mathbf{x})v_n(\mathbf{y})] - \delta(\mathbf{x}-\mathbf{y})\Big\} \quad (A4)$$

and

$$B = \iint d\mathbf{x}\, d\mathbf{y}\Big\{B(\mathbf{x},\mathbf{y})\sum_n [u_n(\mathbf{x})v_n^*(\mathbf{y}) - v_n^*(\mathbf{x})u_n(\mathbf{y})] + \text{c.c.}\Big\}. \quad (A5)$$

The functional derivative of Λ with respect to $v_n^*(\mathbf{x})$ should be zero:

$$0 = \frac{\delta\Lambda}{\delta v_n^*(\mathbf{x})}$$
$$= \int d\mathbf{y} U^*(\mathbf{x},\mathbf{y}) v_n(\mathbf{y}) + \tfrac{1}{2}\int d\mathbf{y}\Delta(\mathbf{x},\mathbf{y}) u_n(\mathbf{y})$$
$$- \int d\mathbf{y} A(\mathbf{x},\mathbf{y}) v_n(\mathbf{y}) + \int d\mathbf{y}[B(\mathbf{y},\mathbf{x}) - B(\mathbf{x},\mathbf{y})] u_n(\mathbf{y}). \quad \text{(A6)}$$

The functional derivative of Λ with respect to $u_m^*(\mathbf{x})$ should also be zero for minimum energy, that is,

$$0 = \frac{\delta\Lambda}{\delta u_m^*(\mathbf{x})}$$
$$= \tfrac{1}{2}\int d\mathbf{y}\Delta^*(\mathbf{x},\mathbf{y}) v_m(\mathbf{y}) + \int d\mathbf{y} A(\mathbf{y},\mathbf{x}) u_m(\mathbf{y})$$
$$+ \int d\mathbf{y}[B^*(\mathbf{x},\mathbf{y}) - B^*(\mathbf{y},\mathbf{x})] v_m(\mathbf{y}). \quad \text{(A7)}$$

If Eq. (A6) is multiplied by $u_m(\mathbf{x})$ and integrated, Eq. (A7) is multiplied by $v_n(\mathbf{x})$ and integrated, and the two equations are added, the Lagrangian multiplier A can be eliminated to yield

$$0 = \iint d\mathbf{x}\, d\mathbf{y} u_m(\mathbf{x}) U^*(\mathbf{x},\mathbf{y}) v_n(\mathbf{y})$$
$$+ \tfrac{1}{2}\iint d\mathbf{x}\, d\mathbf{y} u_m(\mathbf{x})\Delta(\mathbf{x},\mathbf{y}) u_n(\mathbf{y})$$
$$+ \tfrac{1}{2}\iint d\mathbf{x}\, d\mathbf{y} v_n(\mathbf{x})\Delta^*(\mathbf{x},\mathbf{y}) v_m(\mathbf{y})$$
$$+ \tfrac{1}{2}\iint d\mathbf{x}\, d\mathbf{y}\{[u_n(\mathbf{x}) B(\mathbf{x},\mathbf{y}) u_m(\mathbf{y})$$
$$+ v_n(\mathbf{x}) B^*(\mathbf{x},\mathbf{y}) v_m(\mathbf{y})] - [n \leftrightarrow m]\}. \quad \text{(A8)}$$

In order to eliminate the Lagrangian multiplier B, it is necessary to add Eq. (A8) to itself with the indices interchanged; the result is

$$0 = \tfrac{1}{2}\iint d\mathbf{x}\, d\mathbf{y}\{u_m(\mathbf{x}) U^*(\mathbf{x},\mathbf{y}) v_n(\mathbf{y})$$
$$+ u_n(\mathbf{x}) U^*(\mathbf{x},\mathbf{y}) v_m(\mathbf{y}) + u_m(\mathbf{x})\Delta(\mathbf{x},\mathbf{y}) u_n(\mathbf{y})$$
$$+ v_n(\mathbf{x})\Delta^*(\mathbf{x},\mathbf{y}) v_m(\mathbf{y})\}. \quad \text{(A9)}$$

Equation (A9) is the same as Eq. (2.22) for h_{02} when use is made of Eq. (2.23) for U. Therefore, the variation principle leads to the same result as "compensating the dangerous diagrams" in lowest order in the time-independent case.

Equation (A9) can be satisfied if

$$E_n u_n(\mathbf{x}) = \int d\mathbf{y} U(\mathbf{x},\mathbf{y}) u_n(\mathbf{y}) + \int d\mathbf{y}\Delta^*(\mathbf{x},\mathbf{y}) v_n(\mathbf{y}) \quad \text{(A10)}$$

and

$$-E_n v_n(\mathbf{x}) = \int d\mathbf{y} U^*(\mathbf{x},\mathbf{y}) v_n(\mathbf{y}) + \int d\mathbf{y}\Delta(\mathbf{x},\mathbf{y}) u_n(\mathbf{y}). \quad \text{(A11)}$$

The orthogonality condition

$$\int d\mathbf{x}[u_m(\mathbf{x}) v_n(\mathbf{x}) - u_n(\mathbf{x}) v_m(\mathbf{x})] = 0 \quad \text{(A12)}$$

must also be imposed if Eq. (A9) is to be satisfied. The E_n in Eqs. (A10) and (A11) is the bogolon energy since Eq. (2.21) becomes

$$h_{11}(n, m) = E_n \delta_{nm}, \quad \text{(A13)}$$

if the additional orthonormality condition

$$\int d\mathbf{x}[u_n^*(\mathbf{x}) u_m(\mathbf{x}) - v_m(\mathbf{x}) v_n^*(\mathbf{x})] = \delta_{nm} \quad \text{(A14)}$$

is satisfied.

Thus, by minimizing the ground-state energy in Eq. (A1), all of the dangerous diagrams in Eq. (2.12) have been eliminated and the Hamiltonian is in diagonal form if H_{int} is neglected. Equations (A10) and (A11) are the same as Eqs. (3.7) and (3.8) when steady-state solutions are used.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Reciprocity Principle and the Lorentz Transformations

VITTORIO BERZI AND VITTORIO GORINI
Istituto di Scienze Fisiche dell'Università, Milano, Italy, and Istituto Nazionale di Fisica Nucleare, Sezione di Milano, Milan, Italy

(Received 7 June 1968)

By using the principle of relativity, together with the customary assumptions concerning the nature of the space–time manifold in special relativity, namely, space–time homogeneity and isotropy of space, a simple but rigorous proof is given of the reciprocity relation for the relative motion of two inertial frames of reference, which is usually assumed as a postulate in the standard derivations of the Lorentz transformations without the principle of invariance of light velocity. A critical discussion is set forth of the question of eliminating the transformations with invariant imaginary velocity, which one unavoidably obtains together with the Lorentz transformations and the Galilean ones in adopting a procedure of this kind.

I. INTRODUCTION

Since the appearance of the classical Einstein paper,[1] in which the foundations of the theory of relativity were first laid down, several other derivations of the Lorentz transformations have been published in the attempt to throw full light on the underlying principles and to clarify both the physical content and the mathematical implications of the latter.[2] In particular, it has been shown as far back as 1911 by Frank and Rothe[2a] that the assumption of the existence of an invariant velocity is not necessary in order to arrive at the correct transformation equations. This is rather a remarkable result, since it shows that the principle of relativity (which establishes the equivalence of all inertial frames of reference in regard to the description of physical phenomena) together with the customary assumptions concerning the nature of the space–time manifold in special relativity, namely, its homogeneity and the isotropy of physical space, point towards the existence of a universal constant which has the meaning of an invariant velocity, so that there is no need to introduce this constant into the theory at the beginning.

Without imposing from the outset the principle of constancy of light velocity, many of the existing standard derivations of the Lorentz transformations make more or less explicit use of the so called *reciprocity principle* which, as is well known, states simply that the velocity of an inertial frame of reference S with respect to another inertial frame of reference S' is the opposite of the velocity of S' with respect to S.[3]

The use of this principle is not strictly necessary to the scope, but it has the advantage of greatly simplifying the derivation of the transformation equations, which would otherwise require rather lengthy calculations and the resort to nonelementary results of the theory of Lie transformation groups.[4] It appears, however, that in the existing literature no sufficiently convincing arguments have been put forward to justify the use of the reciprocity principle. Indeed, it is generally assumed as a justification that the reciprocity relation is a consequence of the principle of relativity, *whereas the latter merely implies the invariance of the relation between direct and reciprocal velocity.*

It is the aim of the present paper to give a simple but rigorous deduction of the reciprocity relation, starting from the three basic postulates of the special theory of relativity, namely, the principle of equivalence of inertial frames, the homogeneity of space–time, and the isotropy of space.[5]

[1] A. Einstein, Ann. Phys. **17**, 891 (1905).

[2] The existing literature is very wide and rather unrelated and it would be almost impossible to give a fairly complete summary of it. We draw attention to the following references: (a) P. Frank and H. Rothe, Ann. Phys. **34**, 825 (1911); (b) L. A. Pars, Phil. Mag. **42**, 249 (1921); (c) A. S. Eddington, *The Mathematical Theory of Relativity* (Cambridge University Press, London, 1923), Sec. 4; (d) Y. Mimura and T. Iwatsuki, J. Sci. Hiroshima Univ. **A1**, 111 (1931); (e) V. V. Narliker, Proc. Cambridge Phil. Soc. **28**, 460 (1932); (f) G. J. Whitrow, Quart. J. Math. **4**, 161 (1933); (g) L. R. Gomes, Lincei Rend. **21**, 433 (1935); (h) F. Severi, Proc. Phys.-Math. Soc. Japan **18**, 257 (1936); (i) E. Esclangon, Compt. Rend. **202**, 708 (1936); (j) E. Le Roy, *ibid.* **202**, 794 (1936); (k) V. Lalan, *ibid.* **203**, 1491 (1936); Bull. Soc. Math. France **65**, 83 (1937); (l) G. J. Whitrow and E. A. Milne, Z. Astrophys. **15**, 270 (1938); (m) G. Temple, Quart. J. Math. **9**, 283 (1938); (n) H. E. Ives, Proc. Am. Phil. Soc. **95**, 125 (1951); (o) K. D. Stiegler, Compt. Rend. **234**, 1250 (1952); (p) A. W. Ingleton, Nature **171**, 618 (1953); (q) J. Aharoni, *The Special Theory of Relativity* (Oxford University Press, London, 1965), Chap. 1; (r) V. Fock, *The Theory of Space, Time and Gravitation* (Pergamon Press Ltd., London, 1959), Chap. 1 and Appendix A; (s) H. M. Schwartz, Am. J. Phys. **30**, 697 (1962); *Introduction to Special Relativity* (McGraw-Hill Book Co., New York, 1968), Chap. 3; (t) E. C. Zeeman, J. Math. Phys. **5**, 490 (1964); (u) R. Weinstock, Am. J. Phys. **32**, 261 (1964); **33**, 640 (1965); **35**, 892 (1967); (v) V. Mitvalský, *ibid.* **34**, 825 (1966); (w) E. Drake, *ibid.* **34**, 899 (1966); (x) J. L. Strecker, *ibid.* **35**, 13 (1967); (y) L. J. Eisenberg, *ibid.* **35**, 649 (1967); (z) H. Almström, J. Phys. A (Proc. Phys. Soc.) **1**, 331 (1968); (aa) Ya. P. Terletskii, *Paradoxes in the Theory of Relativity* (Plenum Press, Inc., New York, 1968), Chap. 2.

[3] See, for example, Refs. 2(h–j, q, s, u–w, z).

[4] See, for example, Refs. 2(a, k).

[5] A critical analysis of the literature quoted in Footnote 2 and a general discussion concerning the axiomatic derivation of the extended inhomogeneous Lorentz group is the subject of a forthcoming paper.

Once the reciprocity relation has been established, the transformation equations can easily be deduced, as is well known, by making use of their group property, which follows from the principle of relativity.[6] Nevertheless, it is not superfluous to present this deduction again here, since this gives us an opportunity to discuss critically the arguments which have been put forward in favor of excluding the Lorentz transformations with imaginary invariant velocity, which one unavoidably obtains together with the Galilei and the ordinary Lorentz transformations by following a procedure of this kind.

We confine ourselves throughout this paper to the consideration of a two-dimensional space–time and to transformations which conserve the space–time origin. This implies no loss of generality, since any transformation can always be reduced to a homogeneous velocity transformation along an axis by means of a suitable space–time translation, together with suitable rotations of the space-axes of the two observers.

II. THE RECIPROCITY RELATION

As indicated in the introduction, we start from the following assumptions:

(i) *the principle of relativity, which states the equivalence of all inertial systems as regards the formulation of the laws of nature;*

(ii) *the homogeneity of space–time;*

(iii) *the isotropy of space.*

We denote by x the position at which an event takes place and by t the time at which it happens, as viewed by an inertial observer S, and by x', t' the corresponding space–time coordinates of the same event, as viewed by another inertial observer S'.

The homogeneity assumption comes into our considerations in that it implies that the transformation equations which furnish x' and t' as functions of x and t are linear.[7]

In order to prove this assertion, let us employ the notation ξ for the two-vector (x, t) and write the transformation which connects S to S' as

$$\xi' = f(\xi). \tag{1}$$

Since we have decided to restrict our considerations to transformations which conserve the space–time origin, we should require $f(0) = 0$. Here, however, this condition is dropped for the sake of generality.

The homogeneity of space–time requires that a space–time translation T not affect the relation between the two observers and thus leaves Eq. (1) invariant. Denoting by T_α and $T_{\alpha'}$ the representations of T relative to S and S', respectively, we express this property by the relation

$$f(T_\alpha \xi) = T_{\alpha'} f(\xi) \tag{2}$$

or

$$f(\xi + \alpha) = f(\xi) + \alpha', \tag{3}$$

where $\alpha = (\alpha_x, \alpha_t)$, $\alpha' = (\alpha_{x'}, \alpha_{t'})$, and α' depends on f and α but not on ξ.

Taking $\xi = 0$ in (3), we get

$$f(\alpha) = f(0) + \alpha'. \tag{4}$$

Substituting (4) into (3), we obtain

$$f(\xi + \alpha) = f(\xi) + f(\alpha) - f(0). \tag{5}$$

Subtracting $f(0)$ from both sides and setting $g(\xi) = f(\xi) - f(0)$, we have

$$g(\xi + \alpha) = g(\xi) + g(\alpha), \tag{6}$$

for arbitrary ξ and α. From this equation, provided we only assume that g is continuous at the origin, we get that

$$g(k\xi) = kg(\xi), \tag{7}$$

where k is a real number. The proof is quite standard and is given in Appendix A.

Relations (6) and (7) state that g is linear and homogeneous.

We thus write the relation between the pair (x, t) and the pair (x', t') in the form

$$\begin{aligned} x' &= a(v)x + b(v)t, \\ t' &= c(v)x + d(v)t, \end{aligned} \tag{8}$$

where v denotes the velocity of the frame S' with respect to the frame S. This velocity, which we call the *direct velocity* for the pair (S, S'), is given by

$$v = -b(v)/a(v). \tag{9}$$

For the sake of simplicity, we confine ourselves in the rest of this section to the consideration of the case when the space axes of the two observers have the same orientation and their times flow in the same direction,

[6] See, for example, Refs. 2(q, s, aa).

[7] The question of the linearity of the transformation formulas has long been debated in the literature [see, for example, Refs. 1, and 2(d–g, k, m, r, s, y, aa)]. If one does not impose from the outset the existence of an invariant velocity, then the principle of inertia, which implies that a motion which appears uniform to an inertial observer ($x = a_1 t + a_2$) must appear uniform to any other inertial observer ($x' = a_1' t' + a_2'$), is not sufficient to ensure that the transformations are linear. To obtain this result, an additional assumption is needed; namely, that an event of finite space–time coordinates is transformed into an event of finite space–time coordinates. Further, one has to require that the transformation functions be differentiable up to the third order (compare Ref. 2r). It was Einstein (Ref. 1) who first justified the linearity property by an appeal to space–time homogeneity. The argument, however, is rather obscure. Here we give a simple proof of linearity which utilizes a formulation of the homogeneity principle first given by Lalan in Ref. 2k and which is particularly appealing.

which implies the relations[8]

$$\partial x'/\partial x = a(v) > 0,$$
$$\partial t'/\partial t = d(v) > 0, \tag{10a}$$
$$\partial x/\partial x' = d(v)/[a(v)(d(v) + vc(v))] > 0,$$
$$\partial t'/\partial t = 1/(d(v) + vc(v)) > 0. \tag{10b}$$

Then, supposing that both observers use the same unit of time and the same unit of length,[9] the coefficients of (8) are uniquely determined functions of v which, by the principle of relativity, *do not depend on S*.

Denoting by w the *reciprocal velocity*, namely, the velocity of S with respect to S', we have

$$w = b(v)/d(v) = \varphi(v), \tag{11}$$

and our purpose then is to show that the principles of relativity and of isotropy of space, together with some continuity assumptions to be specified later, are sufficient by themselves to arrive at the conclusion that

$$\varphi(v) = -v. \tag{12}$$

The principle of relativity implies that the set Γ of the allowed velocities of S' relative to S does not depend on S and that the reciprocal velocity is the same function of the direct velocity for all pairs of inertial systems. Hence, together with (11), we can write

$$v = \varphi(w) \tag{13}$$

or

$$\varphi(\varphi(v)) = v. \tag{14}$$

Since $w \in \Gamma$, it is clear from (13) that the range of the function φ is equal to its domain Γ. Then φ is a one-to-one mapping of Γ onto Γ. Indeed, if $\varphi(a) = \varphi(b)$, we get from (14) that $a = \varphi(\varphi(a)) = \varphi(\varphi(b)) = b$.

Contrary to a widely held opinion,[3] Eq. (14) is the *only* condition imposed by the principle of relativity on the function φ. This condition is already strongly restrictive on the possible forms of φ, but it by no means alone implies relation (12). For example, the equation

$$w = \varphi(v) = -v/[1 - (v/c)], \tag{15}$$

which is pertinent to one of the cynematics which are compatible with the principle of relativity,[2k] satisfies relation (14) without having the form (12).

[8] Note that conditions (10b) are *a priori* independent of (10a).

[9] It is easy to devise conceptual experiments by which the standards of length and time of the two observers can be made the same. For example, we can make sure that S and S' use the same time standard, if both observers assume as unit of time the mean life of a given unstable particle measured at rest in the laboratory of each of the two observers.

It is precisely with the hope of eliminating solutions of this kind that we resort to the principle of isotropy of space. The main result of applying this principle is the proof that φ is an odd function of v and we see that this property, together with (14) and a physically reasonable assumption concerning the domain of φ and its continuity properties, is sufficient to obtain the result that the reciprocal velocity is given by (12).

In our case, space is one-dimensional and its isotropy means that no one orientation along the x axis should be considered in preference to the other. This assertion is now made precise by stating the isotropy principle in two equivalent forms. The first has a more formal character and concerns the effect that the inversion of the space axes has on the set of transformations (8). The second one, which might be physically more appealing, is based on simple conceptual experiments of a type frequently employed in discussions of the theory of relativity.

We state the isotropy principle in the first form by asserting that if two frames S and S' are connected by a transformation (8), then the two frames $\bar{S}$ and $\bar{S}'$ obtained from the preceding ones by inverting the direction of the x axis are connected by a transformation of the same type. Therefore,

$$\bar{x}' = a(\bar{v})\bar{x} + b(\bar{v})\bar{t},$$
$$\bar{t}' = c(\bar{v})\bar{x} + d(\bar{v})\bar{t}, \tag{16}$$

where $\bar{v}$ is the velocity of $\bar{S}'$ relative to $\bar{S}$. On the other hand, $\bar{x}' = -x'$, $\bar{t}' = t'$, $\bar{x} = -x$, $\bar{t} = t$, so that

$$\bar{x}' = a(v)\bar{x} - b(v)\bar{t},$$
$$\bar{t}' = -c(v)\bar{x} + d(v)\bar{t}, \tag{17}$$

from which we conclude that $\bar{v} = b(v)/a(v) = -v$. Hence, Γ is symmetric and, by comparison with (17),

$$\begin{aligned} a(-v) &= a(v), \\ b(-v) &= -b(v), \\ c(-v) &= -c(v), \\ d(-v) &= d(v). \end{aligned} \tag{18}$$

Then, by (11),

$$\varphi(-v) = -\varphi(v), \tag{19}$$

i.e., φ is an odd function of v.

Consider now the following conceptual experiments:

1. Let T be a rod at rest in S', the end points of which occupy the positions x_1' and x_2'. S measures the length of T by marking the positions x_1 and x_2 that the end points of the rod occupy at a given time t. From the first of the equations in (8) and, again, in (10a) we see that the ratio between the length l' of the rod at

rest (as measured by S') and the length l of the rod in motion (as measured by S) is given by

$$l'/l = a(v). \tag{20}$$

2. Next, let Φ be a phenomenon which takes place at the point x and lasts from time t_1 to time t_2 as observed by S (e.g., we may think of the life of an unstable particle produced at rest at x at time t_1 and decaying at time t_2). By the second of the equations in (10a), the same phenomenon, as observed by S', starts at x_1' at time t_1' and ends at x_2' at time t_2', where (x_1', t_1') and (x_2', t_2') are the transformed coordinates of (x, t_1) and (x, t_2), respectively (in our example, the particle is produced in flight at x_1' at time t_1' and then moves to point x_2', where it decays at time t_2'). From the second of the equations in (8) we see that the ratio between the durations D' and D of Φ (lifetimes of the particle) as measured by S' and S, respectively, is given by

$$D'/D = d(v). \tag{21}$$

3. Finally, let $\tilde{\Phi}$ be another phenomenon which takes place at the point x' and lasts from time t_1' to time t_2' as observed by S', and let (x_1, t_1) and (x_2, t_2) be the transformed coordinates of (x', t_1') and (x', t_2'), respectively. The duration of Φ as measured by S can be determined by means of the following equations:

$$\begin{aligned} a(v)x_1 + b(v)t_1 &= a(v)x_2 + b(v)t_2, \\ t_1' &= c(v)x_1 + d(v)t_1, \\ t_2' &= c(v)x_2 + d(v)t_2, \end{aligned} \tag{22}$$

and, by (9) and by the second of the equations in (10b), the ratio of the durations $\tilde{D}'$ and $\tilde{D}$ of $\tilde{\Phi}$ as measured by the two observers is readily seen to be

$$\tilde{D}'/\tilde{D} = d(v) + vc(v). \tag{23}$$

We state the second version of the principle of isotropy of space by assuming that if v is an allowed velocity, $-v$ is allowed as well (hence the symmetry of Γ), and by requiring that the ratios (20), (21), and (23) are independent of the direction of the motion of S' relative to S, provided that the magnitude of the velocity remains the same, and thus are left unaltered when v is changed to $-v$.[10] This condition implies

$$\begin{aligned} a(-v) &= a(v), \\ d(-v) &= d(v), \\ d(-v) - vc(-v) &= d(v) + vc(v). \end{aligned} \tag{24}$$

Taking into account (9), we see that relations (24) are equivalent to relations (18), and then (19) follows.

We make two further assumptions before we derive the reciprocity relation. These are:

(a) The domain Γ of the function φ is an interval on the real line;

(b) φ is continuous on Γ.

In other words, it is assumed that if v_1 and v_2 are two allowed velocities, any velocity v which is comprised between v_1 and v_2 is again allowed, and that the reciprocal velocity is a continuous function of the direct velocity. The physical plausibility of these two conditions is obvious.

Since φ is a continuous one-to-one mapping of Γ onto itself and Γ is connected, then, from a well-known theorem of analysis, we can state that φ is either a strictly increasing or a strictly decreasing function of v.[11]

Suppose first that φ is strictly increasing. Let $v \in \Gamma$; then $w = \varphi(v) \in \Gamma$. Assume that $v < w$; then $\varphi(v) < \varphi(w)$ and, by (14), $w < v$, which is absurd. We can conclude in the same way $v > w$, so that

$$\varphi(v) = v. \tag{25}$$

If φ is supposed to be strictly decreasing, set $\psi = -\varphi$. Then ψ is strictly increasing and, by (19) and (14), satisfies $\psi(\psi(v)) = v$. Applying to ψ the same argument as before, we obtain $\psi(v) = v$, i.e., Eq. (12).

The choice of Eq. (25) leads to the transformation formulas

$$\begin{aligned} x' &= a(v)x - va(v)t, \\ t' &= c(v)x - a(v)t, \end{aligned} \tag{26}$$

while the choice of Eq. (12) leads to

$$\begin{aligned} x' &= a(v)x - va(v)t, \\ t' &= c(v)x + a(v)t. \end{aligned} \tag{27}$$

Formulas (26) are incompatible with (10). Hence, for two observers whose space axes have the same orientation and whose times flow in the same direction, (12) must necessarily hold and the transformation formulas are given by (27).

Our task of proving the reciprocity relation has thus been completed.

III. EXPLICIT FORM OF THE TRANSFORMATION EQUATIONS

Formulas (27) contain the two as yet undetermined functions $a(v)$ and $c(v)$. However, it is seen at once that $c(v)$ can be expressed in terms of v and $a(v)$.

[10] The same requirement, as regards only the ratio (20), was originally imposed by Frank and Rothe in their derivation of the Lorentz transformations (cf. Ref. 2a).

[11] See, for example, J. Dieudonné, *Foundations of Modern Analysis* (Academic Press Inc., New York, 1960), Theorem 4.2.2. For the reader's convenience the proof is given with some detail in Appendix B.

Indeed, consider the inverse transformations

$$x = \Delta^{-1}(v)a(v)x' + \Delta^{-1}(v)va(v)t',$$
$$t = -\Delta^{-1}(v)c(v)x' + \Delta^{-1}(v)a(v)t', \qquad (28)$$

where

$$\Delta(v) = a(v)\{a(v) + vc(v)\}. \qquad (29)$$

By the reciprocity relation, (28) can also be written in the form

$$x = a(-v)x' + va(-v)t',$$
$$t = c(-v)x' + a(-v)t', \qquad (30)$$

whereby, using (18),

$$a(v) = 1/(a(v) + vc(v)),$$

so that

$$c(v) = (1/v)\{a^{-1}(v) - a(v)\}.$$

Then the transformations (27) read

$$x' = a(v)x - va(v)t,$$
$$t' = (1/v)\{a^{-1}(v) - a(v)\}x + a(v)t. \qquad (31)$$

To interpret (26) we proceed as above, by using (25) instead of (12), and obtain

$$x' = a(v)t - va(v)t,$$
$$t' = -[(1/v)\{a^{-1}(v) - a(v)\}x + a(v)t]. \qquad (32)$$

Hence, (26) is obtained from (27) by an inversion of the time of S', and this explains why (26) corresponds to the choice $\varphi(v) = v$.

The transformation formulas, which connect S to an observer obtained from S' by inverting the orientation of the space-axis, are

$$x' = -[a(v)x - va(v)t],$$
$$t' = (1/v)\{a^{-1}(v) - a(v)\}x + a(v)t, \qquad (33)$$

whereas if S' is subjected to both a space and a time inversion, then

$$x' = -[a(v)x - va(v)t],$$
$$t' = -[(1/v)\{a^{-1}(v) - a(v)\}x + a(v)t]. \qquad (34)$$

It is the essence of the principle of relativity that the set of *all* transformations (31)–(34), as v varies in Γ, forms a group $\mathfrak{L}$. From this property one can derive the explicit form of $a(v)$. In fact, let us compose two transformations of type (31), such that

$$\begin{pmatrix} a(v) & -va(v) \\ (1/v)(a^{-1}(v) - a(v)) & a(v) \end{pmatrix}\begin{pmatrix} a(v') & -v'a(v') \\ (1/v')(a^{-1}(v') - a(v')) & a(v') \end{pmatrix}$$
$$= \begin{pmatrix} a(v)a(v') - (v/v')a(v)(a^{-1}(v') - a(v')) & -(v + v')a(v)a(v') \\ (a(v')/v)(a^{-1}(v) - a(v)) + (a(v)/v')(a^{-1}(v') - a(v')) & a(v)a(v') - (v'/v)a(v')(a^{-1}(v) - a(v)) \end{pmatrix}$$
$$= A(v, v'). \qquad (35)$$

The resulting transformation must be of one of the four types (31)–(34) for some relative velocity v''. However, since the determinant of transformations (31) and (34) is $+1$, whereas the determinant of transformations (32) and (33) is -1, then types (32) and (33) must be ruled out.

Since the diagonal elements of the matrices of both transformations (31) and (34) are equal, we must have

$$a(v)a(v') - (v/v')a(v)(a^{-1}(v') - a(v')) = a(v)a(v') - (v'/v)a(v')(a^{-1}(v) - a(v)),$$

i.e.,

$$(1/v^2)\{1 - a^{-2}(v)\} = (1/v'^2)\{1 - a^{-2}(v')\}, \qquad (36)$$

whence

$$(1/v^2)\{1 - a^{-2}(v)\} = K, \qquad (37)$$

where K is a universal constant having the dimensions of an inverse-square velocity. Then, since $a(v)$ is positive,

$$a(v) = 1/(1 - Kv^2)^{\frac{1}{2}}. \qquad (38)$$

The composite velocity v'' is the negative ratio between the second and the first element of the matrix $A(v, v')$:

$$v'' = (v + v')/[1 - (v/v')\{a^{-2}(v') - 1\}]$$
$$= (v + v')/(1 + Kvv'). \qquad (39)$$

Three cases are to be considered:

(A) $K > 0$. Set $c = (K)^{-\frac{1}{2}}$ and formulas (31) become

$$x' = [1 - (v^2/c^2)]^{-\frac{1}{2}}x - \{v[1 - (v^2/c^2)]^{-\frac{1}{2}}\}t,$$
$$t' = -\{(v/c^2)[1 - (v^2/c^2)]^{-\frac{1}{2}}\}x + [1 - (v^2/c^2)]^{-\frac{1}{2}}t, \qquad (40)$$

and v varies in the domain $\Gamma = (-c, c)$. Equations (40) are the ordinary proper orthochronous Lorentz transformations.

(B) $K = 0$. Formulas (31) become

$$x' = x - vt,$$
$$t' = t, \qquad (41)$$

and v varies in the domain $\Gamma = (-\infty, +\infty)$. These are the Galilean transformations.

(C) $K < 0$. Set $c = (-K)^{-\frac{1}{2}}$ and formulas (31) become

$$x' = [1 + (v^2/c^2)]^{-\frac{1}{2}}x - \{v[1 + (v^2/c^2)]^{-\frac{1}{2}}\}t,$$
$$t' = \{(v/c^2)[1 + (v^2/c^2)]^{-\frac{1}{2}}\}x + [1 + (v^2/c^2)]^{-\frac{1}{2}}t, \qquad (42)$$

and v varies in the domain $\Gamma = (-\infty, +\infty)$.

Set $x^1 = x$, $x^0 = ct$, and $tg\alpha = v/c$ $(-\pi/2 < \alpha < \pi/2)$, and (42) becomes

$$x'^1 = (\cos\alpha)x^1 - (\sin\alpha)x^0,$$
$$x'^0 = (\sin\alpha)x^1 + (\cos\alpha)x^0. \tag{43}$$

Hence, in contrast to the Lorentz transformations (40), which, as is well known, are hyperbolic rotations in the plane (x, t), transformations (42) are ordinary circular rotations. Since α is confined to the interval $(-\pi/2, \pi/2)$, it is clear that *they do not form a group*. If we let α vary from $-\pi/2$ to $3\pi/2$, so as to obtain the full group, we can easily see that we are led to introduce also the transformations

$$x' = -([1 + (v^2/c^2)]^{-\frac{1}{2}}x - \{v[1 + (v^2/c^2)]^{-\frac{1}{2}}\}t),$$
$$t' = -(\{(v/c^2)[1 + (v^2/c^2)]^{-\frac{1}{2}}\}x + [1 + (v^2/c^2)]^{-\frac{1}{2}}t), \tag{44}$$

which are obtained from (42) by inverting both the space and the time axis of S'.

The rotation group (43) translates into mathematical form a complete isotropy of space–time, so that the two directions in time are completely equivalent as well as the two directions in space. On the other hand, if one believes that *there is an intrinsic arrow in the direction of flow of time*, so that time reversal is regarded as a purely mathematical operation which cannot be physically realized, one obtains a strong argument to rule out the transformations (42). Close to this argument is the one set forth by Lalan,[2k] who postulates that if two events take place at the same point in space with respect to a given observer, their time order must be the same for all observers. Alternatively, we could postulate that the relation which states that the space axes have the same orientations and that the times flow in the same direction is transitive, which amounts to assuming that the set of proper orthochronous transformations (31) is by itself a group.

Two other curious features of transformations (43) can be read out in the formula of composition of velocities (39) which, in the present case, has the form

$$v'' = (v + v')/[1 - (vv'/c^2)]. \tag{45}$$

First, by composing two finite velocities v and v' such that $vv' = c^2$, one obtains an infinite velocity v''. Second, by composing two positive velocities v and v' such that $vv' > c^2$, one obtains a negative velocity v''. Some authors[2s,w] use these properties as an argument to exclude transformations (42). In our opinion, however, an argument of this kind is not so convincing as the preceding ones in such a general context, because there are not sufficient reasons of principle to exclude the appearance of phenomena such as those described above. Besides, it is to be noted that peculiarities of this type also appear in the Lorentz case, for which (39) reads

$$v'' = (v + v')/[1 + (vv'/c^2)]. \tag{46}$$

Indeed, if, following some recent ideas,[12] one conjectures the existence of faster-than-light particles (tachyons) and interprets (46) as the transformation formula for the tachyon velocity (v' = particle velocity as measured by S'; v'' = particle velocity as measured by S), it is easily seen that, fixing v very small and negative, we can transform a very large, greater than c, and positive v' into a very large, greater than c, and negative v''. Further, there always exists a reference frame relative to which a tachyon propagates instantaneously. These features are just as curious as those which have been discussed above in connection with formula (45). Notwithstanding, this has not prevented some authors from considering the possibility that faster-than-light particles really exist, on the grounds that the usual objections to such particles are ultimately found to be unconvincing when subjected to critical analysis.

Once we agree to reject formulas (42), we are left with the problem of the choice between the Lorentz transformations (40) and the Galilean transformations (41). As is well known, the Lorentz transformations admit one and only one invariant velocity which is equal to c. In the limit when this velocity is taken to be infinite, one obtains the Galilean transformations. Hence the above problem of choice can be solved only by experience and involves the search for an invariant velocity in nature. The experimental evidence for the existence of signals which travel with a *finite invariant velocity* (such as the electromagnetic waves in vacuo) leads us to rule out the Galilean transformations in favor of the Lorentz ones. In these, of course, the numerical value to be assigned to c is the experimentally measured value of this invariant velocity, namely, the value of the velocity of propagation of electromagnetic disturbances in empty space.

Formally, the rotation transformations (42) correspond instead to the appearance of an invariant imaginary velocity c. This is expressed by the property that they are the linear transformations which conserve the positive-definite quadratic form $x^2 + c^2t^2$, while the Lorentz transformations are those which conserve the indefinite form $x^2 - c^2t^2$. In a

[12] O. M. P. Bilaniuk, V. K. Deshpande, and E. C. G. Sudarshan, Am. J. Phys. **30**, 718 (1962); S. Tanaka, Progr. Theoret. Phys. (Kyoto) **24**, 171 (1960); G. Feinberg, Phys. Rev. **159**, 1089 (1967), and unpublished; R. Newton, Phys. Rev. **162**, 1274 (1967); M. E. Arons and E. C. G. Sudarshan, *ibid.* **173**, 1622 (1968).

four-dimensional space–time the corresponding conserved forms are $x^2 + y^2 + z^2 + c^2t^2$ and $x^2 + y^2 + z^2 - c^2t^2$ and the appropriate groups are the orthogonal group in four dimensions $O(4)$ and the Lorentz group $O(3, 1)$. The characteristic of $O(4)$, that a transformation containing both space reflection and time inversion can be joined continuously to the identity, corresponds to the topological property that, while $O(3, 1)$ has four connected components, $O(4)$ has only two.

IV. CONCLUSION

By making use of the principle of relativity and of the isotropy of space, we have deduced in a simple but rigorous way the reciprocity relation for the relative motion of two inertial reference frames, which is usually assumed as a postulate in the standard derivations of the Lorentz transformations without the principle of invariance of light velocity. For completeness we have then given the usual deduction of the transformation equations by using their group property. We have put forward some alternative arguments to rule out the transformations with invariant imaginary velocity. From a logical viewpoint these arguments might seem more appealing than those previously given by other authors.

APPENDIX A

Let g be a mapping of R^n into itself such that

$$g(\xi + \zeta) = g(\xi) + g(\zeta). \tag{A1}$$

If n is a positive integer, we get by induction, from (A1), that

$$g(n\xi) = ng(\xi). \tag{A2}$$

As $g(0) = 0$, $g(-\xi) = -g(\xi)$, so that (A2) holds equally well for n any integer.

Next, for any rational $r = m/n$, set $m\xi = n\eta$. Then

$$mg(\xi) = g(m\xi) = g(n\eta) = ng(\eta)$$

and thus

$$g(r\xi) = rg(\xi). \tag{A3}$$

Assume now that g is continuous at the origin. This property, together with (A1), implies that g is continuous everywhere. Then, let k be any real number and $\{k_n\}$ be a sequence of rationals which converges to k. So

$$k_n\xi \xrightarrow[n\to\infty]{} k\xi$$

and, by continuity,

$$g(k_n\xi) \xrightarrow[n\to\infty]{} g(k\xi).$$

But

$$g(k_n\xi) = k_n g(\xi) \xrightarrow[n\to\infty]{} kg(\xi),$$

so that

$$g(k\xi) = kg(\xi). \tag{A4}$$

(A1) and (A4) state that g is an endomorphism of the vector space $\mathbf{R}^n$.

APPENDIX B

We recall the following two results of general topology [cf. Ref. (11), Theorems 3.19.7 and 3.19.1].

Proposition 1: The continuous image of a connected topological space is connected.

Proposition 2: A necessary and sufficient condition for a subset A of the real line to be connected is that A is an interval.

In the following, if s and t are any two real numbers, $[s, t]$ will denote the closed interval $\{x : s \leq x \leq t\}$, if $s \leq t$, and the closed interval $\{x : t \leq x \leq s\}$, if $t \leq s$.

In order to prove that the mapping φ is strictly monotone, consider two fixed points p and q of Γ such that $p < q$. Since φ is one to one, we can exclude $\varphi(p) = \varphi(q)$ and suppose, for instance, that $\varphi(p) < \varphi(q)$. Let r be any other point of Γ, $r \neq p$, $r \neq q$. We prove

$$r < p \Rightarrow \varphi(r) < \varphi(p), \tag{B1a}$$

$$p < r \Rightarrow \varphi(p) < \varphi(r), \tag{B1b}$$

and

$$r < q \Rightarrow \varphi(r) < \varphi(q), \tag{B2a}$$

$$q < r \Rightarrow \varphi(q) < \varphi(r). \tag{B2b}$$

Indeed, let, for example, $p < r$. We have $\varphi(r) \neq \varphi(p)$, as implied by φ being one to one, and suppose it to be $\varphi(r) < \varphi(p)$. Since φ is continuous, by propositions 1 and 2, $\varphi([r, q])$ is an interval, so that $\varphi([r, q]) \supseteq [\varphi(r), \varphi(q)]$, whereby $\varphi(p) \in \varphi([r, q])$ because $\varphi(r) < \varphi(p) < \varphi(q)$. Then there is a $p' \in [r, q]$ such that $\varphi(p') = \varphi(p)$, and this is incompatible with φ being one to one because $p' \neq p$, as implied by $p < r$, $p < q$. (B1b) is thus proved.

(B1a), (B2a), and (B2b) are proved in a similar way.

Let now y and y' be any two points of Γ with $y < y'$. Choose s such that $y < s < y'$. Three cases are possible: $s = p$, $s < p$, and $p < s$. In the first case, apply (B1a) and (B1b) to get $\varphi(y) < \varphi(y')$. In the second case, (B1a) implies $\varphi(s) < \varphi(p)$ and we can again apply (B1a) and (B1b) with s in place of p to obtain $\varphi(y) < \varphi(y')$. In the last case, (B1b) gives $\varphi(p) < \varphi(s)$, and use of (B2a) and (B2b) with s in place of q gives again $\varphi(y) < \varphi(y')$. Hence $y < y' \Rightarrow \varphi(y) < \varphi(y')$, and φ is strictly increasing. One can show in the same way that the alternative $\varphi(q) < \varphi(p)$ implies that φ is strictly decreasing.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Equal-Time Pseudolimits of Current Commutators and Schwinger Terms*

PAUL OTTERSON†
Center for Theoretical Physics, Department of Physics and Astronomy
University of Maryland, College Park, Maryland

(Received 21 February 1969)

The equal-time limits of some current commutators are studied in perturbation theory. Although the orthodox equal-time limit does not exist as a distribution, a Schwinger term may be defined as a pseudolimit—a continuous linear functional on a proper subspace of $\mathcal{S}(R^3)$. The results are compared with those obtained from alternative approaches. Simple examples are given to explain why these methods give incorrect results for the cases we consider.

1. INTRODUCTION

A number of authors[1,2] have calculated current commutators and Schwinger terms directly in space–time without using spectral representations or Fourier transformations.[3]

In case of lowest-order (scalar or spinor) quantum electrodynamics, such calculations are equivalent (see the Appendix) to defining equal-time pseudolimits [i.e., limits defined only on a proper subspace of the Schwartz space $\mathcal{S}(R^3)$] of products of free-field Green's functions. These cases are of considerable interest for (i) they contain the most singular Schwinger terms known, (ii) the mathematics may be made perfectly rigorous, and (iii) they thus provide stringent testing grounds for various methods of calculation.

The purpose of this paper is to study these examples in detail—we study the precise mechanism by which various methods of calculation succeed or fail in giving correct Schwinger terms. It should be noted that there exist examples in which the equal-time limiting process does not commute with a spatial limiting process used to define the currents.[4] We thus contribute to an understanding of the conditions under which each method may be appropriately used. We now outline the contents of the following sections.

In Sec. 2 we describe a simple and correct method of calculation based on expressing the Green's functions as limits of a parameterized family of continuous functions, multiplying these functions, and taking the limits (as distributions) of the products. This is done for all nonzero values of the time (a calculation which does not appear to have been carried out previously[5]). One then takes the equal-time pseudolimit to obtain Schwinger terms.

In Sec. 3A we study various previous methods, generalizing them to work for nonzero times. Among others, we study the method of Brandt,[2] which works in a great number of problems. However, as he points out, it gives the wrong coefficient for the c-number Schwinger term in spinor quantum electrodynamics. (Once this is fixed, the method works to all orders in perturbation theory.) We find a simple explanation for the necessity of taking symmetric averages when using these methods.

In Sec. 3B we search for a quick and easy short-cut method of calculation—in the spirit of previous efforts, we try to apply each method directly at equal times. All our attempts fail to give correct coefficients. We try to make this "understandable" with the help of simple examples.

Section 4 contains a few concluding remarks and observations.

2. THE "$i\epsilon$" METHOD

A. $t \neq 0$ Calculations

To order e_0^2 in scalar electrodynamics one finds formally (see the Appendix)

$$\begin{aligned} C^0_{\mu\nu}(x) &\equiv \langle 0|\,[j_\mu(\tfrac{1}{2}x), j_\nu(-\tfrac{1}{2}x)]\,|0\rangle \\ &= 2e_0^2[\Delta_{+,\mu}(x;m)\Delta_{+,\nu}(x;m) \\ &\quad - \Delta_{-,\mu}(x;m)\Delta_{-,\nu}(x;m) \\ &\quad - \Delta_{+,\mu\nu}(x;m)\Delta_{+}(x;m) \\ &\quad + \Delta_{-}(x;m)\Delta_{-,\mu\nu}(x;m)]. \end{aligned} \tag{2.1}$$

* Part of work to be submitted in partial fulfillment of requirements for Ph.D. degree.

† National Science Foundation Pre-doctoral Fellow.

[1] J. Schwinger, Phys. Rev. Letters **3**, 296 (1959).

[2] R. A. Brandt, Phys. Rev. **166**, 1795 (1968).

[3] Closely related work includes the following: T. Nagylaki, Phys. Rev. **158**, 1534 (1967); B. Hamprecht, Nuovo Cimento **50A**, 449 (1967); K. Johnson and F. E. Low, Progr. Theoret. Phys. (Kyoto) Suppl. **38**, 74 (1966); J. Langerholc, Deutsches Elektron-Synchrotron, Hamburg, Report No. 66/24, 1966; J. C. Polkinghorne, Nuovo Cimento **52A**, 351 (1967); B. Schroer and P. Stichel, Commun. Math. Phys. **8**, 8 (1968); and U. Völkel and A. H. Völkel, Commun. Math. Phys. **7**, 3 (1968).

[4] R. A. Brandt, J. Sucher, and C. H. Woo, Phys. Rev. Letters **19**, 801 (1967).

[5] There is more information than we use here. We show elsewhere how such information can be derived in a more general context from spectral representations and used to derive sum rules.

TABLE I. $t \neq 0$ light-cone singularities of products of Δ's. Where k_1, k_2, and k form a right-handed orthogonal triad, $d_{k\pm} = d_{k_1} + id_{k_2}$ and Ω_k denote the spherical angular coordinates (θ, ϕ) in a system with polar axis $(\theta = 0)$ along k and $\phi = 0$ along k_1. The parameter s is defined by $s = (r + |t|)^{-1}$.

$$(\Delta_+\Delta_+ - \Delta_-\Delta_-)(x;0) = -i(\operatorname{sign} t)(32\pi^3r^2)^{-1}[2s - d_r]\delta(r - |t|)$$

$$(\Delta_+\Delta_{+,tk} - \Delta_-\Delta_{-,tk})(x;0) = -i(32\pi^3r^2)^{-1}(4\pi/3)^{\frac{1}{2}}\, Y_1^0(\Omega_k)[r^{-1}(s^2 + sd_r - \tfrac{1}{2}d_r^2) + (2s^3 - sd_r^2 + (\tfrac{1}{3})d_r^3)]\delta(r - t)$$

$$(\Delta_{+,t}\Delta_{+,k} - \Delta_{-,t}\Delta_{-,k})(x;0) = -i(32\pi^3r^2)^{-1}(4\pi/3)^{\frac{1}{2}}\, Y_1^0(\Omega_k)[r^{-1}(s^2 + sd_r - \tfrac{1}{2}d_r^2) + (\tfrac{1}{6})d_r^3]\delta(r - |t|)$$

$$(\Delta_{+,t}\Delta_{+,k} - \Delta_{-,t}\Delta_{-,k})(x;0) = \mp i(32\pi^3r^2)^{-1}(8\pi/3)^{\frac{1}{2}}\, Y_1^{\pm1}(\Omega_k)[r^{-1}(s^2 + sd_r - \tfrac{1}{2}d_r^2) + (\tfrac{1}{6})d_r^3]\delta(r - |t|)$$

$$\Delta_+\Delta_{+,kk\pm} - \Delta_-\Delta_{-,kk\pm})(x;0) = \mp i(\operatorname{sign} t)(32\pi^3r^2)^{-1}(8\pi/3)^{\frac{1}{2}}\, Y_2^{\pm1}(\Omega_k)[3r^{-2}(2s - d_r) + 4r^{-1}(s^2 - sd_r + \tfrac{1}{2}d_r^2) + (2s^3 + sd_r^2 - (\tfrac{1}{3})d_r^3)]\delta(r - |t|)$$

$$(\Delta_+\Delta_{+,kk\pm} - \Delta_-\Delta_{-,kk\pm})(x;0) = \mp(\operatorname{sign} t)(32\pi^3r^2)^{-1}(\tfrac{2}{3})(4\pi/5)^{\frac{1}{2}}\, Y_2^0(\Omega_k)[3r^{-2}(2s - d_r) + 4r^{-1}(s^{-2} - sd_r + \tfrac{1}{2}d_r^2) + (2s^3 + sd_r^2 - (\tfrac{1}{3})d_r^3)]\delta(r - |t|)$$
$$\mp i(\operatorname{sign} t)(32\pi^3r^2)^{-1}(\tfrac{1}{3})[r^{-1}(s^2 - sd_r + \tfrac{1}{2}d_r^2) + (2s^3 + sd_r^2 - (\tfrac{1}{3})d_r^3)]\delta(r - |t|)$$

$$(\Delta_+\Delta_{+,tk} - \Delta_-\Delta_{-tk})(x;0) = \pm i(32\pi^3r^2)^{-1}(8\pi/3)^{\frac{1}{2}}\, Y_1^{\pm1}(\Omega_k)[r^{-1}(s^2 + sd_r - \tfrac{1}{2}d_r^2) + (2s^3 - sd_r^2 + (\tfrac{1}{3})d_r^3)]\delta(r - |t|)$$

The analogous expression in spinor electrodynamics is

$$\begin{aligned} C^{\frac{1}{2}}_{\mu\nu}(x) &\equiv \langle 01[j_\mu(\tfrac{1}{2}x), j_\nu(-\tfrac{1}{2}x)]10\rangle \\ &= 4e_0\{(2g^r_\mu g^s_\nu - g_{\mu\nu}g^{rs})[\Delta_{+,r}(x;m)\Delta_{+,s}(x;m) \\ &\quad - \Delta_{-,r}(x;m)\Delta_{-,s}(x;m)] \\ &\quad + m^2 g_{\mu\nu}[\Delta_+(x;m)\Delta_+(x;m) \\ &\quad - \Delta_-(x;m)\Delta_-(x;m)]\}, \end{aligned} \tag{2.2}$$

where, e.g.,

$$\Delta_\pm(x;m) \equiv \frac{1}{(2\pi)^4}\int e^{ikx}\delta(k^2 - m^2)\theta(k_0)\, d^4k,$$

$$\Delta_{\pm,\mu} = \frac{\partial}{\partial x^\mu}\Delta_\pm(x;m). \tag{2.3}$$

The immediate task is to carry out the indicated multiplications of $\Delta_\pm\Delta_\pm$, utilizing only the space–time representations of these distributions.

Let $t \equiv x_0$ and $r = (\mathbf{x}^2)^{\frac{1}{2}}$. Then,

$$\Delta_\pm(x;0) = \lim_{\epsilon\to 0_+}\Delta_\pm(x;0;\epsilon),$$

where

$$\Delta_\pm(x;0;\epsilon) \equiv \left(\frac{-1}{8\pi^2 r}\right)\left(\frac{1}{r + |t|} + \frac{1}{r - |t| \pm i\epsilon \operatorname{sign} t}\right). \tag{2.4}$$

These are distributions on $\mathcal{S}(R^4)$ which may also be considered as t-parameterized distributions on $\mathcal{S}(R^3)$; we adopt the latter interpretation. Now, since the derivative of a limit of distributions is always the limit of the differentiated distributions, we have

$$\partial_\alpha\Delta_\pm(x;0) = \lim_{\epsilon\to 0_+}\Delta_{\pm,\alpha}(x;0,\epsilon)$$

for any distribution-theoretic derivative ∂_α. These differentiated $\Delta_\pm$ distributions may be considered as t-parameterized distributions [on $\mathcal{S}(R^3)$] for every nonzero value of t, which interpretation we again follow.

For each fixed nonzero value of t, let us define the products

$$(\Delta_{+,\alpha}\Delta_{+,\beta})(x;0) \equiv \lim_{\epsilon\to 0_+}(\Delta_{+,\alpha}(x;0;\epsilon)\Delta_{+,\beta}(x;0;\epsilon)) \tag{2.5}$$

and the analogous products of $\Delta_{-,\alpha}\Delta_{-,\beta}$. (The limits do not exist when one tries to define $\Delta_+\Delta_-$ similarly.[6])

The relevant limits may be evaluated by means of the relation

$$\lim_{\epsilon\to 0_+}\left[\frac{1}{(u + i\epsilon)^n} - \frac{1}{(u - i\epsilon)^n}\right] = \frac{-2\pi i}{(n-1)!}\left(\frac{-\partial}{\partial u}\right)^{n-1}\delta(u) \tag{2.6}$$

and are listed in Table I, which, together with Eqs. (2.1) and (2.2), completely specifies the highest $(t \neq 0)$ light-cone singularities of $C^0_{\mu\nu}(x)$ and $C^{\frac{1}{2}}_{\mu\nu}(x)$. We need but take the $t \to 0$ limits to obtain Schwinger terms.

B. $t \to 0$ Limits and Schwinger Terms

For $t \neq 0$, the products of Δ's are distributions on $\mathcal{S}(R^3)$; because of the presence of inverse powers of $(r + |t|)$, they have no corresponding distribution limit as t approaches zero. However, if we restrict attention to testing functions which vanish at least as rapidly as r^3 at the origin, then, on this proper subspace $\mathcal{S}_3 \subset \mathcal{S}(R^3)$, the limit does exist as a continuous linear functional. Let us denote this (pseudo-) limit

[6] The one-dimensional analog is $1/(x + i\epsilon)$.

$$\lim_{\epsilon\to 0_+}\left(\frac{1}{x + i\epsilon}\right)^n$$

does exist as a distribution for each n. However, since x^{-2} is not locally integrable at $x = 0$,

$$\lim_{\epsilon\to 0_+}\left(\frac{1}{x + i\epsilon}\right)\left(\frac{1}{x - i\epsilon}\right)$$

does *not* exist as a distribution.

by $\lim|_3$; the bar and subscript specify the subspace on which the limit is defined.

A typical term entering in C_{0k} is

$$\left(\frac{4\pi}{3}\right)^{\frac{1}{2}} \frac{Y^0(\Omega_k)\delta(r-|t|)}{r^3(r+|t|)^2}, \tag{2.7}$$

where Y_R^m denote spherical harmonics, e.g.,

$$Y_1^0(\Omega_k) = (3/4\pi)^{\frac{1}{2}} \cos\theta_k.$$

Acting on an arbitrary testing function w (which we express in spherical harmonics),

$$w(\mathbf{x}) = \sum_{l,m} w_{l,m}(r) Y_l^{m*}(\Omega_k) = \sum_{n,l,m} w_{n,l,m} r^n Y_l^{m*}(\Omega_k), \tag{2.8}$$

we have, for $t \neq 0$,

$$\begin{aligned}\left[\left(\frac{4\pi}{3}\right)^{\frac{1}{2}} \frac{Y^0(\Omega_k)\delta(r-|t|)}{r^3(r+|t|)^2}\right](w) &= \int_0^\infty r^2\, dr \left(\frac{4\pi}{3}\right)^{\frac{1}{2}} \frac{w(\mathbf{x})Y_1^0(\Omega_k)\delta(r-|t|)\, d\Omega}{r^3(r+|t|)^2} \\ &= \left(\frac{4\pi}{3}\right)^{\frac{1}{2}} \frac{w_{1,0}(|t|)}{4|t|^3},\end{aligned} \tag{2.9}$$

where in this notation, the term in square brackets is a distribution acting on the testing function in parentheses immediately to its right. To calculate $\lim|_3$, consider only testing functions $w(x)$ for which the coefficients $w_{n,l,m}$ vanish for $n < 3$. Thus,

$$\begin{aligned}\left[\lim_{t\to 0_\pm}\Big|_3 \left(\left(\frac{4\pi}{3}\right)^{\frac{1}{2}} \frac{Y_1^0(\Omega_k)\delta(r-|t|)}{r^3(r+|t|)^2}\right)\right](w) &\equiv \lim_{t\to 0_\pm}\left(\left(\frac{4\pi}{3}\right)^{\frac{1}{2}} \left(\frac{w_{1,0}(|t|)}{4|t|^3}\right)\right)\left(\frac{4\pi}{3}\right)^{\frac{1}{2}} (\tfrac{1}{4} w_{3,1,0}) \\ &= \left[\frac{-1}{40}\frac{\partial}{\partial x_k}\nabla^2\delta(\mathbf{x})\right](w),\end{aligned} \tag{2.10}$$

which is to say

$$\lim_{t\to 0_\pm}\Big|_3 \left(\frac{\delta(r-|t|)Y_1^0(\Omega_k)}{r^3(r+|t|)^2}\left(\frac{4\pi}{3}\right)^{\frac{1}{2}}\right) = \frac{-1}{40}\frac{\partial}{\partial x_k}\nabla^2\delta(\mathbf{x}) \tag{2.11}$$

(where ∇^2 denotes the Laplacian operator).

Evaluating the remaining terms in similar fashion, we find, for example,

$$\begin{aligned}\lim_{t\to 0_\pm}\Big|_3 (\Delta_{+,0}\Delta_{+,k} - \Delta_{-,0}\Delta_{-,k})(x;0) &= \tfrac{1}{2}\lim_{t\to 0_\pm}\Big|_3 (\Delta_{+,0k}\Delta_+ - \Delta_{-,0k}\Delta_-)(x;0) \\ &= \frac{-i}{96\pi^2}\frac{\partial}{\partial x_k}\nabla^2\delta(\mathbf{x}).\end{aligned} \tag{2.12}$$

Combining these limits with expressions (2.1) and (2.2), we obtain the Schwinger terms

$$\lim_{t\to 0_\pm}\Big|_3 C_{0k}^0(x) = \tfrac{1}{4}\lim_{t\to 0_\pm}\Big|_3 C_{0k}^{\frac{1}{2}}(x) = \frac{ie_0^2}{48\pi^2}\frac{\partial}{\partial x_k}\nabla^2\delta(\mathbf{x}). \tag{2.13}$$

This result agrees with the value for $C_{0k}^{\frac{1}{2}}$ obtained by Brandt using spectral functions.[2]

For all other Lorentz indices, $C_{\mu\nu}$ is antisymmetric in t, and $\lim|_3 C_{\mu\nu}$ is zero on $\mathcal{S}_3$. For example,

$$\lim_{t\to 0_+}\Big|_3 C_{00}^0(x) = -\lim_{t\to 0_-}\Big|_3 C_{00}^0(x) \simeq r^{-2}\left(\frac{\partial}{\partial r}\right)^3\delta(r) = 0, \tag{2.14}$$

where the last equality holds because $\mathcal{S}(R^3)$ contains no spherically symmetric function going as r^k for k odd. [r itself is not an element of $\mathcal{S}(R^3)$. On a suitably enlarged space of testing functions the limits are different and nonzero.] This situation naturally reverses with each new time derivative.

Thus consider $dC_{\mu\nu}(x)/dx_0$: The $(0, \mathbf{k})$ components are antisymmetric in t and have zero limits on $\mathcal{S}_4$; the $(0, 0)$ and $(\mathbf{k}, \mathbf{k}')$ components are symmetric in t and have nonzero limits as continuous linear functionals on $\mathcal{S}_4 \subset \mathcal{S}(R^3)$.

3. OTHER METHODS AND DISTRIBUTIONS

Let us try to obtain the same results by manipulating with the Green's functions Δ and Δ_1 rather than $\Delta_\pm$. To do this, one may use relations such as

$$(\Delta_+\Delta_+ - \Delta_-\Delta_-) = (\Delta_+ + \Delta_-)(\Delta_+ - \Delta_-) = -i\Delta\Delta_1.$$

Equivalently, in calculating the commutators of currents from their definitions in terms of fields, rather than just writing $[AB, CD] = ABCD - CDAB$ as in the Appendix, one may use commutator identities to obtain products of commutators and anticommutators of free fields, i.e., Δ and Δ_1. (We have already noted that the product $\Delta_+\Delta_-$ appears undefinable, but we may hope that the implicit use of the equation $\Delta_+\Delta_- - \Delta_-\Delta_+ = 0$ will not lead to difficulties.)

There remains the problem of multiplying distributions which have overlapping singularities. A possible approach is to displace one distribution from the other, so that the singularity supports are not so strongly overlapping. Then one multiplies the relatively displaced pair of distributions and calculates the limit as the displacement approaches zero.

As Figs. 1 and 2 indicate, the geometry is generally unfavorable for such methods; there is no real vector displacement which will make two light cones disjoint. Moreover, as the displacement shrinks, the region of difficulty approaches the origin, the very point with

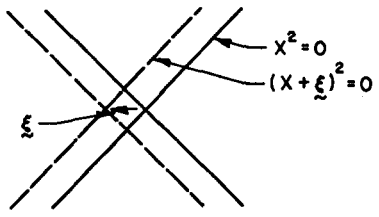

FIG. 1. Two light cones, displaced by a purely spacelike vector, $\xi = (0, \epsilon)$. They intersect in a three-dimensional hyperboloid. At fixed x_0, the singular surfaces are pairs of spheres, each of radius $|x_0|$. (See also Fig. 3.)

which we are ultimately concerned. Finally, except for a purely timelike displacement, the geometry becomes rather complicated and computation becomes difficult. In fact, the real-displacement method was intended only for use on the $t = 0$ plane. Nevertheless, we try these methods for $t \neq 0$, for, as we have seen, the $t \neq 0$ answer is a simple distribution, not merely a linear functional on a proper subspace of $\mathcal{S}(R^3)$. These methods, if they are to work at all, would certainly be expected to work for $t \neq 0$, and any subtleties required to make them work should be more comprehensible for $t \neq 0$. If subtleties are discovered for $t \neq 0$, they may then be taken into account in a $t = 0$ calculation.

A. Real Vector Displacements

1. *Purely Spatial Vector Displacement; $t \neq 0$ Calculation of* $\Delta(x; 0)\Delta_1(x; 0)$

This method is closely related to the limiting procedures of Brandt.[2] In order to multiply $\Delta(x; 0)$ and $\Delta_1(x; 0)$, we consider first the displaced product $\Delta_1(x + \xi; 0)\Delta(x; 0)$ and then let $\xi = (0, \boldsymbol{\xi})$ approach zero. [The answer we are looking for is

$$(\Delta_+(x; 0)\Delta_+(x; 0) - \Delta_-(x; 0)\Delta_-(x; 0))$$

as previously defined.]

With variables as before, letting P denote the Cauchy principal part,

$$\Delta(x; 0) = \frac{(\text{sign } t)}{4\pi r}(\delta(r - |t|)), \tag{3.1}$$

$$\Delta_1(x; 0) = \frac{1}{4\pi^2 r}\frac{1}{(r + |t|)} + P\left(\frac{1}{(r - |t|)}\right). \tag{3.2}$$

The following fact is also useful: If $\{T_\alpha\}$ is a bounded family of distributions and, for each α, w_α is a testing function locally on U_α (supp T_α) and $w_\alpha \to w$ (a testing function) and $T_\alpha \to T$ (a distribution), then

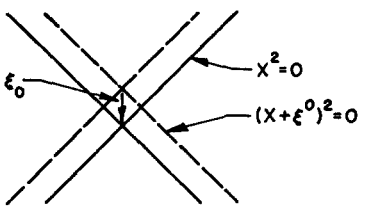

FIG. 2. Two light cones, displaced by a purely timelike vector, $\xi = (\epsilon^0, \mathbf{0})$. They intersect in a three-dimensional sphere of radius $\frac{1}{2}|\xi_0|$ at $x_0 = -\frac{1}{2}\xi_0$. At other values of x_0, the singular surfaces are concentric spheres.

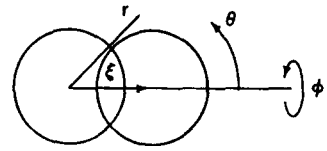

FIG. 3. Coordinate system used in Sec. 3A.

$T_\alpha(w_\alpha) \to T(w)$. The statement remains valid on replacing "distribution" by "continuous linear functional on a subspace of testing functions to which all w belong."

Now let $\Delta_1(x + \xi; 0)\Delta(x; 0)$ act on a testing function $w(x)$ at fixed $t = x_0 > |\boldsymbol{\xi}|$. The natural coordinate system is spherical, with polar axis aligned with $\boldsymbol{\xi}$ (see Fig. 3). A part of the product is given by[7]

$$\left[r^{-1}\delta(r - |t|)\left(\frac{1}{|\mathbf{x} + \boldsymbol{\xi}|}\right)P\left(\frac{1}{|x + \xi| - t}\right)\right](w(r, \Omega))$$
$$\simeq \frac{1}{t}\int d\Omega w(|t|, \Omega)$$
$$\times P\left(\frac{1}{[1 + 2(|\boldsymbol{\xi}|/t)\cos\theta + |\boldsymbol{\xi}|^2/t]^{\frac{1}{2}} - 1}\right). \tag{3.3}$$

Since the denominator of the latter integrand blows up independently of θ ($\neq \pi/2$) as $\boldsymbol{\xi} \to 0$, the expression

$$\lim_{\xi \to 0} \Delta_1(x + \xi; 0)\Delta(x; 0)$$

is meaningless for fixed $t \neq 0$. Suppose one averages over $\pm\xi$ before letting $\boldsymbol{\xi} \to 0$. Then one finds that the principal part contributes zero and (letting $\text{Av}_{\pm\xi}$ denote this average)

$$\lim_{|\xi| \to 0}(\text{Av}_{\pm\xi}\Delta_1(x + \boldsymbol{\xi}; 0)\Delta(x; 0)) = \frac{(\text{sign } t)\delta(r - |t|)}{2|t|^3}. \tag{3.4}$$

This is the wrong answer for $\Delta_1\Delta$, since (in contrast to the first entry in Table I) it contains no derivative of $\delta(r - |t|)$.

$$\Delta(x + \tfrac{1}{2}\xi; 0)\Delta_1(x - \tfrac{1}{2}\xi; 0)$$

behaves differently. We find

$$\lim_{|\xi| \to 0} \text{Av}_{\pm\xi}[\Delta(x + \tfrac{1}{2}\xi; 0)\Delta_1(x - \tfrac{1}{2}\boldsymbol{\xi}; 0)]$$
$$= \frac{\delta(r - |t|)}{2|t|^3} - \frac{1}{2}\frac{\partial}{\partial r}\delta(r - |t|). \tag{3.5}$$

This does give the correct answer as in Table I. Thus,

$$\Delta(x; 0)\Delta_1(x; 0)$$
$$= \lim_{|\xi| \to 0} \text{Av}_{\pm\xi}[\Delta_1(x + \tfrac{1}{2}\xi; 0)\Delta(x - \tfrac{1}{2}\xi; 0)]$$
$$\neq \lim_{|\xi| \to 0} \text{Av}_{\pm\xi}[\Delta(x + \xi; 0)\Delta_1(x)]. \tag{3.6}$$

[7] We have used the fact mentioned in the preceding paragraph to replace $|\mathbf{x} + \boldsymbol{\xi}|^{-1}$ by t^{-1}.

Later, when we work with $t = 0$ directly, we will recall this distinction and attempt to use the former expression.

Here is a culprit to answer for these subtle complications: In the expression

$$[\Delta_+(x) + \Delta_-(x)][\Delta_+(y) - \Delta_-(y)]$$

the $\Delta_+\Delta_-$ terms cancel upon replacing x by $x + \xi$ and y by $x - \xi$ and then averaging over $\pm\xi$. Terms involving the dangerous expression $\Delta_+\Delta_-$ remain unless the averaging is performed.

At this point we leave the $t \neq 0$ real spatial-displacement method—when spatial derivatives are present, geometric problems make it difficult to proceed.

2. *Purely Timelike Displacements; $t \neq 0$ Calculations of $\Delta_1(x;0)\Delta(x;0)$*

In strict analogy with the spatial limiting process, one may consider the expression $\Delta(x + \xi; 0)\Delta_1(x; 0)$ for $\xi = (\xi^0, \mathbf{0})$. At fixed $t \neq 0$, the singularity surfaces are concentric spheres (see Fig. 3). The results are strictly analogous to the spatial limiting results. We find that

$$\begin{aligned}\Delta(x;0)\Delta_1(x;0) &= \lim_{|\xi^0|\to 0} \mathrm{Av}_{\pm\xi^0}[\Delta_1(x + \tfrac{1}{2}\xi^0; 0)\Delta(x - \tfrac{1}{2}\xi^0)] \\ &\neq \lim_{|\xi^0|\to 0} \mathrm{Av}_{\pm\xi^0}[\Delta_1(x + \xi^0; 0)\Delta(x; 0)]. \quad (3.7)\end{aligned}$$

As above, we attribute the inequality to the undefined nature of $\Delta_+\Delta_-$.

Happily, the geometry is not complicated and we may attempt to multiply derivatives of these distributions using real time-like displacements. Upon doing so, we uncover apparently insurmountable new difficulties. These occur simply because there are too many inverse powers of $(r - |t|)$ for a symmetric averaging to remove—the limit, even after averaging, does not exist as a distribution.

B. Calculations Directly at $x_0 = 0$

It would be very nice to be able to obtain the Schwinger terms of Eq. (2.14) by an easier method, such as using the canonical commutation relations for fields. In searching for a method, we take the $t \to 0$ limits of $\Delta_\pm$, Δ, Δ_1, and their derivatives, *then* multiply them. This is purely experimental; there is no *a priori* reason to expect the results to agree with (2.14). (Neither will there be an *a posteriori* reason—the results do not agree.) Similar phenomena have been noted by Brandt, Sucher, and Woo.[3]

Recalling (3.6), we calculate[8]

$$\lim_{|\xi|\to 0}\Big|_3 \mathrm{Av}\left\{\left[\lim_{t\to 0}\Delta_{,0}(x + \tfrac{1}{2}\xi; 0)\right]\left[\lim_{t\to 0}\Delta_{1,k}(x - \tfrac{1}{2}\xi; 0)\right]\right\},$$

where

$$\lim_{t\to 0}\Delta_{,0}(x; 0) = -\delta(x)$$

and

$$\lim_{t\to 0}\Delta_{1,k}(x; 0) = -\frac{\partial}{\partial x_k}\frac{1}{2\pi^2 r^2}, \quad \text{for} \quad r \neq 0. \quad (3.8)$$

Let $w(x) = w(r, \Omega)$ be a testing function in $\mathcal{S}_3 \subset \mathcal{S}(R^3)$. Then,

$$\begin{aligned}&\left[\lim_{|\xi|\to 0}\Big|_3 \mathrm{Av}\frac{1}{2\pi^2}\delta(\mathbf{x} + \tfrac{1}{2}\boldsymbol{\xi})\frac{\partial}{\partial x_k}\frac{1}{|\mathbf{x} - \tfrac{1}{2}\boldsymbol{\xi}|^2}\right][w(x)] \\ &\quad= \left[\frac{1}{2\pi^2}\lim_{|\xi|\to 0}\mathrm{Av}\,\delta(\mathbf{x})\frac{2\hat{\xi}_k}{|\mathbf{x} - \boldsymbol{\xi}|^2}\right][w(\mathbf{x} - \tfrac{1}{2}\boldsymbol{\xi})] \\ &\quad= \frac{1}{2\pi^2}\lim_{|\xi|\to 0}\mathrm{Av}\frac{2\hat{\xi}_k w(-\frac{1}{2}\boldsymbol{\xi})}{|\boldsymbol{\xi}|^3} \\ &\quad= \left[\frac{1}{240\pi^2}\frac{\partial}{\partial x_k}\nabla^2\delta(\mathbf{x})\right](w). \quad (3.9)\end{aligned}$$

Similarly,

$$\lim_{|\xi|\to 0}\Big|_3 \mathrm{Av}\frac{\partial}{\partial x_k}\frac{\delta(x + \frac{1}{2}\xi)}{2\pi^2}\frac{1}{(x - \frac{1}{2}\xi)^2} = \frac{-1}{480\pi^2}\frac{\partial}{\partial x_k}\nabla^2\delta(\mathbf{x}). \quad (3.10)$$

These answers are incorrect by factors of five—that is, they give Schwinger terms with coefficients only one tenth of the correct value for the spinor case and three tenths the correct value in the scalar case.[7]

Analogous results are obtained when one simply sets $t = 0$ then $\epsilon = 0$ in the "$i\epsilon$" method. An example may clarify the mechanism by which the "$i\epsilon$" method breaks down when one sets $t = 0$ formally. Consider the t-parameterized distribution $r^{-2}(r + |t|)^{-n}\delta(t - |t|)$. Letting this act on the space $\mathcal{S}_n$ of testing functions which fall as rapidly as r^n at the origin, we find

$$\lim_{t\to 0\pm}\Big|_n \frac{\delta(r - |t|)}{(r + |t|)^n r^2} = \frac{(-\frac{1}{2})^n}{n!\, r^2}\frac{\partial^n}{\partial r}\delta(r). \quad (3.11)$$

On the other hand, if we just set $t = 0$, divide $\omega \in \mathcal{S}_n \subset \mathcal{S}(R^3)$ by r^n, and define $r^{-(n+2)}\delta(r)$ as a bounded linear functional on $\mathcal{S}_n$ by

$$[r^{-(n+2)}\delta(r)]\big|_n(\omega) = (\omega/r^n)(0), \quad (3.12)$$

[8] Here the average is taken over all angles. When the limit is taken without averaging, the answer varies in the interval $(0, \frac{3}{10})$ times the correct answer in the spinor case [correspondingly in the interval $(0, \frac{6}{10})$ in the scalar case], depending on the direction of $\hat{\xi}$.

then we obtain

$$r^{-(n+2)}\delta(r)\big|_n(\omega) = \frac{(-)^n}{n!}\frac{\partial^n}{\partial r}\frac{\delta(r)}{r^2}(\omega), \quad (3.13)$$

which differs from (3.11) by a factor of $(\frac{1}{2})^n$.

If one considers a real-displacement method augmented with the "$i\epsilon$" method, then the correct order of limiting is

$$\lim_{t\to 0}\left(\lim_{\epsilon\to 0}\lim_{\xi\to 0}\right)$$

and the limits must be taken in precisely the order given.

6. CONCLUDING REMARKS

When a matrix element of a current commutator $C_{\mu\nu}(x)$ may be written as a sum of products of two distributions, such that in each (i) one factor is locally a testing function at the origin and (ii) the other factor has a well-defined equal-time limit, then the equal-time limit of $C_{\mu\nu}(x)$ does exist and may be taken either before or after various spatial limits used in defining currents are taken.

However, in the examples we have studied, condition (ii) is violated. Under these circumstances, we have seen that (a) the equal-time limit of $C_{\mu\nu}(x)$ does not exist as a distribution (but does exist as a suitably defined pseudolimit) and (b) it is not generally permissible to change orders of taking limits.

These examples are sufficiently simple and transparent and reasonably realistic, so that one may regard the phenomena under consideration as understood.[9]

ACKNOWLEDGMENTS

I am greatly indebted to Professor Joseph Sucher and Professor Richard Brandt for suggesting the problem, for many enlightening discussions, and for constant encouragement. I also wish to thank Professor Claudio Orzalesi for a number of helpful discussions.

[9] After this work was completed, we came across a preprint by F. Kashlvhn, E. Wieczorek, and W. Zoellner (reported at the International Seminar on the Theory of Elementary Particles in Varna, 1968), in which a matrix element similar to (2.1) is studied for the case of a charged scalar field coupled to a massive vector field.

APPENDIX A

In scalar electrodynamics, it is convenient to use the independent Hermitian fields ϕ_1 and ϕ_2, defined by

$$\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix} = \frac{1}{(2)^{\frac{1}{2}}}\begin{bmatrix}1 & 1\\ -i & i\end{bmatrix}\begin{pmatrix}\phi\\ \phi^+\end{pmatrix}. \quad (A1)$$

In terms of these, the current $j_\mu \equiv ie_0 :\phi^+(\overrightarrow{d_\mu} - \overleftarrow{d_\mu})\phi:$ may be written

$$j_\mu = e_0(\phi_1(x)\phi_{2,\mu}(x) - \phi_{1,\mu}(x)\phi_2(x)). \quad (A2)$$

(The advantage of using ϕ_1 and ϕ_2 is the absence of normal-order symbols in the latter expression.)

To calculate formally the commutator of two currents, we write, for example,

$$\begin{aligned}&[(\phi_1(x)\phi_{2,\mu}(x)), (\phi_1(y)\phi_{2,\nu}(y))]\\ &= \phi_1(x)\phi_{2,\mu}(x)\phi_1(y)\phi_{2,\nu}(y) - \phi_1(y)\phi_{2,\nu}(y)\phi_1(x)\phi_{2,\mu}(x)\\ &= \phi_1(x)\phi_1(y)\phi_{2,\mu}(x)\phi_{2,\mu}(y) - \phi_1(y)\phi_1(x)\phi_{2,\nu}(y)\phi_{2,\mu}(x).\end{aligned} \quad (A3)$$

(The latter equality holds because independent free scalar fields commute.) Similarly, expanding the other terms and noting that the only state connected to the vacuum by both $\phi_1\phi_1$ and $\phi_2\phi_2$ is the vacuum, we obtain Eq. (2.1).

It is amusing that in the spinor case the analogous fields are ψ_1 and ψ_2, defined by

$$\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} = \frac{1}{(2)^{\frac{1}{2}}}\begin{bmatrix}1 & \gamma^2\\ \gamma^2 & 1\end{bmatrix}\begin{pmatrix}\psi\\ \psi^+\end{pmatrix}, \quad (A4)$$

where $\gamma^2 = \beta\alpha_2$. These are independent in the sense that

$$(0|\ \psi_1^a(x)\psi_2^b(y)\ |0) = 0 = \langle 0|\ \psi_2^a(x)\psi_1^b(y)\ |0\rangle.$$

In terms of them, the current $j_\mu = e:\bar{\psi}\gamma_\mu\psi:$ may be written

$$j_\mu = e\psi_2\gamma^0\gamma_\mu\psi_1. \quad (A5)$$

With these, one may proceed to calculate the current commutators just as in the scalar case. Equation (2.2) is the answer.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

Generalized Triangular Ising Lattice

R. G. J. MILLS
Department of Physics and Astronomy, The University of Rochester, Rochester, New York
AND
C. A. HURST
The University of Adelaide, Adelaide, South Australia

(Received 20 February 1969)

The partition function and critical equations for the generalized triangular Ising lattice are determined in terms of weight factors associated with the decorating lattice. As an example, a lattice which incorporates the Kagome, hexagonal, triangular, and rectangular lattices is solved by the method developed.

1. INTRODUCTION

Since the calculation of the partition function of the rectangular Ising lattice by Onsager,[1] many other two-dimensional Ising lattices have been solved.[2] Use of the combinatorial method involves the counting of closed polygons with a weight factor which can be drawn on the lattice. Hurst and Green[3] have shown how a Pfaffian can be used directly to perform the counting. Kasteleyn[4] has shown how the Ising problem is related to the dimer problem and that all planar Ising lattices can be solved in principle by using a Pfaffian whose elements are found by correct orientation of the bonds of the related dimer lattice. Fisher[5] showed that Kasteleyn's method can be simplified by taking as the dimer lattice one in which there is a one-to-one correspondence between polygons on the Ising lattice and dimer configurations on the dimer lattice. He gives a scheme by which any planar Ising lattice can be solved in principle. The purpose of this paper is to give a method by which the solution can be carried out in detail. We use the method of Hurst since the weight factors of the lattice-point polynomial and the consistency conditions on the weight factors defined in I give a simple method of expanding the determinant in the partition function.

In I it was shown how all the two-dimensional lattices could be classified such that to each class there corresponds a single basic lattice point or cell. For a planar lattice these cells must be connected by parallel bonds to form a generalized triangular lattice as shown in Fig. 1 of I. Thus the class is given by the number of horizontal, diagonal, and vertical bonds adjacent to the cell. Particular members of a class differ in the internal structure of the cell, which must be the same for all cells since we consider only periodic lattices. Bonds connecting points inside a cell will be called internal bonds, while bonds between cells will be called external bonds.

We label the cells in the helical ordering with i running from 1 to N, and we label the bonds connected to the ith cell from 1 to $2g$. We adopt the convention that bonds numbered from 1 to g connect the ith cell to cells with lower index (except cells in the first row) and those numbered from $g + 1$ to $2g$ are connected to lattice points with higher index. So, a bond connecting cells i and i' will, therefore, have two labels, one fixing its ordering with respect to i and the other its ordering with respect to i'. We call these two labels *associated labels*. In the language of Ref. 2, these labels refer to terminals rather than bonds and the latter are specified by the labels of the terminals which form their end points.

In the generalized triangular lattice considered here, the bonds can be grouped into six classes according to the following convention:

(1) Terminals labeled from 1 to h which connect horizontal bonds entering the ith cell from the $(i - 1)$th cell.

(2) Terminals labeled from $h + 1$ to $h + d$ which connect diagonal bonds entering the ith cell from the $(i - m - 1)$th cell.

(3) Terminals labeled from $h + d + 1$ to $h + d + v = g$ which connect vertical bonds entering the ith cell from the $(i - m)$th cell.

(4) Terminals labeled from $g + 1$ to $g + h$ which connect horizontal bonds leaving the ith cell for the $(i + i)$th cell.

(5) Terminals labeled from $g + h + 1$ to $g + h + d$ which connect diagonal bonds leaving the ith cell for the $(i + m + 1)$th cell.

(6) Terminals labeled from $g + h + d + 1$ to $2g$ which connect vertical bonds leaving the ith cell for the $(i + m)$th cell.

Here, m denotes the number of cells in a row. The bonds can be labeled from 1 to g according to the

[1] L. Onsager, Phys. Rev. **65**, 117 (1944).
[2] See, for example, C. A. Hurst, J. Chem. Phys. **38**, 2558 (1963), hereafter referred to as I.
[3] C. A. Hurst and H. S. Green, J. Chem. Phys. **33**, 1059 (1960).
[4] P. W. Kasteleyn, J. Math. Phys. **4**, 287 (1963).
[5] M. E. Fisher, J. Math. Phys. **7**, 1776 (1966).

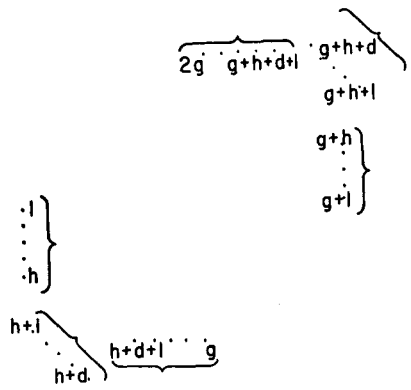

FIG. 1. Ordering of terminals at a vertex.

terminal by which they enter the cell. The ordering of the terminals is as shown in Fig. 1. Because of the structure of the lattice, to every bond numbered j connecting $i-1$ to i and ordered with respect to i, there will be an *associated label* j' for the same bond ordered with respect to $i-1$. It can be seen from Fig. 1 that

$$\begin{aligned} j' &= g+h-j+1, && \text{if } 1 \le j \le h, \\ j' &= g+2h+d-j+1, && \text{if } h < j \le h+d, \\ j' &= g+2h+2d+v-j+1, && \text{if } h+d < j \le g. \end{aligned}$$

We see that j and j' are the terminals by which the jth bond enters and leaves the ith cell, respectively. We denote by $c_{rst\cdots w}$ the lattice-point weight factor corresponding to the set of diagrams drawn on the cell which are consistent with the terminals labeled $rst\cdots w$ of the cell being connected by external bonds and with an even number of bonds being connected to every spin of the cell. The weight factors have an even number of indices, ranging from zero, when there are no external bonds connected, to $2g$ when all possible external bonds are connected. These weight factors will be functions of the internal parameters $z_1, z_2, \cdots, z_r$ which depend on the internal bond energies J and the temperature T. For subsequent work, it is useful to define new weight factors

$$C_{ab\cdots h} = c_{rs\cdots w}, \tag{1.1}$$

where $\{ab\cdots h\}$ is that subset of the $2g$ indices which is complementary (in the set-theoretical sense) to the subset $\{rs\cdots w\}$.

In I it was shown how a lattice-point polynomial involving the weight factors $c_{rs\cdots w}$ and creation and annihilation operators could be associated with each cell. For planar lattices, the weight factors satisfy the consistency conditions

$$(C_{ab\cdots d})^{\nu-\mu-1} C_{ab\cdots h} = |C_{ab\cdots dij}|, \tag{1.2}$$

where $C_{ab\cdots h}$ is a weight factor with 2ν indices, $C_{ab\cdots d}$ is a weight factor with 2μ indices which are a subset of $\{ab\cdots h\}$ and $C_{ab\cdots dij}$ is a weight factor with $2(\mu+1)$ indices. The indices are restricted by the inequalities

$$a < b < \cdots < d < \cdots < e < \cdots < g < \cdots < h$$

and the Pfaffian on the right-hand side of (1.2) is of order $2(\nu-\mu)$. The polynomial is factorized into a product of linear fermion operators and the expectation value of a product of polynomials gives a Pfaffian which is evaluated as the square root of an antisymmetric determinant. The combinatorial equivalence of this method and the counting of closed polygons gives the partition function for the lattice considered.

For the generalized triangular lattice, the partition function is

$$N^{-1}\log Z = \log(2\cosh H_1 \cosh H_2 \cdots \cosh H_g) + (2\pi)^{-2}\int_0^{2\pi} d\theta \int_0^{2\pi} d\phi \log|\Delta(e^{i\theta}, e^{i\phi})|, \tag{1.3}$$

where

$$\Delta = C_0^2 \Delta' \tag{1.4}$$

is a matrix of order $2g$ with elements

$$\Delta'_{ij} = A_{ij} + \Lambda_{ij}, \tag{1.5}$$

$$A_{ij} = -A_{ji} = (-1)^{i+j-1} C'_{ij}, \tag{1.6}$$

and, for $i < j$,

$$\begin{aligned} \Lambda_{ij} &= -x_i e^{i\theta}\delta_{ji'}, && \text{for } 1 \le i \le h, \\ &= -x_i e^{i(\theta+\phi)}\delta_{ji'}, && \text{for } h < i \le h+d, \\ &= -x_i e^{i\phi}\delta_{ji'}, && \text{for } h+d < i \le g, \\ &= \Lambda^*_{ji}. \end{aligned} \tag{1.7}$$

Here,

$$C'_{ij} = C_{ij}/C_0 \tag{1.8}$$

and

$$\begin{aligned} \delta_{ji'} &= 1, \quad \text{if } j = i', \\ &= 0, \quad \text{otherwise}, \end{aligned}$$

where i' is the associated label of i.

In Sec. 2 we show how the determinant can be expanded in a compact expression which is quadratic in the weight factors. The partition function for each member of a class of lattices can then be found by calculating the weight factors. In Sec. 3 we consider the critical equations for the singularities in the partition function corresponding to physical singularities found in I and show how the Pfaffian method gives conditions on the weight factors. In Sec. 4 we solve a new lattice with two spins to a cell by the method developed. Finally, we conclude with a discussion of the method and comparison with other methods.

2. EXPANSION OF THE DETERMINANT

The determinant Δ can be written so that all the C'_{ij} have a positive sign by using $C_{ij} = -C_{ji}$, for $i < j$, and then multiplying the odd columns and even rows by -1 giving

$$\Delta = C_0^2 |C'_{ij} + (-1)^P x_i e^{ix} \delta_{ji'}|, \tag{2.1}$$

where

$$P = i + i', \quad \text{if } i < j,$$
$$= i + i' + 1, \quad \text{if } i > j,$$

and x is the appropriate angle given in Eq. (1.7).

The special structure of Δ enables it to be reduced to a compact expression depending on the coefficients $C_{ab\cdots h}$ and which is quadratic in these coefficients. From Eq. (1.5) it can be seen that Δ is the determinant of a matrix which is the sum of a real antisymmetric and an anti-Hermitian matrix. The latter matrix is such that all the nonzero elements above the main diagonal have row indices which range from 1 to g and column indices which range from $g + 1$ to $2g$. Hence the row and column indices of Λ_{ij} for $i < j$ form disjunct sets. Furthermore, because Δ is an even-order anti-Hermitian determinant, it is a real number. Hence the phase factors can only contribute to Δ in the form $\cos(p\theta + q\phi)$, where $-(h + d) \leq p \leq (h + d)$ and $-(d + v) \leq q \leq d + v$. For the moment, we shall put $(\theta + \Phi) = \chi$ and consider the coefficient of

$$\exp i\{(p_1 - q_1)\theta + (p_2 - q_2)\chi + (p_3 - q_3)\phi\},$$

with

$$0 \leq p_i, q_i \leq h, \quad 0 \leq p_2, q_2 \leq d, \quad 0 \leq p_3, q_3 \leq v.$$

The integers p_i define the number of phase factors of the various types which come from the portion of Δ above the diagonal whilst the numbers q_i give the number of phase factors which come from below the diagonal.

In order to evaluate the determinant Δ, we consider the coefficient of $x_1^{\alpha_1}, x_2^{\alpha_2}, \cdots, x_g^{\alpha_g}$, where $0 < \alpha < 2$. Let $\lambda_{21}, \lambda_{22}, \cdots, \lambda_{2n}$ denote the indices λ for which $\alpha_\lambda = 2$, let $\lambda^+, \lambda^+, \cdots, \lambda_p^+$ denote the indices λ corresponding to $\alpha_\lambda = 1$ and a positive sign in the phase factor, let $\lambda_{11}^-, \lambda_{12}^-, \cdots, \lambda_{1q}^-$ denote the indices corresponding to $\alpha_\lambda = 1$ and a negative sign in the phase factor, and let $\lambda_{01}, \lambda_{02}, \lambda_{0m}$ denote the indices corresponding to $\alpha = 0$. All the indices satisfy the inequality $1 \leq \lambda_2, \lambda_1^+, \lambda_1^-, \lambda_0 \leq g$ and associated with the indices are the associated indices $\lambda'_2, \lambda_1^{+\prime}, \lambda_1^{-\prime}$, and λ'_0 which satisfy $g \leq \lambda'_2, \lambda_1^{+\prime}, \lambda_1^{-\prime}, \lambda'_0 \leq 2g$. Let n_1, p_1, q_1, and m_1 denote the number of bonds selected in the horizontal direction such that $1 \leq n_1, p_1, q_1, m_1 \leq h$; n_2, p_2, q_2, and m_2 denote the number of bonds selected in the diagonal direction such that $1 \leq n_2, p_2, q_2, m_2 \leq d$ and n_3, p_3, q_3, and m_3 denote the number selected in the vertical direction with $1 \leq n_3, p_3, q_3, m_3 \leq v$. The required coefficient can be uniquely denoted by

$$D(\lambda_2; \lambda_1^+; \lambda_1^-; \lambda_0)$$
$$\equiv D(\lambda_{21} \cdots \lambda_{2n}; \lambda_{11}^+ \cdots \lambda_{1p}^+; \lambda_{11}^- \cdots \lambda_{1q}^-; \lambda_{01} \cdots \lambda_{0m}) \tag{2.2}$$

and the determinant can be expanded as

$$\Delta = \sum_p \sum_q \sum_{m+n=g-p-q} \exp i\{(p_1 - q_1)\theta + (p_2 - q_2)\chi + (p_3 - q_3)\phi\} \times x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_g^{\alpha_g} D(\lambda_2; \lambda_1^+; \lambda_1^-; \lambda_0). \tag{2.3}$$

The coefficient (2.2) can be geometrically represented by a cluster of $2g$ terminals ordered from 1 to $2g$ as described at the beginning of this section. The labels denoted by λ_2 and λ'_2 represent the bonds from one of the cells $i - 1$, $i - m - 1$, $i - m$ to i and from i to one of the cells $i + 1$, $i + m + 1$, $i + m$, respectively. The labels denoted by $\lambda_1^{+\prime}$ represent bonds from i to one of $i + 1$, $i + m + 1$, $i + m$, while the labels λ_1^- represent bonds from $i - m$, $i - m - 1$, $i - 1$ to i. The labels λ_0 and λ'_0 represent no bonds to those terminals of the cell.

The coefficient $D(\lambda_2; \lambda_1^+; \lambda_1^-; \lambda_0)$ is a determinant of order $2g - 2n - p - g = 2m + p + q$ which is obtained from Δ by striking out the rows and columns containing the factors x_λ appearing in the product $x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_g^{\alpha_g}$. The striking out of the $2n$ rows and columns for which $\alpha_\lambda = 2$ eliminates the indices λ_{2n}, λ'_{2n} from the determinant. The striking out of the p rows and columns above the diagonal and the q rows and columns below the diagonal for which $\alpha_\lambda = 1$ removes one set of the indices $\lambda_1^+, \lambda_1^+, \lambda_1^-, \lambda_1^{-\prime}$ from the determinant. As all indices occur twice in the original determinant, these indices are still present but only occur once. The rows of the determinant are labeled by $\lambda_1^-, \lambda_0, \lambda'_0, \lambda_1^{+\prime}$ and the columns by $\lambda_1^+, \lambda_0, \lambda'_0, \lambda_1^{-\prime}$ and, from the manner of construction, the sets $\{\lambda_1^-, \lambda_1^{+\prime}\}$ and $\{\lambda_1^+, \lambda_1^{-\prime}\}$ are disjunct. The rows and columns labeled by λ_0, λ'_0 for which $\alpha_\lambda = 0$ form a real antisymmetric submatrix of $D(\lambda_2; \lambda_1^+; \lambda_1^-; \lambda_0)$. The sign factors arising from these and subsequent operations are discussed in Appendix B. We ignore the sign factors in the following discussion, giving the total sign for D in Eq. (2.8).

The determinant can be reduced in order by removing the antisymmetric submatrix by the method shown in Appendix A. At each step of the reduction, the elements of the determinant become 4×4 Pfaffians, which can be replaced by a higher-order weight factor by using the consistency conditions (1.2). On the first

reduction, D is reduced to a determinant of order $2m + p + g - 2$:

$$D = C_0^2 |C'_{ij}| = C_0^2 (C'_{\lambda_{01}\lambda_{01}'})^{4-2m-p-q} |D'|, \quad (2.4)$$

where

$$D'_{ij} = \begin{vmatrix} C'_{\lambda_{01}\lambda_{01}'} & C'_{\lambda_{01}j} & C'_{i\lambda_{01}} \\ & C'_{\lambda_{01}'j} & C'_{i\lambda_{01}'} \\ & & C'_{ij} \end{vmatrix}.$$

Using the consistency conditions Eq. (1.2) with $\nu = 2$ and $\mu = 0$ gives

$$D_{ij} = C'_0 C'_{\lambda_{01}\lambda_{01}'ij} = C'_{\lambda_{01}\lambda_{01}'ij}.$$

In Appendix A we show that the new determinant still contains a real antisymmetric submatrix so the process can be continued until all the antisymmetric part which is labeled by λ_0 is removed. The result is a determinant of order $p + q$:

$$D = C_{\lambda_0}^{2-p-q} |C_{\lambda_0 ij}|, \quad (2.5)$$

where

$$\boldsymbol{\lambda}_0 \text{ denotes } \lambda_{01}\lambda'_{01}\lambda_{02}\lambda'_{02} \cdots \lambda_{0m}\lambda'_{0m}$$

and ij are the indices of the sets $\{\lambda_1^-, \lambda_1^{+\prime}\}$ and $\{\lambda_1^+, \lambda_1^{-\prime}\}$, respectively.

The terms of the determinant of order $(p + q)$ form a subset of the terms of the Pfaffian $|C_{\lambda_0 rs}|$ of order $2(p + q)$, where the indices r, s range over the combined set

$$\{\lambda_1^+, \lambda_1^-, \lambda_1^{+\prime}, \lambda_1^{-\prime}\}.$$

Thus, the determinant in Eq. (2.5) is equal to the Pfaffian if the terms in the Pfaffian with both indices from the set $\{\lambda_1^+, \lambda_1^{-\prime}\}$ are made zero. Thus the determinant $|C_{\lambda_0 ij}|$ can be written, apart from a sign factor discussed in Appendix B, as

$$|C_{\lambda_0 ij}| = \{|C_{\lambda_0 rs}| - \sum |C_{\lambda_0 ab}|\,|C_{\lambda_0 ru}| + \sum\sum |C_{\lambda_0 ab}|\,|C_{\lambda_0 tu}|\}. \quad (2.6)$$

The Pfaffians in Eq. (2.6) are of order $2(p + q)$, 2 and $2(p + q - 1)$, 4 and $2(p + q - 2)$, respectively. The summations are over the 0, 1, 2, $\cdots$ pairs of indices a, b which can be selected from the set $\{\lambda_1^+, \lambda_1^{-\prime}\}$. By using the consistency conditions (1.2), these Pfaffians can be replaced by weight factors

$$C_{\lambda_0}^{p+q-1} C_{\lambda_0 \lambda_1^+ \lambda_1^- \lambda_1^{+\prime} \lambda_1^{-\prime}},\ (-1)^R C_{\lambda_0 ab} C_{\lambda_0}^{p+q-2} C_{\lambda_0 tu \cdots w}$$

and

$$(-1)^R C_{\lambda_0 abcd},\ (-1)^R C_{\lambda_0} C_{\lambda_0 abcd},\ C_{\lambda_0}^{p+q-3} C_{\lambda_0 v \cdots w}.$$

Here $ab \cdots d$ are indices from the set $\{\lambda_1^+, \lambda_1^{-\prime}\}$ and $ru \cdots w$ are the complement of $ab \cdots d$ in the set $\{\lambda_1^+, \lambda_1^-, \lambda_1^{+\prime}, \lambda_1^{-\prime}\}$. The sign factor R is the parity of the permutation required to arrange $\{ab \cdots d\}$ and its complement in numerical order.

Finally, we combine Eqs. (2.5) and (2.6) and use Eq. (1.1) to express the determinant in terms of the original weight factors giving

$$D = (-1)^S \{C_{\lambda_2 \lambda_1^+ \lambda_1^- \lambda_1^{+\prime} \lambda_1^{-\prime} C_{\lambda_2}} - (-)^R \sum C_{\lambda_2 tu \cdots w c \lambda_2 ab} + (-)^R \sum\sum C_{\lambda_2 v \cdots w} C_{\lambda_2 abcd \cdots}\}. \quad (2.7)$$

Here, all the indices are in numerical order and S is the sign factor discussed in Appendix B given by

$$S = \tfrac{1}{2}(p_1 + q_1)(2n + p_1 + q_1) + \tfrac{1}{2}(p_1 + q_2)(2n_2 + p_2 + q_2) + \tfrac{1}{2}(p_3 + q_3)(2n_3 + p_3 + q_3) + \tfrac{1}{2}(p + q)(2n + p + q). \quad (2.8)$$

The structure of (2.7) can best be seen by giving a geometric representation. Each term on the right-hand side can be represented by two clusters of $2g$ terminals labeled as in Fig. 1. The first term corresponds to all the terminals ($\boldsymbol{\lambda}_2$, $\boldsymbol{\lambda}_1^+$, $\boldsymbol{\lambda}_1'$, and $\boldsymbol{\lambda}_2$) being connected by external bonds in the first and second cluster, respectively. The second term corresponds to the terminals ($\boldsymbol{\lambda}_2 tu \cdots w$ and $\boldsymbol{\lambda}_2 a, b$) being connected by external bonds in the first and second cluster, respectively. The second term differs from the first in that the terminals a, b are connected by external bonds in the second cluster rather than in the first. The third term corresponds to the terminals ($\boldsymbol{\lambda}_2 v \cdots w$ and $\boldsymbol{\lambda}_2 abcd$) being connected by external bonds in the first and second cluster, respectively. In no cluster are the terminals labeled by $\boldsymbol{\lambda}_0$ connected.

So, finally, we see that the determinant Δ can be expanded as

$$\begin{aligned} \Delta = & \sum_{m+n=g} D(\lambda_{2n}; 0; 0; \lambda_{0m}) x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_g^{\alpha_g} \\ & + \sum_p \sum_{q=p} \sum_{m+n+p+q=g} D(\lambda_{2n}; \lambda_{1p}^+; \lambda_{1q}^-; \lambda_{0m}) x_1^{\alpha_1} \cdots x_g^{\alpha_g} \\ & \times \exp i\{(p_1 + p_2 - q_1 - q_2)\theta \\ & + (p_2 + p_3 - q_2 - q_3)\phi\} \\ & + 2 \sum_p \sum_{q<p} \sum_{m+n+p+q=g} D x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_g^{\alpha_g} \\ & \times \cos\{(p_1 + p_2 - q_1 - q_2)\theta \\ & + (p_2 + p_3 - q_2 - q_3)\phi\}, \end{aligned} \quad (2.9)$$

where D is given by (2.7). The first summation is over the 2^g ways in which the bonds $x_1 x_2 \cdots x_g$ can be separated into two groups $\lambda_{21}, \cdots, \lambda_{2n}$ with $\alpha_\lambda = 2$ and $\lambda_{01}, \cdots, \lambda_{0m}$ with $\alpha_\lambda = 0$. In the second summation, the p bonds $\lambda_{11}^+, \cdots, \lambda_{1p}^+$ can be selected from above the diagonal in $\binom{g}{p}$ ways, the bonds

$\lambda_{11}^-, \cdots, \lambda_{1p}^-$ can be selected from below the diagonal in $\binom{g}{p}$ ways, and the remaining $g-2p$ bonds can be separated into the group $\lambda_{21}, \cdots, \lambda_{2n}$ and $\lambda_{01}, \cdots, \lambda_{0m}$ in 2^{g-2p} ways. For a given selection $D(\lambda_{2n}; \lambda_{1p}^+; \lambda_{1p}^-; \lambda_{0m})$, there is another selection $D(\lambda_{2n}; \lambda_{1p}^-; \lambda_{1p}^+; \lambda_{0m})$ and these terms combine together to give a cosine term. The third summation is over the $\binom{g}{p}\binom{g-p}{q}2^{g-p-q}$ ways in which $D(\lambda_{2n}; \lambda_{1p}^+; \lambda_{1q}^-; \lambda_{0m})$ can be selected.

3. CRITICAL BEHAVIOR

In I it was shown that critical equations for a general lattice could be derived by satisfying the conditions

$$\begin{aligned} &\Delta(e^{i\theta}, e^{i\phi}) = 0, \\ &\frac{\partial}{\partial\theta}\Delta(e^{i\theta}, e^{i\phi}) = 0, \\ &\frac{\partial}{\partial\phi}\Delta(e^{i\theta}, e^{i\phi}) = 0. \end{aligned} \tag{3.1}$$

Since Δ is given by (2.9) as a sum of cosine terms in multiple angles of θ and ϕ, the second and third equations will be satisfied when $\theta, \phi = 0, \pi$. Other solutions may exist, but the fact that they were nonphysical in the simpler model considered in I suggests that they will be nonphysical in our model also. Therefore, we confine our attention to the solution of the first equation of (3.1) for $\theta, \phi = 0, \pi$.

When $\theta = \phi = 0$ are substituted in (2.1), the determinant becomes

$$\Delta = C_0^2 \,|C'_{ij} + (-1)^P x_i \delta_{ji'}|, \tag{3.2}$$

since $e^{i\chi} = e^{-i\chi} = 1$ when $\chi = 0$. Since $P = i + i'$ if $i < j$ and $i + i' + 1$ if $i > j$, the determinant is antisymmetric and hence it can be written as a Pfaffian

$$\Delta^{\frac{1}{2}} = C_0 \,|C'_{ij} + (-1)^{i+i'} x_i \delta_{ji'}|. \tag{3.3}$$

This Pfaffian can be expanded as a sum of terms

$$x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_g^{\alpha_g},$$

where $\alpha_i = 0, 1$. Let $\lambda_i, \cdots, \lambda_n$ denote the indices for which $\alpha_\lambda = 1$, where n_1 are chosen such that $1 \le \lambda \le h$, n_2 are chosen such that $h < \lambda \le h + d$, and n_3 are chosen such that $h + d < \lambda \le g$. Let the coefficient of $x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_g^{\alpha_g}$ be denoted by $P(\lambda_1, \cdots, \lambda_n)$, where $n = n_1 + n_2 + n_3$. Then the required expansion is

$$\Delta^{\frac{1}{2}} = \sum_{\alpha_\lambda = 0,1} P(\lambda_1, \cdots, \lambda_n) x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_g^{\alpha_g}. \tag{3.4}$$

The coefficient $P(\lambda_1, \cdots, \lambda_n)$ is the cofactor of the Pfaffian when the rows and columns containing the indices $\lambda_1, \cdots, \lambda_n$ and their associated indices $\lambda'_1, \cdots, \lambda'_n$ are removed from the Pfaffian and the bond weights x_λ for which $\alpha_\lambda = 0$ are made zero. The cofactor itself is a Pfaffian of order $2(g-n)$ with elements C'_{ab} whose rows and columns are labeled by the set of indices $\{a, b, \cdots, h\}$ which is the complement of the set $\{\lambda\}$ in the set $\{1, 2, \cdots, 2g\}$. By the consistency conditions (1.1), the Pfaffian $|C'_{ab}|$ can be replaced by $C_0^{-1} C_{ab\cdots h}$ so that

$$\begin{aligned} P(\lambda_1, \cdots, \lambda_n) &= (-)^T C_0 \,|C'_{ab}| \\ &= (-)^T C_{ab\cdots h}. \end{aligned}$$

Here, T is the sign factor to remove the bond weights for which $\alpha_\lambda = 1$ which is discussed in Appendix C. From Eq. (1.1), which relates the weight factors, we have

$$C_{ab\cdots h} = c_\lambda,$$

so that

$$\Delta^{\frac{1}{2}} = \sum_{\alpha_\lambda = 0,1} (-)^T c_\lambda x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_g^{\alpha_g}. \tag{3.5}$$

The form of the critical equations for the other values of θ and ϕ can be found from the case $\theta = \phi = 0$ by transformation of the bond weights. From the structure of the determinant in Eq. (1.7), it can be seen that the change $\theta = 0$ to $\theta = \pi$ has the effect of changing x_λ to $-x_\lambda$ if $1 \le \lambda \le h + d$. Similarly, the change $\phi = 0$ to $\phi = \pi$ has the effect of changing x_λ to $-x_\lambda$ if $h < \lambda \le g$. In Appendix C we show that those changes can be incorporated in the sign factor T given by

$$\begin{aligned} T &= n + n_1 n_2 + n_2 n_3 + n_1 n_3, \quad \text{if} \quad \theta = \phi = 0, \\ &= n - n_1 n_2 + n_2 n_3 - n_1 n_3, \quad \text{if} \quad \theta = 0 \quad \phi = \pi, \\ &= n + n_1 n_2 - n_2 n_3 - n_1 n_3, \quad \text{if} \quad \theta = \pi \quad \phi = 0, \\ &= n - n_1 n_2 - n_2 n_3 + n_1 n_3, \quad \text{if} \quad \theta = \pi \quad \phi = \pi. \end{aligned} \tag{3.6}$$

4. A LATTICE WITH TWO SPINS TO A CELL

In this section we show how the partition function and the critical equations for the lattice shown in Fig. 2 can be written down by using the results of Secs. 2 and 3. The power of the method lies in the fact that the calculations are straightforward, although tedious. In general, only the method of calculation and the results will be given.

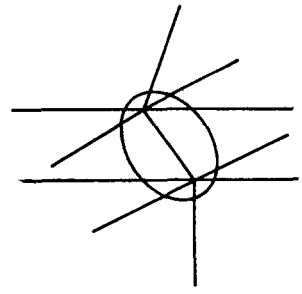

FIG. 2. Example of a lattice with 2 spins to a cell.

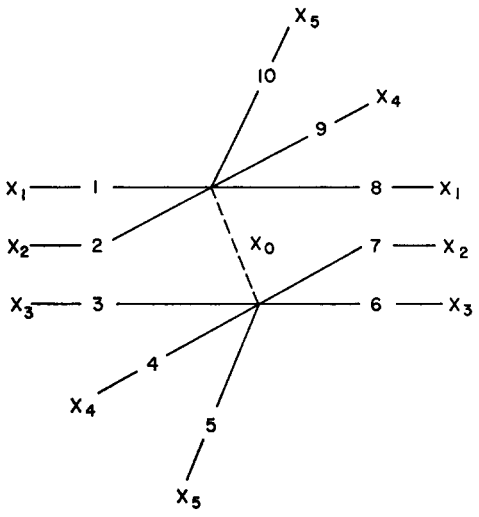

FIG. 3. Cell structure for lattice of Fig. 2.

In order to represent the lattice of Fig. 2, each cell is given ten terminals and these terminals are connected to the spins as shown in Fig. 3. There is a weight factor for each way in which an even number of external bonds are connected to a cell. This weight factor describes the ways in which the internal bonds can be arranged so that there is an even number of external and internal bonds to each spin. The weight factors can easily be found from Fig. 3 to be

$$c_{ab\cdots h} = x_0, \text{ if the indices } ab\cdots h \text{ contain an odd number of the set } \{3, 4, 5, 6, 7\},$$
$$= 1, \text{ otherwise.} \tag{4.1}$$

Thus $c_0 = c_{12} = 1$ and $c_{13} = x_0$, etc. The lattice has five external bonds and one internal bound; the cell has ten terminals and $h = 3$ and $d = v = 1$. There is one terminal for each bond entering or leaving the lattice and the external bonds are labeled $x_1, x_2, \cdots, x_5$ corresponding to the terminal of the set $\{1, 2, \cdots, 5\}$ to which they are connected. The bonds x_1, x_2, and x_3 represent horizontal bonds between cells, x_4 represents a diagonal bond between cells, and x_5 represents a vertical bond between cells. It can be seen that the pairs of terminals (1, 8), (2,7), (3, 6), (4, 9), and (5, 10) are the associated terminals defined in Sec. 1. The coefficient of $x_1x_2^2x_4x_5 \cos(\theta - \chi + \phi)$ is $D(2; 1, 5; 4; 3)$, with $n_1 = p_1 = p_3 = q_2 = m_1 = 1$ and $n_2 = n_3 = p_2 = q_1 = m_2 = m_3 = 0$. From Eq. (2.8), the sign factor is even. Since $1' = 8$, $2' = 7$, $3' = 6$, $4' = 9$, and $5' = 10$, Eq. (2.7) gives

$$D(2; 1, 5; 4; 3) = \{c_{1245789\,10}c_{27} + c_{2478910}c_{1257} + c_{2457810}c_{1279} + c_{1247810}c_{2579}\}.$$

The sign factors R of the second, third, and fourth terms are positive because (4, 8, 9, 10, 1, 5), (4, 5, 8, 10, 1, 9), and (1, 4, 8, 10, 5, 9) are odd permutations of (1, 4, 5, 8, 9, 10). The weight factors are shown in Fig. 4. It can be seen that each spin has an even number of bonds (external and internal) connected to it. We see that $D(2; 15; 4; 3) = 2(1 + x_0^2)$. It can be seen that x_2 is always connected in both clusters, whereas x_3 is never connected. The result of combining each cluster is a bond weight $x_1x_2^2x_4x_5$. Each term in the expansion of the 10×10 determinant can be evaluated as above.

We notice that if x_1 and x_2 are selected from above the diagonal and not selected from below the diagonal, the 1st and 2nd columns of the determinant, which remain when the 1st and 2nd rows and the 7th and 8th columns have been removed, are equal. No matter what further selections are made, the 1st and 2nd columns will remain equal so that $D(\lambda_2; 1, 2, \lambda_1^+; \lambda_1^-; \lambda_0)$ will be zero for all λ_2, λ_1^+, λ_1^-, λ_0. Similarly, we find that

$$D(\lambda_2; 1, 2, \lambda_1^+; \lambda_1^-; \lambda_0) = D(\lambda_2; \lambda_1^+; 1, 2, \lambda_1^-; \lambda_0) = 0,$$
$$D(\lambda_2; 2, 3, \lambda_1^+; \lambda_1^-; \lambda_0) = D(\lambda_2; \lambda_1^+; 2, 3, \lambda_1^-; \lambda_0) = 0,$$
$$D(\lambda_2; 3, 4, \lambda_1^+; \lambda_1^-; \lambda_0) = D(\lambda_2; \lambda_1^+; 3, 4, \lambda_1^-; \lambda_0) = 0,$$
$$D(\lambda_2; 4, 5, \lambda_1^+; \lambda_1^-; \lambda_0) = D(\lambda_2; \lambda_1^+; 4, 5, \lambda_1^-; \lambda_0) = 0,$$
$$D(\lambda_2; 1, 4, \lambda_1^+; \lambda_1^-; \lambda_0) = D(\lambda_2; \lambda_1^+; 1, 4, \lambda_1^-; \lambda_0) = 0,$$
$$D(\lambda_2; 5, \lambda_1^+; 3, \lambda_1^-; \lambda_0) = D(\lambda_2; 3, \lambda_1^+; 5, \lambda_1^-; \lambda_0) = 0,$$
$$D(\lambda_2; 5, \lambda_1^+; 1, \lambda_1^-; \lambda_0) = D(\lambda_2; 1, \lambda_1^+; 5, \lambda_1^-; \lambda_0) = 0.$$

These relationships reduce the number of terms in the expansion of the determinant considerably.

The first summation in Eq. (2.9) produces the 2^5 terms which arise on selecting the 5 external bonds in two groups $\lambda_{21}, \cdots, \lambda_{2n}$ and $\lambda_{01}, \cdots, \lambda_{0m}$. The coefficient of $x_1^{\alpha_1}x_2^{\alpha_2}\cdots x_5^{\alpha_5}$ where $\alpha_\lambda = 0$ or 2 is found from Eq. (2.7) to be

$$D(\lambda_{2n}; 0; 0; \lambda_{0m}) = c_{\lambda_2}^2. \tag{4.2}$$

From Fig. 3 it is easily found that c_{λ_2} is x_0 or 1 if $\boldsymbol{\lambda}_2$ contains an odd or even number of elements from the set $\{3, 4, 5, 6, 7\}$, respectively.

The second summation corresponds to selecting p bonds from above the diagonal and the same number $q = p$ of different bonds from below the diagonal. Since there are only five bonds altogether, only the two cases $p = q = 1$ and $p = q = 2$ need be considered for this lattice. Furthermore, we have $D(\lambda_2; \lambda_1^+; \lambda_1^-; \lambda_0) = D(\lambda_2; \lambda_1^-; \lambda_1^+; \lambda_0)$, since by Eq. (2.7) the coefficient depends on the set $\{\lambda_{11}^+, \cdots, \lambda_{1p}^+, \lambda_{11}^-, \cdots, \lambda_{1q}^-\}$ for a given $\boldsymbol{\lambda}_2$. Thus the coefficients combine together to give cosine terms. For selections

such that $p_i = q_i$, for $i = 1, 2, 3$, the angle

$$\{(p_1 + p_2 - q_1 - q_2)\theta + (p_2 + p_3 - q_2 - q_3)\phi\}$$

is zero.

The third summation corresponds to selecting p bonds from above the diagonal and $q = p$ bonds from below the diagonal. For selections such that $p_1 + p_2 = q_1 + q_2$ and $p_2 + p_3 = q_2 + q_3$, the above angle is zero and these contribute to the absolute term.

The coefficients $D(\lambda_2; \lambda_1^+; \lambda_1^-; \lambda_0)$ are evaluated by the method illustrated in the example given above and, after some tedious algebraic manipulations, the determinant is found to be

$$\begin{aligned}
\Delta = {} & \tfrac{1}{2}(1 + x_1^2)(1 + x_3^2) \\
& \times [(1 + x_0^2)(1 + x_2^2)(1 + x_4^2)(1 + x_5^2) \\
& + (1 - x_0^2)(1 - x_2^2)(1 - x_4^2)(1 - x_5^2)] \\
& + 4x_0x_2(1 + x_4^2)(1 + x_5^2) \\
& \times [x_1(1 + x_3^2) + x_3(1 + x_1^2)] \\
& + 2x_1x_3(1 + x_0^2)(1 + x_2^2)(1 + x_4^2)(1 + x_5^2) \\
& + 4x_4x_5(1 + x_0^2)(1 + x_2^2) \\
& \times [x_1(1 + x_3^2) + x_3(1 + x_1^2)] \\
& + 8x_0x_2x_4x_5(1 + x_1^2)(1 + x_3^2) + 32x_0x_1x_2x_3x_4x_5 \\
& - [x_1(1 + x_3^2) + x_3(1 + x_1^2)] \\
& \times [(1 + x_0^2)(1 + x_2^2)(1 - x_4^2)(1 - x_5^2) \\
& + (1 - x_0^2)(1 - x_2^2)(1 + x_4^2)(1 + x_5^2)] \cos\theta \\
& - 2(1 + x_1^2)(1 + x_3^2)[x_0x_2(1 - x_4^2)(1 - x_5^2) \\
& + x_4x_5(1 - x_0^2)(1 - x_2^2)] \cos\theta \\
& - 8x_1x_3[x_0x_2(1 - x_4^2) + (1 - x_5^2) \\
& + x_4x_5(1 - x_0^2)(1 - x_2^2)] \cos\theta \\
& - 2(1 - x_1^2)(1 - x_3^2)[x_0x_4(1 - x_2^2)(1 - x_5^2) \\
& + x_2x_5(1 - x_0^2)(1 - x_4^2)] \cos(\theta + \phi) \\
& - 2x_0x_5(1 - x_1^2)(1 - x_2^2)(1 - x_3^2)(1 - x_4^2) \cos\phi \\
& + 2x_1x_3(1 - x_0^2)(1 - x_2^2)(1 - x_4^2)(1 - x_5^2) \cos 2\theta \\
& - 2x_2x_4(1 - x_0^2)(1 - x_1^2)(1 - x_3^2)(1 - x_5^2) \\
& \times \cos(2\theta + \phi). \qquad (4.3)
\end{aligned}$$

This result can be checked against the known results for the rectangular, triangular, hexagonal, and Yamamoto lattices. A particular bond x_i of the lattice can be removed by putting $x_i = 0$. A bond can also be made redundant if we coalesce the spins which form the end points of the bond by putting $x_i = 1$. From Fig. 2 we can see that $x_0 = 1$ and any of the choices $x_2 = x_3 = x_4 = 0$ or $x_1 = x_3 = x_4 = 0$ or $x_1 = x_2 = x_4 = 0$ or $x_2 = x_3 = x_5 = 0$ will produce a rectangular lattice. For the first choice the determinant is

$$\begin{aligned}
\Delta_R = {} & (1 + x_1^2)(1 + x_5^2) \\
& - 2x_1(1 - x_5^2)\cos\theta - 2x_5(1 - x_1^2)\cos\phi.
\end{aligned}$$

The critical equation for this lattice can be found by the method developed in Sec. 3. For a singularity at $\theta = \phi = 0$, the critical equation is given by Eq. (3.5) as

$$\begin{aligned}
& c_0 - c_{18}x_1 - c_{27}x_2 - c_{36}x_3 - c_{49}x_4 - c_{510}x_5 \\
& + c_{1278}x_1x_2 - c_{1368}x_1x_3 - c_{1489}x_1x_4 + c_{15810}x_1x_5 \\
& - c_{2367}x_2x_3 - c_{2479}x_2x_4 - c_{25710}x_2x_5 - c_{3469}x_3x_4 \\
& - c_{35610}x_3x_5 - c_{45910}x_4x_5 - c_{123678}x_1x_2x_3 \\
& - c_{124789}x_1x_2x_4 - c_{1257810}x_1x_2x_5 - c_{134689}x_1x_3x_4 \\
& + c_{1356810}x_1x_3x_5 - c_{1456910}x_1x_4x_5 - c_{234679}x_2x_3x_4 \\
& + c_{2356710}x_2x_3x_5 + c_{2457910}x_2x_4x_5 - c_{3456910}x_3x_4x_5 \\
& - c_{12346789}x_1x_2x_3x_4 - c_{123567810}x_1x_2x_3x_5 \\
& - c_{124578910}x_1x_2x_4x_5 - c_{134568910}x_1x_3x_4x_5 \\
& - c_{234567910}x_2x_3x_4x_5 + c_{12345678910}x_1x_2x_3x_4x_5 = 0.
\end{aligned}$$

The sign of the terms have been found by using Eq. (3.6). The weight factors substituted from Eq. (4.1) give the critical equation after some algebraic manipulation as

$$\begin{aligned}
& (1 - x_0x_2)(1 - x_4x_5)(1 - x_1)(1 - x_3) \\
& \quad - (x_0 + x_2)(x_4 + x_5)(1 + x_1)(1 + x_3) = 0. \qquad (4.4)
\end{aligned}$$

The other critical equations corresponding to the other values of θ and ϕ at which a singularity may occur can be found by transforming the bond weights according to the sign factor given in Eq. (3.6). One can show that Eq. (4.4) gives the correct equation for the rectangular, triangular, hexagonal, and Yamamoto lattices by the same method as used in the previous paragraph. For example, if we put $x_1 = x_3 = 0$, we obtain the equation

$$1 + x_0x_2x_4x_5 = x_0x_2 + x_4x_5 + x_0x_5 + x_2x_4 + x_2x_5,$$

which corresponds with the Eq. (47a) of I for the Yamamoto lattice.

5. DISCUSSION

The above method of expanding the determinant in the partition function as a quadratic expression in the weight factors is applicable to any two-dimensional Ising lattice without crossed bonds possessing a periodic structure of similar cells. Fisher has shown how such lattices can be solved in principle in terms of a Pfaffian. However, for a complicated cell structure, his determinant is very large being of order $3(\Sigma_i q_i - 2r)$, where r is the number of spins in the Ising lattice cell and q_i is the number of bonds with spin. For the example of Sec. 4, a determinant of order 24 would require evaluation. Our method evaluates a determinant of order 10 using weight factors which can be found graphically.

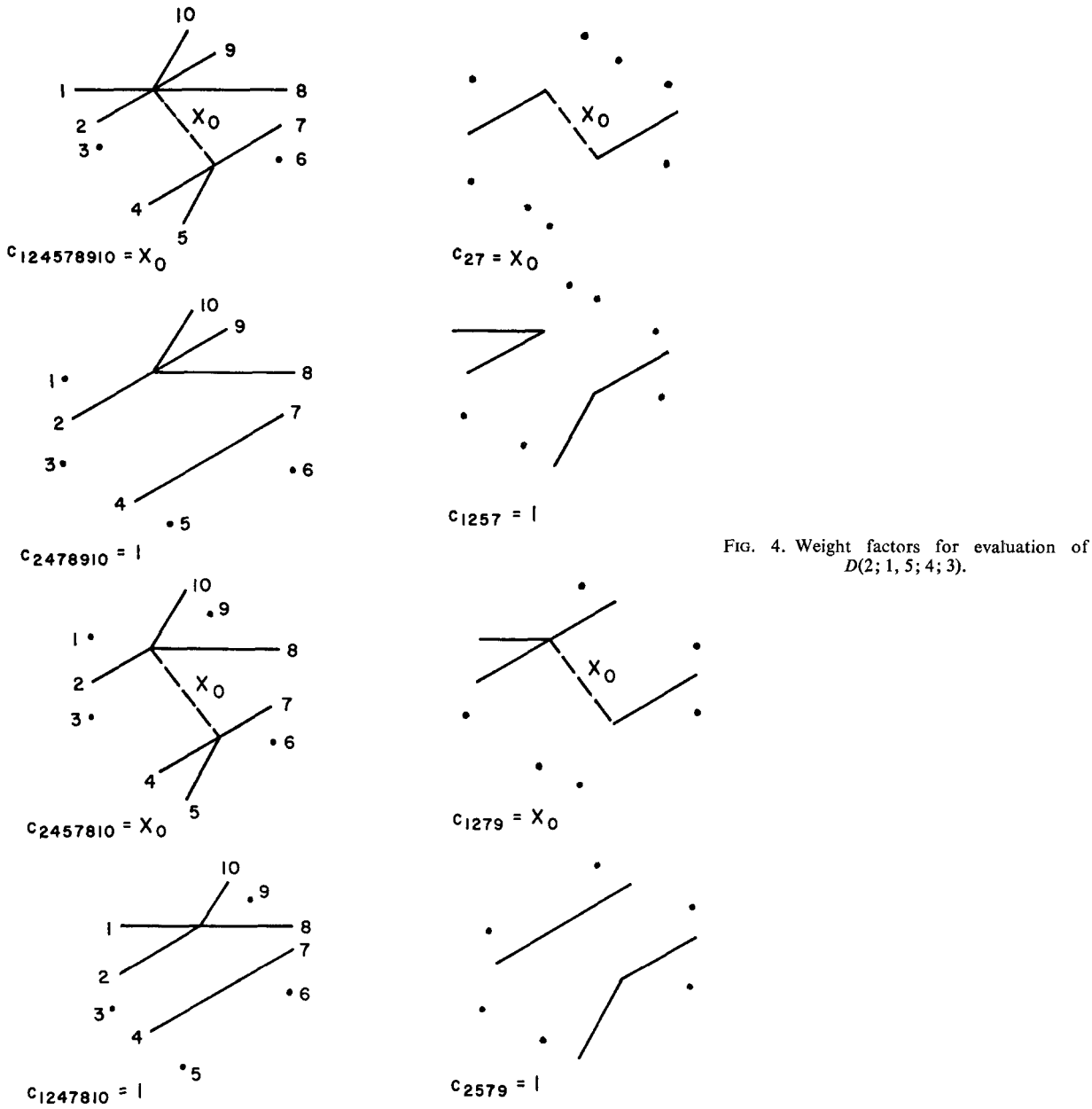

FIG. 4. Weight factors for evaluation of $D(2; 1, 5; 4; 3)$.

A cell with f spins has been considered as a multiple-state model with 2^f states for each cell.[6] An interaction energy between states can be defined and the conditions which must be imposed of the interaction energy for the problem to reduce to the combinatories of closed polygons on the lattice has been investigated.[7] If the cell structure is simple enough, the weight factors can be easily found and a model with many bonds between cells can be constructed. By choosing the strength of the bonds to decrease as the number of bonds increases, we hope to investigate a continuum model in a future paper.

[6] H. S. Green, Z. Physik 171, 129 (1963).
[7] R. G. J. Mills, Ph.D. thesis, University of Adelaide, 1966.

APPENDIX A: REDUCTION OF A DETERMINANT WITH AN ANTISYMMETRIC SUBMATRIX

We consider the reduction of a partly antisymmetric determinant $|C_{\lambda\lambda'}|$ of order $2m + r$. By partly antisymmetric we mean that the determinant has an antisymmetric submatrix of order $2m$. We select the smallest integer λ_{01} such that $C_{\lambda_{01}\lambda_{01}'}$ and $C_{\lambda_{01}'\lambda_{01}}$ are in this submatrix. Then to every row of the determinant other than the λ_{01}'th we add a constant multiple α_μ of the λ_{01}'th row so as to make all the elements of the λ_{01}'th column zero except $C_{\lambda_{01}\lambda_{01}'}$:

$$|C'_{\lambda\lambda'}| = |C_{\lambda\lambda'} + \Sigma_\mu \delta_{\lambda\mu}\alpha_\mu C_{\lambda_0'\lambda'}|,$$

with

$$C'_{\mu\lambda_{01}} = C_{\mu\lambda_{01}} + \alpha_\mu C_{\lambda_{01}'\lambda_{01}} = 0,$$

and hence

$$\alpha_\mu = -\frac{C_{\mu\lambda_{01}}}{C_{\lambda_{01}'\lambda_{01}}} = \frac{C_{\mu\lambda_{01}}}{C_{\lambda_{01}\lambda_{01}'}}. \tag{A1}$$

Then to every column of the determinant other than the λ'_{01}th we add a constant multiple β_μ of the λ'_{01}th column so as to make all the elements of the λ_{01}th row except $C_{\lambda_{01}\lambda_{01}'}$:

$$|C''_{\lambda\lambda'}| = |C'_{\lambda\lambda'} + \Sigma_\mu \delta_{\mu\lambda'}\beta_\mu C'_{\lambda\lambda_{01}'}|,$$

i.e.,

$$\begin{aligned}|C''_{\lambda\lambda'}| &= |C_{\lambda\lambda'} + \Sigma_\mu \delta_{\lambda\mu}\alpha_\mu C_{\lambda_{01}'\lambda'} + \Sigma_\mu \delta_{\mu\lambda'}\beta_\mu \\ &\quad \times \{C_{\lambda\lambda_{01}'} + \Sigma_\mu \delta_{\lambda\mu}\alpha_\mu C_{\lambda_{01}'\lambda_{01}'}\}| \\ &= |C_{\lambda\lambda'} + \Sigma_\mu \delta_{\lambda\mu}\alpha_\mu C_{\lambda_{01}'\lambda'} + \Sigma_\mu \delta_{\mu\lambda'}\beta_\mu C_{\lambda\lambda_{01}'}|,\end{aligned}$$

since $C_{\lambda_{01}'\lambda_{01}'} = 0$. To find β_μ we require $C''_{\lambda_{01}\mu} = 0$, i.e.,

$$C_{\lambda_{01}\mu} + \beta_\mu C_{\lambda_{01}\lambda_{01}'} = 0,$$

and hence

$$\beta_\mu = -C_{\lambda_{01}\mu}/C_{\lambda_{01}\lambda_{01}'}.$$

Thus

$$\begin{aligned}C''_{\lambda\lambda'} &= C_{\lambda\lambda'} + \frac{C_{\lambda\lambda_{01}}}{C_{\lambda_{01}\lambda_{01}'}} C_{\lambda_{01}'\lambda'} - \frac{C_{\lambda_{01}\lambda'}}{C_{\lambda_{01}\lambda_{01}'}} C_{\lambda\lambda_{01}'} \\ &= \frac{1}{C_{\lambda_{01}\lambda_{01}'}} \begin{vmatrix} C_{\lambda_{01}\lambda_{01}'} & C_{\lambda_{01}\lambda} & C_{\lambda_{01}\lambda'} \\ & C_{\lambda_{01}'\lambda} & C_{\lambda_{01}'\lambda'} \\ & & C_{\lambda\lambda'} \end{vmatrix}.\end{aligned} \tag{A2}$$

As a check, we note that $C''_{\lambda_{01}\lambda'} = C''_{\lambda\lambda_{01}} = 0$. The determinant is expanded by the λ_{01}th row and then the λ_{01}th column, giving

$$D = |C_{\lambda\lambda'}| = C_{\lambda_{01}\lambda_{01}'}^{2-(2m+r-2)} \, |\backslash C_{\lambda_{01}\lambda_{01}'} C_{\lambda_{01}'\lambda'} C_{\lambda\lambda'}|\,|. \tag{A3}$$

We note that the new submatrix formed from the original antisymmetric submatrix in this way is still antisymmetric because

$$\begin{aligned}C''_{\lambda_0\lambda_0'} &= \begin{vmatrix} C_{\lambda_{01}\lambda_{01}'} & C_{\lambda_{01}\lambda_{01}'} & C_{\lambda_0\lambda_{01}} \\ & C_{\lambda_{01}'\lambda_0'} & C_{\lambda_0\lambda_{01}'} \\ & & C_{\lambda_0\lambda_0'} \end{vmatrix} \\ &= \begin{vmatrix} C_{\lambda_{01}\lambda_{01}'} & C_{\lambda_{01}\lambda_0} & -C_{\lambda_0'\lambda_{01}} \\ & C_{\lambda_{01}'\lambda_0} & -C_{\lambda_0'\lambda_{01}'} \\ & & -C_{\lambda_0'\lambda_0} \end{vmatrix} \\ &= -C''_{\lambda_0'\lambda_0},\end{aligned}$$

and so the reduction process may be continued until all of the antisymmetric part of the original determinant has been removed.

APPENDIX B: DETERMINATION OF THE SIGN OF $D(\lambda_2; \lambda_1^+; \lambda_1^- \lambda_0)$

In this appendix we list the operations and their sign factors for the selection of the coefficient $D(\lambda_{2n}; \lambda_{1p}^+; \lambda_{1q}^-; \lambda_{0m})$ from Eq. (2.1) where all the c_{ij} are positive and the sign on the bond weights x_i above and below the diagonal are $(-1)^{i+i'}$ and $(-1)^{i+i'+1}$, respectively. From the relationship between the terminals and their associated terminals, we have

$$\begin{aligned}\text{sign } (i + i') &= \text{sign } (g + h + 1), \quad \text{if } 1 \le i \le h, \\ &= \text{sign } (g + d + 1), \quad \text{if } h < i \le h + d, \\ &= \text{sign } (g + v + 1), \quad \text{if } h + d < i \le g.\end{aligned}$$

(1) Selection of $\{\lambda_{2n}\}$ $(n = n_1 + n_2 + n_3)$ with $\alpha_\lambda = 2$. We select n bonds from the set g above the diagonal corresponding to $\lambda_{21} \cdots \lambda_{2n}$ and n bonds from the set g below the diagonal corresponding to $\lambda'_{21} \cdots \lambda'_{2n}$. Associated with each selection is a sign from the bond weight and a sign from the position in the determinant. The total sign change for all these selections is zero.

(2) Selection of $\{\lambda_{1p}^+\}$ and $\{\lambda_{1q}^-\}$ with $\alpha_\lambda = 1$. We select p bonds from those remaining above the diagonal and then q bonds with different labels from the p bonds from below the diagonal. Again we get a sign factor from the bond weight and a sign factor from the position in the determinant. The total sign factor is

$$\begin{aligned}&\tfrac{1}{2}p_1(1 - p_1) + \tfrac{1}{2}p_2(1 - p_2) + \tfrac{1}{2}p_3(1 - p_3) \\ &\quad + \tfrac{1}{2}q_1(1 - q_1) + \tfrac{1}{2}q_2(1 - q_2) \\ &\quad + \tfrac{1}{2}q_3(1 - q_3) + n(p + q) + n_1(p_1 + q_1) \\ &\quad + n_2(p_2 + q_2) + n_3(p_3 + q_3) + q(p + 1).\end{aligned}$$

(3) Selection of $\{\lambda_{0m}\}$ with $\alpha_\lambda = 0$. This selection is done while reducing the antisymmetric submatrix as in Appendix A. There is no bond weight so the sign comes only from the position in the determinant as

$$\begin{aligned}&(p + q)m + (p_1 + q_1)m_1 \\ &\qquad + (p_2 + q_2)m_2 + (p_3 + q_3)m_3.\end{aligned}$$

(4) Comparison of the reduced determinant Eq. (2.5) with the determinant (2.6) can be found by comparing the terms from the principal diagonal

$$\tfrac{1}{2}(p + q)(p + q - 1).$$

(5) Combining the row indices $\{\lambda^-, \lambda^{+\prime}\}$ and the column indices $\{\lambda^+, \lambda^{-\prime}\}$ in numerical order gives a

sign factor

$$pq + p_1q_1 + p_2q_2 + p_3q_3.$$

(6) Since D is a quadratic expression in the weight factors with $\boldsymbol{\lambda}_2$ appearing in both weight factors, there is no sign factor resulting from bringing the $\boldsymbol{\lambda}_2$ to numerical order.

(7) Finally, there is a sign factor for combining the numerical set $\boldsymbol{\lambda}_2$ and the numerical set $\{\lambda^-, \lambda^+, \lambda^{+\prime}, \lambda^{-1}\}$ into numerical order:

$$(p_1 + q_1)(m_2 + m_3) + (p_2 + q_2)(m_1 + m_3) + (p_3 + q_3)(m_1 + m_2).$$

If we combine the signs found above and use the relations

$$h = n_1 + p_1 + q_1 + m_1,$$
$$d = n_2 + p_2 + q_2 + m_2,$$
$$v = n_3 + p_3 + q_3 + m_3,$$
$$g = n + p + q + m,$$

we obtain the sign factor for D given in Eq. (2.8).

APPENDIX C: DETERMINATION OF THE SIGN OF $P(\lambda_1, \cdots, \lambda_n)$

The sign T of the Pfaffian $P(\lambda_1, \cdots, \lambda_n)$ given in Eq. (3.6) after selecting out the bonds $x_1^{\alpha_1}, \cdots, x_g^{\alpha_g}$ for which $\alpha_\lambda = 1$ is the product of the sign factor $(-1)^{\lambda+\lambda'}$ of the bond weight and the sign factor $(-1)^{i+j+1}$ corresponding to its position in the Pfaffian. We shall list the operations and their associated sign factors for the expansion of the Pfaffian by the method of Sec. 3.

(1) Selection of $\lambda_1, \cdots, \lambda_{n_1}$ from the set $\{1, 2, \cdots, h\}$. Here, $\lambda + \lambda' = g + h + 1$:

$$\begin{aligned}\text{sign} &= n_1(g + h + 1) + (1 + g + h + 1) \\ &\quad + (1 + g - 1 + h - 1 + 1) \\ &\quad + \cdots + (1 + g - n + 1 + h - n_1 + 1 + 1) \\ &= n_1(g + h + 1) + \tfrac{1}{2}n_1(8 + 4g + 4h - 2n_1) \\ &= n_1.\end{aligned}$$

In finding the position of the bond weight in the Pfaffian, one must remember to remove all the rows and columns containing the indices and the associated indices of the previously selected bond weights.

(2) Selection of $\lambda_{n_1+1}, \ldots \lambda_{n_1+n_2}$ such that $h < i < h + d$. Here, $\lambda + \lambda' = g + d + 1$ (mod. 2):

$$\begin{aligned}\text{sign} &= n_2(g + d + 1) \\ &\quad + (h - n_1 + 1 + g - n_1 + h - n_1 + d + 1) \\ &\quad + \cdots + (h - n_1 + 1 + g - n_1 \cdots n_2 \\ &\quad + 1 + h - n_1 + d - n_2 + 1 + 1) \\ &= n_1n_2 + n_2 \text{ (mod 2)}.\end{aligned}$$

(3) Selection of $\lambda_{n_1+n_2+1}, \ldots \lambda_n$ such that $h + d < i < g$. Here, $\lambda + \lambda' = g + v + 1$ (mod 2):

$$\begin{aligned}\text{sign} &= n_3(g + v + 1) + (h - n_1 + d - n_2 \\ &\quad + 1 + g + g - n_1 - n_2 + h - n_1 + d - n_2 \\ &\quad + v + \cdots + (h - n_1 + d - n_2 + 1 \\ &\quad + g - n_1 - n_2 - n_3 + 1 + h - n_1 + d \\ &\quad - n_2 + v - n_3 + 1 + 1) \\ &= n_1n_3 + n_2n_3 + n_3.\end{aligned}$$

Thus the total sign factor on selecting $\lambda_1, \cdots, \lambda_n$ is

$$T = n + n_1n_2 + n_2n_3 + n_1n_3. \qquad \text{(C1)}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 8 AUGUST 1969

One-Dimensional Phase Transition

JOSEPH L. STRECKER
Department of Physics, Wichita State University, Wichita, Kansas

(Received 30 January 1969)

The Berlin–Witten–Gersch model of a classical gas gives a phase transition in one dimension for a potential

$$V = -g\gamma e^{-\gamma R}, \quad R > \delta,$$
$$= \infty, \quad R < \delta,$$

in the limit $\gamma \to 0$. The correlation function is nonvanishing in the transition region; it vanishes outside. Investigation of the P–L isotherms (where L is the one-dimensional volume) shows a discontinuity in the first derivative at the transition points.

1. INTRODUCTION

Recent discussions of the theory of superconductivity as to whether a one-dimensional system such as a whisker can become superconducting[1] makes the question of one-dimensional phase transitions more than academic. Further, interest in phase transitions in one-dimensional systems has gained momentum recently due to Little's work[2] on long-chained polymers which may become superconducting. It is not too amiss to approximate these long polymers by one-dimensional systems.

Models which undergo a phase transition in one dimension are of interest especially because proof or evidence for the existence of one-dimensional phase transitions is still inconclusive. Existence or nonexistence depends markedly upon the model employed. We do have Van Hove's[3] famous work wherein he shows that a one-dimensional system cannot exhibit a phase transition if the forces are of finite range. The delicate nature of this type of transition is appreciated when one makes a comparison of our model with other similar ones. The Kac model[4] gives no phase transition whereas those of Kac, Uhlenbeck, and Hemmer[5] and Gersch[6] do.

We describe here a simple model of a classical gas which has long-range forces such that we can vary the range parameter, at the same time holding the total energy of the system constant. In the limit of infinitely long range, a phase transition is obtained.

We approach the problem of a classical gas with an attractive-pair potential $-g\gamma e^{-\gamma R}$ for $R \geq \delta$, where R is the distance between particles and g is a coupling constant. For $R < \delta$, the potential is infinite (positive) as in a hard-sphere gas. In particular, we are interested in the case when γ becomes arbitrarily small.

The role played by g can be seen by considering the potential energy. In a first approximation we may regard the density as uniform about a given particle and neglect the correlation function. The average value of the potential energy $\bar{U}$ per particle will then be

$$\bar{U} = \int_{\delta}^{\infty} U(R)\rho \, dR,$$

since in a given length element there will be $\rho \, dR$ particles (where ρ is the density) with energy $U(R)$. Adding up the energy over all the particles gives, for the total potential energy U_T,

$$U_T = \tfrac{1}{2}N\bar{U},$$

where the factor $\frac{1}{2}$ prevents our counting the same pair twice. Letting $\rho = N/L$, we obtain in the limit as $\gamma \to 0$

$$U_T = -\frac{N^2}{2L}\int_{\delta}^{\infty} g\gamma e^{-\gamma R}\, dR = -\frac{N^2}{2L}\, g.$$

That is, g measures the average potential energy of the system.

It is easy to show that a gas of hard spheres shows no condensation phenomena (even in three dimensions where phase transitions abound). A first guess might be deceiving in our problem, because with $\gamma = 0$ we are left with a potential identical to the hard-sphere gas. But the calculated results show that there is a transition region for a temperature below a critical temperature. As the parameter γ grows smaller, the "long-rangeness" of the potential "predominates" to effect a many-particle collective action.

Investigation of the P–L curves discloses the first derivative to be discontinuous at the transition points. The correlation function is nonvanishing in the transition region, whereas outside it vanishes. We have a first-order phase transition.

[1] R. A. Ferrel, Bull. Am. Phys. Soc. **2**, 315 (1966), AI1.
[2] W. A. Little, Phys. Rev. **134**, A1416 (1964).
[3] L. Van Hove, Physica **16**, 137 (1950).
[4] M. Kac, Phys. Fluids **2**, 8 (1959).
[5] M. Kac, G. E. Uhlenbeck, and P. C. Hemmer, J. Math. Phys. **4**, 216, 229 (1963); **5**, 60 (1964).
[6] H. A. Gersch, Phys. Fluids **6**, 599 (1963).

In comparing our results with those of Kac, Uhlenbeck, and Hemmer,[5] we note that our model leaves no uncertainty that in the limit $\gamma \to 0$ there is a sharp distinction in the correlation function between the condensed-phase and the single-phase region.

Our correlation function and the rate at which $x \to 0$ (Appendix A) both on the vapor dome and in the two-phase region agrees with the work of Gersch.[6] Our method does not permit comparison in the region of the transition for small γ, since it allows us to obtain only the limit functions.

2. THE CLASSICAL PARTITION FUNCTION

The configurational partition function for an imperfect gas is a sum over all the allowed states weighted by a Boltzman factor $\exp(-E/kT)$; in our case, this becomes an integral in configuration space:

$$Z_N = \frac{1}{N!}\int_V \cdots \int d\tau_1 \cdots d\tau_N \exp[-\beta W(r_1 \cdots r_N)], \tag{2.1}$$

where $W(r_1 \cdots r_N)$ is the potential energy of a configuration of particles, r_j the position vector of the ith particle, and $\beta = 1/kT$. The total free energy Ψ of the system composed of one type of particle is

$$\exp[-\beta\Psi] = [2\pi m/(\beta h^2)]^{\frac{3}{2}N} Z_N.$$

We will have more occasion to use the limiting value of the free energy per particle ψ, defined to be

$$\psi(v, \beta) = \lim_{N\to\infty} \frac{\Psi(N, V, \beta)}{N}, \qquad V/N = v, \tag{2.2}$$

where v is the specific volume. The thermodynamic pressure is given by

$$p = -\partial\psi/\partial v.$$

We now divide the total volume V into K cells of equal volume τ such that $K\tau = V$ and such that the cells are numbered from 1 to K. The number of particles in the jth cell is n_j. We can then write

$$Z_N = \frac{1}{N!}\sum_{\{n_j\}} \left(\frac{N!}{\Pi_{j=1}^k n_j!}\right) \int d\tau_1 \cdots \int d\tau_N \exp[-\beta W]. \tag{2.3}$$

The sum is over all combinations of integers $\{n_j\}$, $\eta_j \geq 0$, subject to the restriction

$$\sum_{j=1}^{k} n_j = N.$$

In principle, the specific subdivision of the space into cells should be irrelevant for the final result. We make the choice $\tau = v$, where v is the specific volume in n dimensions. Then $\langle n_j \rangle = 1$ for all values of β, v.

3. AN APPROXIMATION TO THE IMPERFECT GAS

We now seek to approximate Eq. (2.3). We proceed to list the assumptions and their direct consequences.[7]

(1) First we assume that the interaction between particles is pairwise,

$$W(r_1 \cdots r_N) = \sum_{1\leq i<j\leq N}' U(r_{ij}),$$

where r_{ij} is the distance between particles i and j, and $U(r)$ is the potential energy of interaction between a pair of particles.

(2) When two particles i and j are in different cells k and m, we assume that $U(r_{ij}) = U(R_{km})$, where R_{km} is the distance between centers of the kth and mth cells. The total interaction energy with respect to particles in different cells becomes

$$W(r_1 \cdots r_N) \cong \sum_{1\leq k<m\leq N}' n_k n_m U(R_{km}).$$

Note that this energy is now independent of the coordinates of the particles; it can be taken outside the integral.

Within a cell we define an average interaction by

$$\exp[-\beta W_k] \equiv \frac{1}{v^{n_k}} \int d\tau_1 \cdots \int d\tau_{n_k} \exp[-\beta W(r_1 \cdots r_{n_k})].$$

(3) Let W_0 be defined by

$$W_k \equiv \tfrac{1}{2} n_k(n_k - 1)W_0.$$

In general, W_0 will be a function of v, β, and n_k. We now assume that the dependence of W_0 on n_k can be neglected.

The consequences of assumptions (1), (2), and (3) lead to

$$Z_N = \sum_{\{n_j\}} \left(\frac{v^N}{\Pi_{j=1}^N n_j!}\right) \exp\left[\tfrac{1}{2}N\beta W_0 - \beta \sum_{j,k} a_{jk} n_j n_k\right],$$

where use of the following identity has been made,

$$\sum_k W_k = \sum_k n_k(n_k - 1)W_0 = W_0 \sum_k (n_k^2 - n_k) = -NW_0 + \sum_k a_{kk} n_k n_k,$$

and where

$$a_{jk} = \tfrac{1}{2}U(R_{jk}), \quad j \neq k, \tag{3.1}$$

$$a_{jj} = \tfrac{1}{2}W_0, \quad j = k. \tag{3.2}$$

(4) The sum over the set $\{n_j\}$, where n_j are non-negative integers and subject to the condition

$$\sum_{j=1}^{N} n_j = N,$$

[7] T. H. Berlin, L. Witten, and H. A. Gersch, Phys. Rev. 92, 189 (1953).

is a sum over lattice points, in an N-dimensional Euclidean space, bounded by an $(N-1)$-dimensional figure, a "hyperhedron." The center of this hyperhedron is the point $(1, 1, \cdots, 1)$, which describes the configuration with one particle in each cell. This is the average configuration, since $\langle n_j \rangle = 1$.

Our next assumption is to replace the sum over the lattice points by an integration throughout the volume Ω of the hyperhedron. We write

$$Z_N = A(N)v^N \exp\left[\tfrac{1}{2}N\beta W_0\right] \times \int_\Omega dn_1 \cdots \int dn_N \left\{\exp\left[-\beta \sum_{\zeta,k} a_{jk} n_j n_k\right] \Big/ \left[\Pi_{j=1}^N \Gamma(n_j + 1)\right]\right\}, \quad (3.3)$$

where $A(N)$ is a normalization constant.

(5) The division of physical space into cells of volume v implies that the fluctuation in the number of particles per cell, $\sigma = \langle (n_j - 1)^2 \rangle$, is finite for all v, β, unless of course purely attractive forces are present. This suggests that a significant contribution to the above integral arises from points which lie on the intersection of the hyperplane

$$\sum_{j=1}^N n_j = N$$

with the N-dimensional sphere

$$\sum_{j=1}^N (n_j - 1)^2 = R^2,$$

where $R^2 = N\sigma$. The intersection is the surface of an $(N-1)$-dimensional sphere with center at the center of the hyperhedron. The volume of the intersection is of the order of magnitude of the volume of the hyperhedron.

The volume integral in Eq. (3.3) may be written

$$\int_\Omega dn_1 \cdots \int dn_N = \int_0^N dR \int_{\sum_{j=1}^N (n_j-1)^2 = R^2,\ \sum_1^N n_j = N} dn_1 \cdots \int dn_N .$$

The difficulty here is that, for $R > 1$, a good part of the surface of the $(N-1)$-dimensional sphere lies outside the hyperhedron. However, integration becomes feasible if the whole surface of the sphere is an allowed region. We allow this extended region, but we introduce a weighting factor to compensate for the additional unwanted states. We employ a factor which is 1 for $R < 1$, which is of the order of e^{-N} for $R \sim N^{\frac{1}{2}}$, and which is negligible with respect to e^{-N} for $R \sim N$. We, therefore, assume that

$$\int_\Omega dn_1 \cdots \int dn_N \doteq \int_0^\infty dR \exp\left[-\frac{\alpha}{2N} R^4\right] \int_{\sum_1^N (n_j-1)^2 = R^2,\ \sum_1^N n_j = N} dn_1 \cdots \int dn_N$$

where α is a finite constant.

(6) Since in both the gas and liquid regions we can expect $\sigma \sim 1$, the $\Gamma(n_j + 1)$ will be mostly of order 1. Let us then approximate the gamma function by $\exp[an_j(n_j - 1)]$. We could then choose $a = \frac{1}{2}\ln 2$ to fit the point $n_j = 2$. But it is more interesting to let a play the role of a parameter and see its effects upon the isotherms and the critical thermodynamic variables as done in Sec. 6.

(7) The final assumption is to set the Born–von Kármán periodic boundary conditions on the physical space. A one-dimensional gas would be confined to a ring, a two-dimensional gas to the surface of a torus. The mathematical analogy is extended to the three-dimensional gas. The effect of this assumption is to cyclize the interaction matrix a_{ij}.

The partition function of our approximation to the imperfect gas is

$$Z_N = Av^N \exp\left[\tfrac{1}{2}N\beta W_0\right] \int_0^\infty dR \exp\left[-\frac{\alpha}{2N} R^4\right] \times \int dn_1 \cdots \int dn_N \times \exp\left[-a\sum_{j=1}^N n_j(n_j - 1) - \beta \sum_{j,k} a_{jk} n_j n_k\right].$$

We fix the constant A by comparison with the ideal gas; in the limit the potential we use is that of a gas of hard spheres.

4. EVALUATION OF THE PARTITION FUNCTION

Let us transform from the variables $\{n_j\}$ to the variables $\{y_j\}$ by a translation and rotation such that

$$\sum_1^N (n_j - 1)^2 = \sum_1^N y_j^2, \quad (4.1)$$

$$\sum_{j,k} a_{jk}(n_j - 1)(n_k - 1) = \sum_1^N \Lambda_j y_j^2. \quad (4.2)$$

The eigenvalues Λ_k and eigenvectors t_{jk} normalized to unity are given by[8]

$$\Lambda_k = a_{11} + 2\sum_{p=2}^{(N+1)/2} a_{1p} \cos\left[\frac{2\pi}{N}(k-1)(p-1)\right], \quad (4.3)$$

$$t_{jk} = N^{-\frac{1}{2}}\{\cos[(2\pi/N)(j-1)(k-1)] + \sin[(2\pi/N)(j-1)(k-1)]\}. \quad (4.4)$$

[8] T. H. Berlin and M. Kac, Phys. Rev. **86**, 821 (1952).

The subscript p refers to the cell number with respect to cell 1. With the exception of Λ_1, the eigenvalues are doubly degenerate because

$$\Lambda_k = \Lambda_{N-k+2}, \quad k \neq 1.$$

The hyperplane is described by $y_1 = 0$. This can readily be seen from the transformation equation, $x = Ty$, which, upon solving for y_1, becomes

$$y_1 = \sum_k t_{1k}x_k = N^{-\frac{1}{2}} \sum x_k = N^{-\frac{1}{2}} \sum (n_k - 1) = 0.$$

Employing Eqs. (4.1) and (4.2), our partition function now becomes

$$Z_N = B\int_0^\infty dR \exp\left[-\frac{\alpha}{2N}R^4\right] \int\cdots\int_{\sum y_j^2 = R^2} dy_2 \cdots dy_N \times \exp\left[-\sum_{j=2}^N (a + \beta\Lambda_j)y_j^2\right], \quad (4.5)$$

where we define $B \equiv Av^N \exp[N\beta(\frac{1}{2}W_0 - \Lambda_1)]$.

Relaxing the restriction on the region of integration of the variables $\{y_j\}$ by means of the delta function, then

$$Z_N = \frac{B}{2\pi i}\int_{S_0-i\infty}^{S_0+i\infty} dS \exp\left(\frac{NS^2}{2\alpha}\right)\int_{-NS/\alpha}^\infty dw \times \exp\left(-\frac{\alpha}{2N}w^2\right)\left\{\prod_{j=2}^N \pi^{\frac{1}{2}}(S + a + \beta\Lambda_j)^{-\frac{1}{2}}\right\}. \quad (4.6)$$

The product in the integrand can be written as

$$\prod_{j=2}^N (S + a + \beta\Lambda_j)^{-\frac{1}{2}} = \exp\left[-\tfrac{1}{2}\sum_2^N \ln(S + a + \beta\Lambda_j)\right].$$

We are interested in the limit as $N \to \infty$. Let

$$f(S) = \lim_{N\to\infty} \frac{1}{N}\sum_{j=2}^N \ln(S + a + \beta\Lambda_j). \quad (4.7)$$

Then (for one dimension)

$$f(S) = \frac{1}{2\pi}\int_0^{2\pi} d\omega \ln[S + a + \beta\Lambda(\omega)], \quad (4.8)$$

where

$$\Lambda(\omega) = a_{11} + \lambda(\omega) = a_{11} + 2\sum_r a_{1r}\cos\frac{\omega r}{L}. \quad (4.9)$$

The sum $\sum_r$ is over all the cell centers in the physical space, excluding cell 1. The distance r has its origin at the center of cell 1.

Let $\tilde{S}$ denote the algebraically smallest value of $a + \beta\lambda(\omega)$. If the s plane is cut from $s = -\infty$ to $s = -\tilde{S}$ along the real axis, then the integrand is analytic in the cut plane. The behavior of the integral in the neighborhood of $s = -\tilde{S}$ is characterized by the density of eigenvalues $\Lambda(\omega)$ in the neighborhood of the algebraically smallest eigenvalue. This behavior incompletely describes the behavior of Eq. (4.8) in the neighborhood of $s = -\tilde{S}$ because the singular nature of the integral has in general nothing to do with the singular nature of a single term in the sum, Eq. (4.7). We may, therefore, separate out this singular point.

In the integral over w in Eq. (4.6), the important values of w are of the order of $N^{\frac{1}{2}}$. For finite s, the lower limit of integration is of the order of N. In this situation we can extend the lower limit to negative infinity; the integral over w becomes $(2\pi N/\alpha)^{\frac{1}{2}}$ if s is positive. We must, therefore, evaluate

$$Z_N = \frac{B}{2\pi i}\pi^{\frac{1}{2}(N-1)}\left(\frac{2\pi N}{\alpha}\right)^{\frac{1}{2}}\int_{S_0-i\infty}^{S_0+i\infty}\frac{e^{NG(S)}}{S + \tilde{S}}dS, \quad (4.10)$$

where

$$G(S) \equiv (S^2/2\alpha) - \tfrac{1}{2}f(S).$$

Applying the method of steepest descent to the integral, we find that

$$Z_N = \frac{B\pi^{\frac{1}{2}(N-1)}e^{NG(S)}}{(S + \tilde{S})[\alpha(\partial^2 G/\partial S^2)]^{\frac{1}{2}}_{S=S_s}},$$

if a saddle point S_s can be found such that S_s is real, positive, to the right of the singularities of the integrand, and with

$$\left[\frac{\partial G(S)}{\partial S}\right]_{S_s} = 0, \quad \left[\frac{\partial^2 G(S)}{\partial S^2}\right]_{S_s} > 0.$$

The constant A shall be determined by normalizing to the ideal gas, for which $Z_N = V^N/N!$. For the ideal gas $a_{jk} = 0$ for all j, k, so that $\Lambda_j = 0$ for all j. Then

$$G(S) = (S^2/2\alpha) - \tfrac{1}{2}\ln(S + a),$$

$$\left[\frac{\partial G(S)}{\partial S}\right]_{S_s} = \frac{S_s}{\alpha} - \frac{1}{2(S_s + a)} = 0,$$

$$\left[\frac{\partial^2 G(S)}{\partial S^2}\right]_{S_s} = \frac{1}{\alpha} + \frac{1}{2(S_s + a)^2} > 0.$$

The solution of the saddle-point equation that meets the requirements is

$$S_s = S_{id} = \tfrac{1}{2}(a^2 + 2\alpha)^{\frac{1}{2}} - \tfrac{1}{2}a.$$

The positive square root is chosen because $G(S)$ has a cut to left of $s = -a$. It is found that

$$\lim_{N\to\infty}\frac{\ln A}{N} = \ln\frac{e}{\pi^{\frac{1}{2}}} - G(S_{id}).$$

For the free energy per particle this gives

$$-\beta\psi = \tfrac{1}{2}\ln(2\pi m/\beta h^2) + \ln ev + \beta[\tfrac{1}{2}W_0 - \Lambda(0)] + G(S_s) - G(S_{id}). \quad (4.11)$$

For our one-dimensional gas model, the saddle-point equation becomes

$$\frac{2S_s}{\alpha} = \frac{1}{2\pi}\int_0^{2\pi} \frac{d\omega}{S_s + a + \beta\Lambda(\omega)} \equiv I(S_s, \gamma). \quad (4.12)$$

In our calculations we approximate this sum by an integral; essentially we are averaging the interaction over the distant cells:

$$\lambda(\omega) \cong \frac{2}{L}\int' dR a(R)\cos(\omega R/L), \quad (4.13)$$

where the prime signifies that the integration is over all space excluding a length L centered at the origin.

We employ a potential $-g\gamma e^{-\gamma R}$ for $\delta \leq R \leq \infty$, where δ is the diameter of our particles. For $R < \delta$ we assume hard spheres: the potential is infinite (positive). The definition of g has been clarified in Sec. 1. (We have now confined ourselves exclusively to one dimension: $v = L$.) We are interested in the thermodynamic quantities in the limit as $\gamma \to 0$.

From Eq. 4.13 we obtain

$$\lambda(\omega) = -\frac{2g}{L}\frac{\gamma e^{-\gamma L/2}}{\gamma^2 + (\omega/L)^2}\left(\gamma\cos\frac{\omega}{2} - \frac{\omega}{L}\sin\frac{\omega}{2}\right),$$

$$\lambda(0) = -\frac{2g}{L}e^{-\gamma L/2}.$$

These values follow readily from the definition of $\lambda(\omega)$:

$$\lambda(\omega) = -\frac{2g\gamma}{L}\int_{L/2}^{\infty} dR e^{-\gamma R}\cos\frac{\omega R}{L}$$

$$= \frac{2g\gamma}{L}\left[\frac{e^{-\gamma R}\cos[(\omega R/L) + \mathfrak{A}]}{[\gamma^2 + (\omega/L)^2]^{\frac{1}{2}}}\right]_{L/2}^{\infty},$$

where

$$\tan\mathfrak{A} = \omega/\gamma L.$$

Substituting limits and using the equation

$$\cos(x + y) = \cos x\cos y - \sin x\sin y,$$

we obtain $\lambda(\omega)$.

The saddle-point equation can be rewritten as

$$\frac{2S_s}{\alpha} = \frac{1}{2\pi}\int_0^{2\pi}\frac{d\omega}{S_s + a + \beta a_{11} + \beta\lambda(\omega)}$$

$$\equiv \frac{1}{2\pi}\int_0^{2\pi}\frac{d\omega}{-\beta\lambda(0)[1 + x] + \beta\lambda(\omega)},$$

where we define ϵ, x by

$$S_s + \tilde{S} \equiv \epsilon \equiv -\beta\lambda(0)x.$$

After a slight rearrangement we can rewrite this as

$$\frac{2S_s}{\alpha} = \frac{1}{2\pi}\frac{1}{-\beta\lambda(0)}\frac{1}{1 + x}$$

$$\times\left\{2\pi + \int_0^{2\pi}\frac{\lambda(\omega)\,d\omega}{\lambda(0)[1 + x] - \lambda(\omega)}\right\}. \quad (4.14)$$

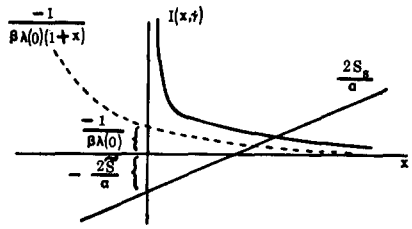

FIG. 1. Behavior of saddle-point equations for small γ. $2S_s/\alpha$ intersects $I(x, \gamma)$ at $x_s(\gamma)$ and the dotted curve, $[-\beta\lambda(0)(1 + x)]^{-1}$, at $x_s(0)$.

Let us define the integral as $K(x, \gamma)$. In Appendix A we prove that for all values of $x > 0$, the limit of K, as $\gamma \to 0$, is zero. However, if x is set equal to zero, the integral $K(x, \gamma)$ blows up. Therefore, if $K(x, \gamma)$ is to be nonvanishing, x must go to zero as $\gamma \to 0$. Just how rapidly x will do this is proven in Appendix A, where K is evaluated.

We evaluate $K(x, \gamma)$ by placing an upper and a lower bound on it and discovering that as $\gamma \to 0$ the two bounds converge to the same limit value.

As $\gamma \to 0$, the value of x can change in order that the saddle-point equation holds.

In Fig. 1 we have plotted qualitatively the monotonically decreasing function $I(x, \gamma)$ in solid line and $2S_s/\alpha$ against x. We have made a transformation:

$$S + \tilde{S} = -\beta\lambda(0)x.$$

For finite γ, I goes to infinity at $x = 0$. It is also obvious that $K(x, \gamma)$ is always nonnegative. By $x_s(\gamma)$ and $x_s(0)$ we denote the saddle points for the two cases, respectively, when γ is finite and when $\gamma \to 0$. In this limit

$$I(x, \gamma) \to \frac{1}{1 + x}\frac{1}{-\beta\lambda(0)}, \quad x > 0,$$

which is the dotted line in Fig. 1.

The function $\tilde{S} \equiv a + \beta a_{11} + \beta\lambda(0)$ for large values of L is about equal to a. It decreases to a minimum value as L decreases, and then increases to infinity as L decreases to a value of the order of the particle radius. That is, in Fig. 1, the intercept $2S/\alpha$ starts at $-a$ for $L = \infty$, rises, and then returns through $-a$ on to $-\infty$ as L goes from infinity to some quantity of the order of δ.

We show in Appendix B that for sufficiently large values of β, the intercept $-2\tilde{S}/\alpha$ crosses and rises above the intercept $[-\beta\lambda(0)]^{-1}$. The range of L for which this happens we call the transition or condensation region, and we call the two points of crossing the transition or condensation points. When this happens, it is clear that $x \to 0$ so that the $K(x, \gamma)$ function absorbs the difference between $2S/\alpha$ and $[-\beta\lambda(0)]^{-1}$. If x is set equal to zero, $K(x, \gamma) = \infty$. This is obvious

from inspection of the denominator of the integrand which defines $K(x, \gamma)$; it vanishes as ω^2. Therefore, the only alternative is for $x \to 0$ as $\gamma \to 0$ in some way which keeps $K(x, \gamma)$ finite. The rate of x is γ^2 and is proven in Appendix A.

Let us now review Fig. 1 as L goes from infinity to δ. At infinity the two intercepts are separated. As L decreases, the two intercepts approach each other. For sufficiently small β, they reach a minimal distance and then separate again. There is no phase transition. If, however, β is sufficiently large, then, as L decreases, the intercepts cross each other; the intercept $-2\tilde{S}/\alpha$ rises above that of $[-\beta\lambda(0)]^{-1}$. For continually decreasing values of L, then, the separation of the intercepts will reach a maximum, then decrease, cross each other again and recede. With L approaching δ, $-2\tilde{S}/\alpha$ goes to negative infinity.

Let us now define a critical temperature β_c and critical volume L_c. For β extremely small, there clearly exists no transition, and for β extremely large, there is a range of L. However, between these extremes, there will be some β_c, L_c such that the two intercepts just touch.

5. CORRELATIONS

We define a correlation C_{jk} between the number of molecules n_j and n_k situated, respectively, in the jth and kth cells as

$$C_{jk} \equiv \frac{\langle n_j n_k \rangle}{\langle n_j^2 \rangle^{\frac{1}{2}} \langle n_k^2 \rangle^{\frac{1}{2}}},$$

where the average value of a function F is defined to be

$$\langle F \rangle \equiv \frac{\int_V d\tau_1 \cdots \int_V d\tau_N F \exp\left[-\beta W(r_1 \cdots r_N)\right]}{\int_V d\tau_1 \cdots \int_V d\tau_N \exp\left[-\beta W(r_1 \cdots r_N)\right]}.$$

Since the average number of particles per cell is 1, it would be more appropriate to use the variable

$$x_j = n_j - 1.$$

We obtain then

$$\begin{aligned}\langle x_j x_k \rangle &= \frac{1}{2}\frac{1}{2\pi}\int_0^{2\pi} \frac{\cos(\omega R/L)\, d\omega}{S_s + a + \beta a_{11} + \beta\lambda(\omega)} \\ &= \frac{1}{2}\frac{1}{2\pi}\int_0^{2\pi} \frac{\cos(\omega R/L)\, d\omega}{-\beta\lambda(0)(1+x) + \beta\lambda(\omega)} \\ &= \frac{1}{2}\frac{1}{2\pi}\frac{1}{-\beta\lambda(0)[1+x]}\left\{\int_0^{2\pi} \cos\frac{\omega R}{L}\, d\omega \right. \\ &\quad \left. - \int_0^{2\pi} \frac{\lambda(\omega)\cos(\omega R/L)\, d\omega}{-\lambda(0)[1+x] + \lambda(\omega)}\right\}.\end{aligned}$$

The first integral vanishes, since $R = nL$, where n is an integer. We compare the second integral with $K(x, \gamma)$. If we expand the cosine function, the first order term gives $K(x, \gamma)$. Higher-order terms vanish as $\gamma \to 0$, since higher powers of ω appear in the numerator of the integrand. Therefore we write

$$\lim_{\gamma \to 0} \langle x_j x_k \rangle = \frac{1}{2}\frac{1}{2\pi}\frac{1}{-\beta\lambda(0)[1+x]} K(x, \gamma).$$

We see, therefore, that the correlation function $\langle x_j x_k \rangle$ vanishes outside the transition region, but inside it takes some finite value.

6. PRESSURE AND ISOTHERMS

The equation of state can now be obtained from Eq. (4.11) by the relation $p = -\partial\psi/\partial L$:

$$\begin{aligned}p\beta &= \frac{1}{L} + \frac{\beta}{2}\frac{\partial W_0}{\partial L} - \beta\frac{\partial\Lambda(0)}{\partial L} \\ &\quad - \frac{1}{2}\frac{1}{2\pi}\int_0^{2\pi} d\omega \left\{\beta\frac{\partial\Lambda(\omega)}{\partial L}\Big/[S_s + a + \beta\Lambda(\omega)]\right\},\end{aligned}$$

where use of the saddle-point Eq. (4.12) has been made.

Rewriting this equation and making use of the results of Appendix C,

$$\begin{aligned}p &= \frac{1}{\beta L} + \frac{1}{2}\frac{\partial W_0}{\partial L} - \frac{\partial\Lambda(0)}{\partial L} + \frac{\partial\Lambda(0)}{\partial L}\left(\frac{\tilde{S} - \epsilon}{\alpha}\right) \\ &\quad + \frac{1}{2\beta L}\left\{1 + \frac{2\epsilon}{\alpha}(\tilde{S} - \epsilon)\right\} \quad (\gamma \to 0) \\ &= \frac{3}{2\beta L} - \frac{2g}{L^2} + \left[\frac{2g}{L^2} - \frac{\delta}{\beta L(L-\delta)} + \frac{\epsilon}{\beta L}\right]\left[\frac{\tilde{S} - \epsilon}{\alpha}\right].\end{aligned} \tag{6.1}$$

Clearly the pressure is a continuous function at the transition point. The derivative of the isotherm is

$$\begin{aligned}\frac{\partial P}{\partial L} &= -\frac{3}{2\beta L^2} + \frac{4g}{L^3} \\ &\quad + \left[-\frac{4g}{L^3} + \frac{\delta(2L-\delta)}{\beta L^2(L-\delta)^2} \right. \\ &\quad \left. - \frac{\epsilon}{\beta L^2} + \frac{1}{\beta L}\frac{\partial\epsilon}{\partial L}\right]\left[\frac{\tilde{S} - \epsilon}{\alpha}\right] \\ &\quad + \left(\frac{1}{\alpha}\right)\left[\frac{2g}{L^2} - \frac{\delta}{\beta L(L-\delta)} + \frac{\epsilon}{\beta L}\right] \\ &\quad \times \left[\frac{2g\beta}{L^2} - \frac{\delta}{L(L-\delta)} - \frac{\partial\epsilon}{\partial L}\right],\end{aligned} \tag{6.2}$$

where

$$\tilde{S} \equiv a - \ln\left[1 - (\delta/L)\right] - 2g\beta/L.$$

We define

$$\Delta\left(\frac{\partial P}{\partial L}\right) = \left(\frac{\partial P}{\partial L}\right)_{\text{normal}} - \left(\frac{\partial P}{\partial L}\right)_{\text{transition}}.$$

It is clear from Eq. (6.2) that there is a singularity only if there is a difference between $\partial\epsilon/\partial L$ in the two regions. Clearly such quantities as $\tilde{S}$ and $\partial\tilde{S}/\partial L$ will be the same on both sides of the transition point.

An equation for ϵ and its derivatives can readily be obtained from the relation for the saddle point:

$$\frac{2}{\alpha}(-\tilde{S} + \epsilon) = \frac{1}{2\pi}\frac{1}{-\beta\lambda(0)}\frac{1}{1+x}\{2\pi + K(x,\gamma)\}. \tag{6.3}$$

For the *normal* region, $K(x,\gamma) \to 0$. We take derivatives of the above equation and remembering that, at the transition point $\epsilon \to 0$,

$$-\left(\frac{\partial\epsilon}{\partial L}\right)_{\text{normal}}(\tilde{S} + \beta\lambda(0)) + \tilde{S}\beta\frac{\partial\lambda(0)}{\partial L} + \beta\lambda(0)\frac{\partial\tilde{S}}{\partial L} = 0, \tag{6.4}$$

which determines $(\partial\epsilon/\partial L)$ for the normal region.

For the *transition* region, $K(x,\gamma)$ does not vanish. In Appendix A we have shown

$$K(x,\gamma) = \frac{\gamma}{(Ax)^{\frac{1}{2}}}\tan^{-1}\frac{2\pi}{\gamma}\left(\frac{A}{x}\right)^{\frac{1}{2}} \qquad A = \frac{1+x}{L^2} - \frac{\gamma}{2L}$$

in the limit of $\gamma \to 0$. A sufficiently accurate approximation to $K(x,\gamma)$ is

$$K(x,\gamma) = \frac{\gamma L}{x^{\frac{1}{2}}}\left(1 - \frac{x}{2}\right)\tan^{-1}\left[\frac{2\pi}{x^{\frac{1}{2}}\gamma L}\left(1 + \frac{x}{2}\right)\right],$$

since on the vapor dome x behaves as $\gamma^{\frac{2}{3}}$. We now calculate $\partial K/\partial L$ as $x \to 0$. The only remaining term is

$$\frac{\partial K(x,\gamma)}{\partial L} = \frac{(-\frac{1}{2})\gamma L}{x^{\frac{3}{2}}}\frac{\partial x}{\partial L}\tan^{-1}\left[\frac{2\pi}{x^{\frac{1}{2}}\gamma L}\left(1 + \frac{x}{2}\right)\right]$$
$$= -\frac{L}{2c}\frac{\partial x}{\partial L}\frac{\pi}{2},$$

where c is introduced by $\lim_{\gamma\to 0} x = c\gamma^{\frac{2}{3}}$.

Using the saddle-point equation (6.3) and taking derivatives, as $\gamma \to 0$, we obtain

$$-\left(\frac{\partial\epsilon}{\partial L}\right)_{\text{trans}}(\tilde{S} + \beta\lambda(0)) + \beta\lambda(0)\frac{\partial\tilde{S}}{\partial L} + \beta\tilde{S}\frac{\partial\lambda(0)}{\partial L}$$
$$= \frac{\alpha L}{16c}\frac{1}{\beta\lambda(0)}\left(\frac{\partial\epsilon}{\partial L}\right)_{\text{trans}}. \tag{6.5}$$

Comparing Eqs. (6.4) and (6.5), we see that there is a singularity at the transition point.

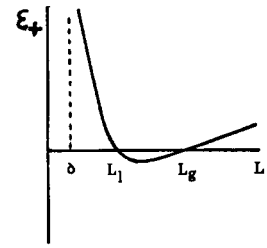

FIG. 2. A qualitative plot of ϵ_+ as a function of L.

In plotting the isotherms and especially the vapor dome, it is instructive to plot qualitatively the curve for ϵ_+ in the transition region, where

$$2\epsilon_\pm = A \pm (A^2 - D)^{\frac{1}{2}}.$$

A and D are defined and discussed in Appendix D. We choose the positive sign, since, by inspecting $\epsilon_\pm$ for L near δ and $L \to \infty$, we see that $\epsilon_+ > 0$ and $\epsilon_- < 0$. (See Fig. 2.)

We now investigate the interplay of the various parameters in determining the critical thermodynamic variables and the isotherms. On the vapor dome the equation of state is

$$P_{vd} = \frac{1}{\beta}\left\{\frac{3}{2L} - \frac{2Z}{L^2} + \left(\frac{L}{4Z}\right)\left[\frac{\delta}{L(L-\delta)} - \frac{2Z}{L^2}\right]\right\}, \tag{6.6}$$

where we have found it convenient to introduce $Z \equiv g\beta$. We note that p_{vd} is explicitly independent of α and a.

We may wish to prevent the pressure over a certain region of the vapor dome from becoming negative; in particular, the pressure at the critical point should be positive. We can accomplish this if we demand in Eq. (6.6) that the positive terms be larger than the negative ones. That is,

$$\frac{1}{L} + \frac{1}{4Z(L-1)} > \frac{2Z}{L^2},$$

which becomes

$$P(Z) \equiv Z^2 - \tfrac{1}{2}ZL - L^2/[8(L-1)] < 0. \tag{6.7}$$

Setting $P(Z) = 0$ and calling the solutions $2\zeta_+$, $2\zeta_-$ for the zeros of $P(Z)$, we obtain

$$2\zeta_\pm = \frac{L}{2} \pm \frac{L}{2}\left(1 + \frac{2}{L-1}\right)^{\frac{1}{2}}.$$

$2\zeta_-$ is clearly always negative. We discard this possibility. Hence a plot of $2\zeta_+$ indicates it has a minimum at $L = \sqrt{2}$ such that $2\zeta_+(\sqrt{2}) = 1 + \sqrt{2}$. It goes to ∞, as $L \to 1$ and $L \to \infty$. For values of Z below $2\zeta_+$, we can always be assured that the pressure on the vapor dome is positive.

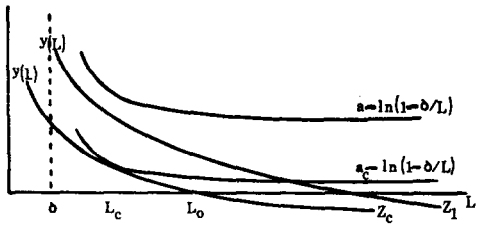

FIG. 3. For a given Z_c ($= g\beta_c$) and α, the text describes the manner in which the critical volume L_c and a_c are determined.

Let us now recall Eq. (4.14) for the vapor dome:

$$\tilde{S}\beta\lambda(0) = \alpha/2 \tag{6.8}$$

or

$$a - \ln\left(1 - \frac{\delta}{L}\right) = \frac{2Z}{L} - \frac{\alpha}{2}\left(\frac{L}{2Z}\right) \equiv y(L).$$

In Fig. 3 we have qualitatively plotted the function $a - \ln(1 - \delta/L)$. We can now decide on a critical value of the temperature Z_c. We plot $y(L)$ for this desired Z_c for some value of α. The zero of $y(L)$, L_0 occurs at

$$L_0 = 2Z_c(2/\alpha)^{\frac{1}{2}}.$$

We now lower the curve $a - \ln(1 - \delta/L)$ by decreasing a until the curve just touches $y(L)$ at one point L_c. This determines a_c and L_c.

Now as Z increases above the critical values, to Z_1 say, the zero L_0 shifts to the right, and the curve $y(L)$ rotates clockwise, both features contributing to $y(L)$ intersecting the curve for the logarithm at two points L_g and L_l, which are the gas- and liquid-phase transition points, respectively. We note that this analysis does not depend upon whether L_c is less than or greater than L_0.

We rewrite Eq. (4.14) for the vapor dome by defining

$$\Delta(Z, L, a) \equiv a - \ln\left(1 - \frac{\delta}{L}\right) - \frac{2Z}{L} + \frac{\alpha}{2}\left(\frac{L}{2Z}\right)$$
$$\equiv a + G(Z, L). \tag{6.9}$$

Clearly, Eq. (6.8) is $\Delta(Z, L, a) = 0$.

A qualitative plot of $\Delta(Z, L, a)$ for some arbitrary value of a and Z is given in Fig. 4. If we make this value of Z our critical temperature Z_c, then, by varying a until the minimum in the curve just touches the L axis, we have $L_{\min} = L_c$. That is, $(\partial\Delta/\partial L) = 0$ gives L_c, where

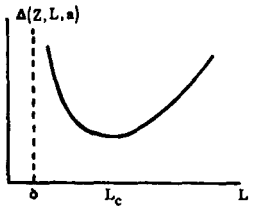

FIG. 4. By choosing a value for Z_c (within certain bounds), we can determine L_c (within certain bounds) by suitably varying the parameter a.

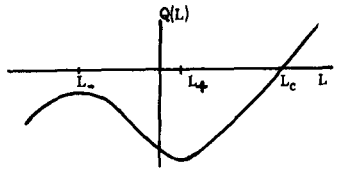

FIG. 5. Qualitative plot of $Q(L) \sim d\Delta/dL$ to show that $\Delta(Z, L, a)$ has only a single minimum in the region $\delta < L < \infty$.

$$L^2(L - \delta)\frac{\partial\Delta}{\partial L} = \frac{\alpha L^2}{4Z}(L - \delta)$$
$$+ (2Z - \delta)L - 2Z\delta \equiv Q(L). \tag{6.10}$$

Equation (6.10) allows us to choose Z_c and L_c independently. Later we see that there are regions for a choice of Z_c and L_c. Then Eq. (6.10) set equal to zero gives us a value for α. Finally, to determine a we need solve Eq. (6.9) for

$$\Delta(Z_c, L_c, a) = 0. \tag{6.11}$$

In Fig. 5 we show a qualitative plot of $Q(L)$. From $Q'(L)$ we can determine that it will have inflection points at L_+ and L_-, given by

$$L_\pm = \tfrac{1}{3}\delta\{1 \pm [1 + (12Z/\alpha\delta^2)(\delta - 2Z)]^{\frac{1}{2}}\}.$$

Let us normalize by setting $\delta = 1$. We note that $Q(1) = -1$, $Q(0) = -2Z_c$. Further,

$$Q'(L) = (\alpha L/4Z)(3L - 2) - (1 - 2Z),$$

where henceforth Z means Z_c. We are interested only in the region $L > 1$. We distinguish the two cases:

$$\text{(A) } (1 - 2Z) \leq 0, \quad \text{(B) } (1 - 2Z) > 0.$$

Case A: $Q'(L)$ is always positive ($L > 1$), and hence the slope of Q is always positive. $Q(1) = -1$ and hence Q steadily increases as L increases. There is only one root of $Q(L) = 0$ and, hence, of $\Delta'(L) = 0$ in the range $L > 1$.

Case B: Here $L_\pm$ can be unbounded. We observe that $Q'(0) = -(1 - 2Z) < 0$. [Recall that $Q(0) = -2Z > -1$.] L_- is negative. The Q curve must pass through $Q(1) = -1$. Either $L_+ \leq 1$ or $L_+ \geq 1$; in either case, again there is only one root of $Q(L)$ and hence of $\Delta'(L)$.

We conclude that $\Delta(L)$ has only one extremum point in the range $\delta \leq L < \infty$. The other two roots of Q (and hence Δ') may be real or complex, depending upon whether the left side of the curve intersects the L axis. This point is of no further interest to us.

Once a and α are determined for some value of L_c and Z_c, then from Fig. 4 we see that, by increasing Z, $\Delta(Z, L, a)$ descends, intersecting smaller and larger values of L for the vapor dome. Clearly, then, L_c can take on values greater than or less than L_0.

We now inspect our equations to determine what values are available for the critical temperature and critical volume. As we shall see, there is a certain region in which we have complete freedom, but we may not move outside. In the remainder of this section all values of L and Z refer to critical values L_c and Z_c, respectively. Frequently we find it convenient to substitute $Z = g\beta$ (again, these are critical values).

Let us set Eq. (6.10) equal to zero to obtain the critical volume L_c. From this equation, as a condition that $\alpha > 0$ we obtain

$$2Z < L/(L-1) \equiv W(L). \tag{6.12}$$

This curve $W \equiv L/(L-1)$ then fixes a boundary for allowed values of $2Z$.

$2\zeta_+$ has been discussed previously. $2\zeta_+$ can intersect $W(L)$ at only one point, $L = \frac{5}{3}$. By considering its derivative, $2\zeta_+$ has a minimum at $L = \sqrt{2}$ when it lies below $W(L)$; it then rises, but always lies below the curve $W(L)$ as $L \to 1$.

For a given value of α, Eq. (6.10) gives the relationship between Z and L when $\Delta'(Z_c, L_c, a) = 0$. The quadratic equation gives

$$2Z_\pm = \frac{L \pm L[1 - 2\alpha(L-1)^2]^{\frac{1}{2}}}{2(L-1)}.$$

To keep $Z_\pm$ real, $L \leq L_1 \equiv 1 + (2\alpha)^{-\frac{1}{2}}$. This sets the "right-hand" boundary to values of L. $2Z_\pm$ are drawn in Fig. 6.

For a finite $\alpha > 0$, $2Z_+$ clearly always lies below $W(L)$; as $\alpha \to 0$, $2Z_+ \to W(L)$. $2Z_+$ always lies above $2\zeta_+$ until it intersects it, after which it always lies below. The point of intersection results from simply setting $2Z_+ = 2\zeta_+$ to obtain the cubic equation

$$t^3(\alpha^2 + 2\alpha) + 3t - 2 = 0,$$

where $t = L - 1$. This equation clearly has only one real root.[9] For $\alpha = 0$ this point of intersection is at $L = \frac{5}{3}$. At L_1, $2Z_+$ meets the curve $(\frac{1}{2}\alpha)^{\frac{1}{2}}L$ for the first time and takes the value $\frac{1}{2}W(L_1)$.

The allowed values of $2Z_+$ then lie in this region from the intersection of $2Z_+$ with $2\zeta_+(L_2)$ to the joining with $(\frac{1}{2}\alpha)^{\frac{1}{2}}L$ at L_1. Now as α increases, the point of intersection (L_2) moves to the left as L_1 moves to the left. This intersection point sweeps a lower range to $2Z_+$, namely, the curve $W/2$. At $\alpha = 8$, $L_1 = L_2 = \frac{5}{4}$. This can be seen by simply solving

$$2Z_+ = \frac{L}{2(L-1)} = 2\zeta_+, \quad \text{for} \quad L = 1 + (2\alpha)^{-\frac{1}{2}}.$$

$2Z_-$ always lies below $(\frac{1}{2}\alpha)^{\frac{1}{2}}L$ $(1 \leq L \leq L_1)$, except for the single point $L = L_1$ where they intersect. $2Z_-$ starts at $L = 1$ with slope $\frac{1}{2}\alpha$. Its derivative can never vanish and is initially positive at $L = 1$; therefore, $2Z_-$ is always positive and intersects $(\frac{1}{2}\alpha)^{\frac{1}{2}}L$ only at L_1 where it takes the value $W/2$. As $\alpha \to 0$, $2Z_- \to 0^+$. When $\alpha = 8$, $2Z_-$ intersects both $2\zeta_+$ and $W/2$ at $L = \frac{5}{4}$. For larger values of α, we investigate the Taylor series expansion for $W/2$ and $2\zeta_+$ near $L = \frac{5}{4}$ (see Fig. 6):

$$\frac{W}{2} \equiv \frac{L}{2(L-1)} = \tfrac{5}{2} - 8(L - \tfrac{5}{4}) + 32(L - \tfrac{5}{4})^2 + \cdots,$$

$$2\zeta_+ = \tfrac{5}{2} - \tfrac{4}{3}(L - \tfrac{5}{4}) + \cdots.$$

And hence $W/2$ rises above $2\zeta_+$ as $L \to 1$. There is no other point of intersection for these two curves. Contrary to the case now for $2Z_+$, α can continue to higher values. As $2Z_-$ moves steeper and steeper, it first cuts across $2\zeta_+$ and then $W/2$. All values of $2Z_-$

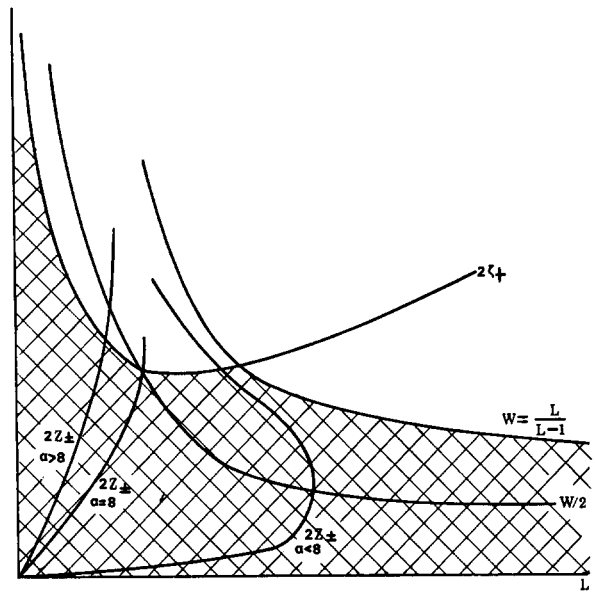

FIG. 6. Qualitative plot of $2\zeta_+$, $W \equiv L_c/(L_c - 1)$, $W/2$ as functions of L_c. $2Z_c$ (for $2Z_{c+}$ and $2Z_{c-}$) is plotted against L_c for three values of α. W intersects $2\zeta_+$ at $L_c = \frac{5}{3}$ where the two functions take the common value $\frac{5}{2}$. $W/2$, $2\zeta_+$, and $2Z_{c\pm}$ for $\alpha = 8$ intersect at a common point for $L_c = \frac{5}{4}$, at which the value of the three functions is $\frac{5}{2}$. The minimum in $2\zeta_+$ occurs at $L_c = \sqrt{2}$ such that $2\zeta_+(\sqrt{2}) = 1 + \sqrt{2}$. Shaded region is allowed values of $2Z_c$. (The subscript has purposely been omitted on L_c in the drawing.)

[9] V. I. Smirnov, *A Course of Higher Mathematics* (Pergamon Press, Inc., New York, 1964), Vol. 1, p. 491.

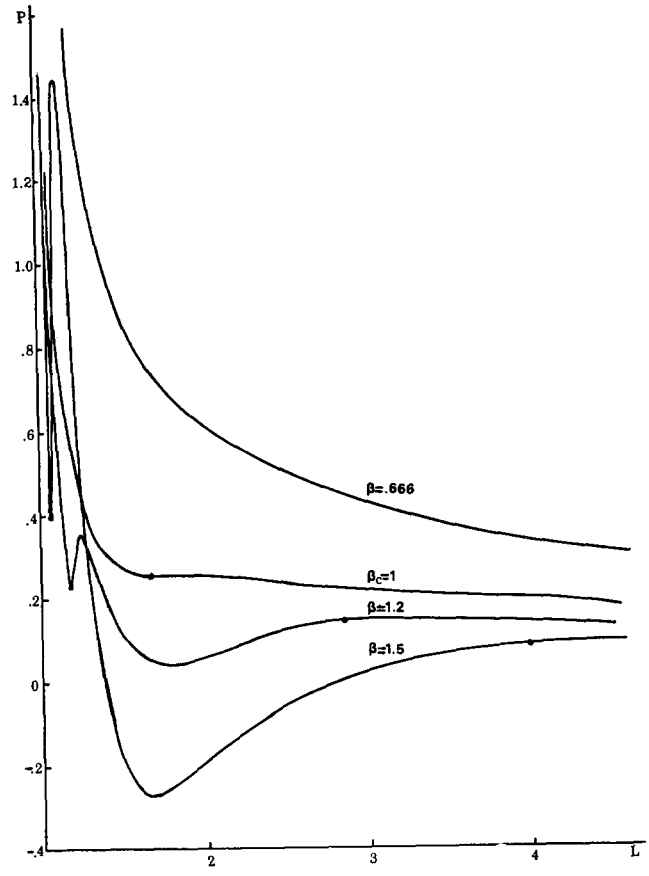

FIG. 7. P–L isotherms for 4 values of the (inverse) temperature. $L_c = \frac{5}{3}, Z_c = 1.0, \alpha = 0.720, a = -0.01629067$. β_c is the critical temperature. The dots on the isotherms locate the transition points.

below $2\zeta_+$ are allowed. As $\alpha \to \infty$, $2Z_-$ climbs steeper, but $2\zeta_+ \to \infty$ as $L \to 1$ so that there is no bound upon $2Z_-$. $2Z_+$ now clearly lies outside the region.

The description of the boundary for Z_c as a function of L_c is complete. The lower bound is zero. The upper bound is $2\zeta_+$ for the range $1 \leq L \leq \frac{5}{3}$; then the curve $W(L)$ takes over for $\frac{5}{3} \leq L \leq L_1 \equiv 1 + (2\alpha)^{-\frac{1}{2}}$. At L_1 there arises a "right-hand" boundary which forbids larger values of L. In the allowed region the curve $2Z_+$ and $2Z_-$ then relates Z_c to L_c. In Fig. 6 those values are allowed for which $2Z$ lies in the shaded region.

From Fig. 6 one can assign initially an arbitrary value of Z_c and L_c as long as the pair of values lies within the bounded region. Depending upon whether this value lies above or below the curve, $W/2$ determines whether one must vary Z_+ or Z_-. By changing α one can force either Z_+ or Z_- to pass through the selected point. Then, with this value of α, a can be determined from Eq. (6.11).

We can obtain a qualitative idea of how a must vary as the parameters change. On the vapor dome and in particular at the critical point, Eq. (6.9) vanishes. Solving for $2Z'_+$, we obtain

$$\begin{aligned} 2Z'_+ &\equiv 2\theta_+ \\ &= L[a - \ln(1 - L^{-1})] \\ &\quad + L\{[a - \ln(1 - L^{-1})]^2 + \tfrac{1}{2}\alpha\}^{\frac{1}{2}}. \end{aligned}$$

We see that if L is kept fixed, then θ_+ increases if either α or a increases. But the curve $2Z_+$ decreases as α increases and hence a must decrease in order to compensate.

In Fig. 7 we plot a few isotherms for the values $L_c = \frac{5}{3}$, $Z_c = 1$. We have set $\delta = g = 1$. Then $\alpha = 0.720$, $a = -0.0162$. The sharp discontinuity in $\partial P/\partial L$ is evident at the liquid transition point. At the gas transition point we use the computer data to calculate the discontinuity and we find it to be remarkably sharp. For example, for $\beta = 1.20$,

$$\left(\frac{\partial P}{\partial L}\right)_{2\text{ phase}} = \frac{0.13990 - 0.13965}{2.8450 - 2.8400} = 0.05,$$

$$\left(\frac{\partial P}{\partial L}\right)_{1\text{ phase}} = \frac{0.13991 - 0.13992}{2.90 - 2.85} = -0.0002.$$

For $\beta = 1.5$,

$$\left(\frac{\partial P}{\partial L}\right)_{2\text{ phase}} = \frac{0.07864 - 0.07862}{3.9935 - 3.9930} = 0.04,$$

$$\left(\frac{\partial P}{\partial L}\right)_{1\text{ phase}} = \frac{0.07865 - 0.07865}{3.9945 - 3.9940} \approx 0.$$

We are plagued with Maxwell loops, negative pressures, and with the pressure not decreasing with decreasing temperature in the two-phase region. Nor is the vapor dome decreasing as one moves toward the liquid side of the critical point.

APPENDIX A

1. Preliminary Discussion

As we see further on in our analysis, $K(x, \gamma)$ always vanishes in the limit as $\gamma \to 0$, for x finite. The only way to keep $K(x, \gamma)$ finite is to allow x to go to zero with γ:

$$K(x,\gamma) \equiv \int_0^{2\pi} \frac{\gamma\left(\gamma\cos\frac{\omega}{2} - \frac{\omega}{L}\sin\frac{\omega}{2}\right)d\omega}{[1+x]\left[\gamma^2 + \left(\frac{\omega}{L}\right)^2\right] - \gamma\left(\gamma\cos\frac{\omega}{2} - \frac{\omega}{L}\sin\frac{\omega}{2}\right)}. \tag{A1}$$

The important terms are the first-order terms in the expansion of the numerator; the higher-order terms, because of the contribution of a term of ω^4 or higher, clearly vanish as $\gamma \to 0$. Call K' these first-order terms, so that

$$K'(x,\gamma) \equiv \int_0^{2\pi} \frac{\gamma^2 - \frac{\gamma\omega^2}{2L}\, d\omega}{[1+x]\left[\gamma^2 + \left(\frac{\omega}{L}\right)^2\right] - \gamma\left(\gamma\cos\frac{\omega}{2} - \frac{\omega}{L}\sin\frac{\omega}{2}\right)}$$
$$\equiv K_1' - K_2'.$$

We now show that even for $x = 0$, one of these terms, K_2', vanishes as $\gamma \to 0$:

$$K_2' \equiv \frac{\gamma}{2L}\int_0^{2\pi} \frac{\omega^2/[\gamma^2 + (\omega/L)^2]\, d\omega}{[1+x] - \left\{\gamma\Big/\left[\gamma^2 + \left(\frac{\omega}{L}\right)^2\right]\right\}\left(\gamma\cos\frac{\omega}{2} - \frac{\omega}{L}\sin\frac{\omega}{2}\right)}.$$

Since $x \geq 0$,

$$K_2' \leq \frac{\gamma}{2L}\int_0^{2\pi} \frac{\omega^2/[\gamma^2 + (\omega/L)^2]\, d\omega}{1 - \left\{\gamma\Big/\left[\gamma^2 + \left(\frac{\omega}{L}\right)^2\right]\right\}\left(\gamma\cos\frac{\omega}{2} - \frac{\omega}{L}\sin\frac{\omega}{2}\right)}, \tag{A2}$$

since in the range $\omega = [0, 2\pi]$, for sufficiently small γ,

$$\left\{\gamma\Big/\left[\gamma^2 + \left(\frac{\omega}{L}\right)^2\right]\right\}\left(\gamma\cos\frac{\omega}{2} - \frac{\omega}{L}\sin\frac{\omega}{2}\right) \leq 1,$$
$$\left|\gamma\cos\frac{\omega}{2} - \frac{\omega}{L}\sin\frac{\omega}{2}\right| \leq \gamma + \frac{\omega^2}{2L}, \tag{A3}$$
$$\left\{\gamma\Big/\left[\gamma^2 + \left(\frac{\omega}{L}\right)^2\right]\right\}\left|\gamma\cos\frac{\omega}{2} - \frac{\omega}{L}\sin\frac{\omega}{2}\right| \leq \frac{\gamma^2 + (\gamma\omega^2/2L)}{\gamma^2 + (\omega/L)^2}.$$

Therefore,

$$\left|[1+x] - \left\{\gamma\Big/\left[\gamma^2 + \left(\frac{\omega}{L}\right)^2\right]\right\}\left(\gamma\cos\frac{\omega}{2} - \frac{\omega}{L}\sin\frac{\omega}{2}\right)\right| \geq [1+x] - \frac{\gamma^2 + (\gamma\omega^2/2L)}{\gamma^2 + (\omega/L)^2}. \tag{A4}$$

Therefore, Eq. (A2) becomes, upon replacing the denominator by Eq. (A4),

$$|K_2'| \leq \frac{\gamma}{2L}\int_0^{2\pi} \frac{d\omega}{\left[\frac{1}{L^2} - \frac{\gamma}{2L}\right]} \to 0.$$

Finally, we are left with K_1' as the only important part of $K(x,\gamma)$.

2. Rate $x \to 0$ in the Transition Region

We now evaluate $K_1'(x,\gamma)$ as $\gamma \to 0$ in the transition region. We shall see that x behaves as γ^2. We evaluate K_1' rigorously. We do this by showing that K_1' is bounded from above and from below. In the limit as $\gamma \to 0$ the two bounds converge to the same value. Before starting, we might note that it is at $\omega = 0$ that the integral can blow up.

We now obtain this value of $K(x,\gamma)$. We set the upper bound by using Eq. (A4):

$$K_1'(x,\gamma) \leq \int_0^{2\pi} \frac{\gamma^2}{\gamma^2 x + A\omega^2}\, d\omega = \frac{\gamma}{(Ax)^{\frac{1}{2}}}\tan^{-1}\frac{2\pi}{\gamma}\left(\frac{A}{x}\right)^{\frac{1}{2}},$$

where

$$A \equiv [(1+x)/L^2] - (\gamma/2L).$$

We see that, to prevent K_1' from vanishing,

$$x \cong \gamma^n, \quad n \geq 2. \tag{A5}$$

We now set a lower bound to $K_1'(x,\gamma)$. Clearly, because of Eq. (A3), the following relation is valid:

$$[1+x] - \frac{\gamma}{\gamma^2 + (\omega/L)^2}\left(\gamma\cos\frac{\omega}{2} - \frac{\omega}{L}\sin\frac{\omega}{2}\right) \geq 0.$$

Let us now construct the following function:

$$F(\omega) \equiv \gamma^2 x + \omega^2\left[\frac{1+x}{L^2} + \frac{\gamma^2}{8} + \frac{\gamma}{2L}\right] - \left[\gamma^2 + \left(\frac{\omega}{L}\right)^2\right][1+x] + \gamma\left(\gamma\cos\frac{\omega}{2} - \frac{\omega}{L}\sin\frac{\omega}{2}\right).$$

We note that $F(0) = F'(0) = 0$. We now show that over a reasonably finite range of ω ($\omega < 2$), $F(\omega)$ is always nonnegative:

$$F(\omega) \geq 0, \quad \omega < 2.$$

We proceed to expand $F(\omega)$ in power series, so that

$$F(\omega) = \gamma \sum_{n=1}^{\infty} (-1)^{n+1} \left(\frac{\omega}{2}\right)^{2n+2} \left[\frac{\gamma}{(2n+2)!} + \frac{2/L}{(2n+1)!}\right].$$

We separate the expansion into two parts:

$$\frac{F(\omega)}{\gamma} = \sum_{\substack{n=1,3,5\cdots \\ \text{odd}}} \left(\frac{\omega}{2}\right)^{2n+2} \left[\frac{\gamma}{(2n+2)!} + \frac{2/L}{(2n+1)!}\right] - \sum_{\substack{n=2,4,6 \\ \text{even}}} \left(\frac{\omega}{2}\right)^{2n+2} \left[\frac{\gamma}{(2n+2)!} + \frac{2/L}{(2n+1)!}\right],$$

$$= \sum_{\substack{n=1,3,5 \\ \text{odd}}} \left(\frac{\omega}{2}\right)^{2n+2} \left\{\left[\frac{\gamma}{(2n+2)!} + \frac{2/L}{(2n+1)!}\right] - \left(\frac{\omega}{2}\right)^{2} \left[\frac{\gamma}{(2n+4)!} + \frac{2/L}{(2n+3)!}\right]\right\}.$$

Clearly $F(\omega) \geq 0$ for $(\omega/2) < 1$.

The equivalent statement is

$$\gamma^2 x + \omega^2 \left[\frac{1+x}{L^2} + \frac{\gamma^2}{8} + \frac{\gamma}{2L}\right] \geq \left[\gamma^2 + \left(\frac{\omega}{L}\right)^2\right][1+x] - \gamma\left(\gamma \cos\frac{\omega}{2} - \frac{\omega}{L}\sin\frac{\omega}{2}\right), \quad \omega < 2.$$

Since both sides of the inequality are nonnegative, we can dispense with absolute value signs. Then

$$K_1'(x,\gamma) \geq \int_0^{2\pi} \frac{\gamma^2\, d\omega}{\gamma^2 x + \omega^2[(1+x)/L^2 + \frac{1}{8}\gamma^2 + \gamma/2L]} = \frac{\gamma^2}{B}\int_0^{2\pi} \frac{d\omega}{(\gamma^2 x/B) + \omega^2} = \frac{\gamma}{(Bx)^{\frac{1}{2}}} \tan^{-1} \frac{2\pi}{\gamma}\left(\frac{B}{x}\right)^{\frac{1}{2}},$$

where the definition of B is clear.

To prevent the blow-up of K_1', $x \cong \gamma^n$, $n \leq 2$. From the previous relation [Eq. (A5)], we conclude that $x \cong \gamma^2$. $B \to A$ as $\gamma \to 0$, and we see that the upper and lower bounds converge to the same value.

We might note that it is possible for $K_1'(x,\gamma)$ to vanish as is necessary on the "vapor dome." Here then, x cannot go as γ^2. This means $x = \gamma^n$, $n < 2$.

3. Rate $x \to 0$ on the Vapor Dome

We wish to calculate the rate $x \to 0$ on the vapor dome. This is the point where the phase transition just sets in: $K(x,\gamma) \to 0$ as the two asymptotes just touch.

The equation of the saddle point is

$$\frac{2S_s}{\alpha} = \frac{1}{2\pi} \frac{1}{-\beta\lambda(0)} \frac{1}{1+x} \{2\pi + K(x,\gamma)\},$$

where now

$$\lim_{\gamma \to 0} K(x,\gamma) = 0.$$

Substituting in the values expanded to first order in γ, we have

$$\frac{2}{\alpha}\left[-a + \ln\left(\frac{L-\delta}{L}\right) + \frac{\beta g \gamma}{2} + \frac{2g\beta}{L} - \frac{2g\beta}{L}\frac{\gamma L}{2} + \frac{2g\beta}{L}x\right] = \frac{L}{2g\beta} + \frac{L}{2g\beta}\frac{\gamma L}{2} - \frac{L}{2g\beta}x + \frac{L}{2g\beta}\frac{1}{2\pi}\frac{\gamma}{(Ax)^{\frac{1}{2}}}\frac{\pi}{2}, \quad \text{(A6)}$$

where $K(x,\gamma)$ is replaced by its limiting value

$$\lim_{\gamma \to 0} K(x,\gamma) = \frac{\gamma}{(Ax)^{\frac{1}{2}}} \tan^{-1} \frac{2\pi}{\gamma}\left(\frac{A}{x}\right)^{\frac{1}{2}} = \frac{\gamma}{(Ax)^{\frac{1}{2}}}\frac{\pi}{2}.$$

Now on the vapor dome as $\gamma \to 0$, $x \to 0$ and $K(x,\gamma) \to 0$. Therefore, those terms on both sides of Eq. (A6) which are independent of γ are equal. We have

$$\frac{2}{\alpha}\left[\frac{\beta g \gamma}{2} - \frac{2g\beta}{L}\frac{\gamma L}{2} + \frac{2g\beta}{L}x\right] = \frac{L}{2g\beta}\frac{\gamma L}{2} - \frac{L}{2g\beta}x + \frac{L}{2g\beta}\frac{1}{4}\frac{\gamma}{(Ax)^{\frac{1}{2}}}.$$

For brevity we write

$$Bx = C\gamma + D\gamma/x^{\frac{1}{2}},$$

where B, C, D are obvious from the previous equation. Now for $\gamma/x^{\frac{1}{2}}$ to vanish, $x = \gamma^{2(1-\epsilon)}$, $0 < \epsilon < 1$. Therefore, our equation becomes

$$B\gamma^{2(1-\epsilon)} = C\gamma + D\gamma^{\epsilon}.$$

γ^{ϵ} is the leading term on the right:

$$B\gamma^{2(1-\epsilon)} = D\gamma^{\epsilon},$$
$$2(1-\epsilon) = \epsilon,$$
$$\epsilon = \tfrac{2}{3}, \quad x \cong \gamma^{\frac{2}{3}}.$$

On the vapor dome $x \cong \gamma^{\frac{2}{3}}$.

APPENDIX B

In this appendix we demonstrate that for sufficiently large β the curve $2S_s/\alpha$ crosses above the curve $[-\beta\lambda(0)]^{-1}(1+x)^{-1}$ at $x = 0$ in the limit $\gamma \to 0$. The significance of this result is that the quantity $K(x,\gamma)$ cannot vanish, but must take on the proper value in order to satisfy the saddle-point equation. The range of values of L for which

$$-\frac{2\tilde{S}}{\alpha} \geq \frac{1}{-\beta\lambda(0)} \quad \text{(B1)}$$

we call the transition or condensation region.

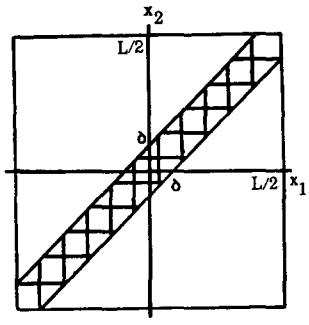

FIG. 8. Region of integration for calculation of $\mathfrak{A}_1$. The cross-hatched region is bounded above by the curve $x_2 - x_1 = \delta$ and below by $x_1 - x_2 = \delta$. The cross-hatched region is excluded from the integration.

We need to calculate βa_{11}. From its definition we have

$$\mathfrak{A}_1 \equiv \exp[-2\beta a_{11}] = \frac{1}{L^2}\int dx_1 \int dx_2 \exp[-\beta W(12)]. \tag{B2}$$

Since $\exp[-\beta W(12)] = 0$ for $|x_1 - x_2| \leq \delta$, the contribution to $\mathfrak{A}_1$ from the cross-hatched region in Fig. 8 vanishes. If we call the remaining region R_1, then

$$\mathfrak{A}_1 = \frac{1}{L^2}\int_{R_1} dx_1 \int dx_2 \exp[\beta g\gamma e^{-\gamma R}], \quad R = |x_1 - x_2|.$$

We expand the exponential in the integrand:

$$\mathfrak{A}_1 = \frac{1}{L^2}\int_{R_1} dx_1 \int dx_2 \times \left\{1 + \beta g\gamma e^{-\gamma R} + \frac{(\beta g\gamma)^2}{2!} e^{-2\gamma R} + \cdots\right\}.$$

The first term in the expansion is simply the area of R_1:

$$2[(L/2) - \delta]^2 + 2(L/2)^2 - \delta^2 \equiv (L - \delta)^2.$$

We are interested, as $\gamma \to 0$. The remaining terms vanish in this limit, since

$$\int_{R_1} dx_1 \int dx_2 \frac{(\beta g\gamma)^n}{n!} e^{-n\gamma R} \leq \frac{(\beta g\gamma)^n}{n!} L^2.$$

We conclude with

$$\mathfrak{A}_1 = \frac{(L-\delta)^2}{L^2} + \mathcal{O}(\gamma) = \left(\frac{L-\delta}{L}\right)^2 [1 + \beta g\gamma] + \mathcal{O}(\gamma^2). \tag{B3}$$

We now find the conditions sufficient for the validity of the relationship in Eq. (B1). By making use of the definition of $\tilde{S}$, Eqs. (B2) and (B3), we have

$$[\lambda(0)]^2 + \lambda(0)\left[\frac{a - \ln\left(1 - \frac{\delta}{L}\right)}{\beta}\right] \geq \frac{\alpha}{2}\frac{1}{\beta^2}. \tag{B4}$$

Since $\lambda(0)$ is independent of β, Eq. (B4) can be satisfied by sufficiently large β.

APPENDIX C

These calculations allow us to simplify the expression for the pressure equation (6.1).

Consider the integral

$$\mathfrak{L} \equiv \int_0^{2\pi} \frac{[\partial\Lambda(\omega)/\partial L] - [\partial\Lambda(0)/\partial L]}{S_s + a + \beta\Lambda(\omega)}\, d\omega = \int_0^{2\pi} \frac{[\partial F(\omega)/\partial L]}{\epsilon + \beta F(\omega)}\, d\omega, \tag{C1}$$

where we define a function

$$F(\omega) \equiv \lambda(\omega) - \lambda(0) \equiv \Lambda(\omega) - \Lambda(0).$$

We calculate

$$\frac{\partial F}{\partial L} = -\left(\frac{1}{L} + \frac{\gamma}{2}\right)F(\omega) + \frac{\lambda(\omega)(2/L)(\omega/L)^2}{\gamma^2 + (\omega/L)^2} - \frac{2g}{L}\frac{e^{-\gamma L/2}\gamma}{\gamma^2 + (\omega/L)^2}\frac{\omega}{L^2}\sin\frac{\omega}{2}. \tag{C2}$$

Placing Eq. (C2) into Eq. (C1), we can easily show that the contributions to the integral from the second and third terms of (C2) vanish as $\gamma \to 0$. Proof is as follows:

$$\int_0^{2\pi} \frac{\omega^2\lambda(\omega)/[\gamma^2 + (\omega/L)^2]}{\epsilon + \beta F(\omega)}\, d\omega \leq \int_0^{2\pi} \frac{\{\omega^2/[\gamma^2 + (\omega/L)^2]\} \times \{[\gamma^2 + (\gamma\omega^2/2L)]/[\gamma^2 + (\omega/L)^2]\}}{1 - [\gamma^2 + (\gamma\omega^2/2L)]/[\gamma^2 + (\omega/L)^2]}\, d\omega,$$

where the inequalities (A3) and (A4) have been used. The right-hand side of the last equation is equal to

$$\frac{L^2}{1 - \frac{1}{2}\gamma L}\int_0^{2\pi} \frac{\gamma^2 + \gamma\omega^2/2L}{\gamma^2 + (\omega/L)^2}\, d\omega \to 0.$$

The last integral is

$$\gamma\int_0^{2\pi} \frac{(\omega/L^2)\{\sin(\omega/2)/[\gamma^2 + (\omega/L)^2]\}}{\epsilon + \beta F(\omega)}\, d\omega \sim \gamma\int_0^{2\pi} \frac{\omega^2/[\gamma^2 + (\omega/L)^2]}{\epsilon + \beta F(\omega)}\, d\omega,$$

using inequalities (A4). This integral behaves as K_2' in Eq. (A2).

Therefore, $\mathfrak{L}$ becomes

$$\mathfrak{L} = -\left(\frac{1}{L} + \frac{\gamma}{2}\right)\int_0^{2\pi} \frac{F(\omega)}{\epsilon + \beta F(\omega)}\, d\omega + \mathcal{O}(\gamma).$$

By simple division of the integrands, we obtain

$$\mathfrak{L} = -\frac{1}{\beta}\left(\frac{1}{L}+\frac{\gamma}{2}\right)\int_0^{2\pi}\left[1-\frac{\epsilon}{\epsilon+\beta F(\omega)}\right]d\omega$$
$$= -\frac{1}{\beta}\left(\frac{1}{L}+\frac{\gamma}{2}\right)\left[2\pi-\epsilon\left(\frac{4\pi S_s}{\alpha}\right)\right],$$

using the saddle-point equation

$$\mathfrak{L} = -\frac{2\pi}{\beta L}\left[1-\frac{2\epsilon}{\alpha}(-\tilde{S}+\epsilon)\right], \quad \text{as} \quad \gamma \to 0.$$

APPENDIX D. BEHAVIOR OF ϵ_+

The definition of ϵ_+ arises from Eq. (6.3) by setting $K = 0$:

$$2\epsilon_\pm = \tilde{S}+\beta\lambda(0) \pm \{[\tilde{S}+\beta\lambda(0)]^2 - 4[\tilde{S}\beta\lambda(0) - \tfrac{1}{2}\alpha]\}^{\frac{1}{2}}$$
$$\equiv A \pm (A^2 - D)^{\frac{1}{2}}.$$

That we must choose the plus sign is clear from inspection of the limits when $L \to \delta$ and $L \to \infty$.

If $L \approx \delta$, the significant term is βa_{11}. Then

$$A \equiv \tilde{S}+\beta\lambda(0) \approx -\ln(1-\delta/L) \to \infty, \quad L \to \delta,$$
$$(D/4) \equiv \tilde{S}\beta\lambda(0) - \tfrac{1}{2}\alpha \approx \beta^2 a_{11}\lambda(0) \to -\infty, \quad L \to \delta,$$

and clearly the positive sign must be chosen.

If $L \to \infty$, then

$$A \approx a,$$
$$D \approx -\alpha/2,$$

and clearly the positive sign must be chosen again.

We now investigate the behavior of ϵ_+ in the transition region. We use Eq. (6.3) and set

$$\frac{2\tilde{S}}{\alpha} = \frac{1}{\beta\lambda(0)}\left(1+\frac{K}{2\pi}\right),$$
$$A = \frac{\alpha/2}{\beta\lambda(0)}\left(1+\frac{K}{2\pi}\right)+\beta\lambda(0) \leq 0,$$
$$D/4 = (\alpha K/4\pi) \geq 0.$$

We see that ϵ_+ is negative in the transition region. A qualitative plot is given in Fig. 2.